Adaptive Minimum Bit-Error Rate Equalization for Binary Signaling

Chen-Chu Yeh and John R. Barry

***Abstract*—We consider the design and adaptation of a linear equalizer with a finite number of coefficients in the context of a classical linear intersymbol-interference channel with Gaussian noise and a memoryless decision device. If the number of equalizer coefficients is sufficient, the popular minimum mean-squared-error (MMSE) linear equalizer closely approximates the optimal linear equalizer that directly minimizes bit-error rate (BER). However, when the number of equalizer coefficients is insufficient to approximate the channel inverse, the minimum-BER equalizer can outperform the MMSE equalizer by as much as 16 dB in certain cases. We propose a simple stochastic adaptive algorithm for realizing the minimum-BER equalizer. Compared to the least-mean-square algorithm, the proposed algorithm can provide a substantial reduction in BER with no increase in complexity.**

***Index Terms*—Adaptive equalizers, intersymbol interference, minimum mean-squared error (MMSE) equalization.**

I. Introduction

We consider the design and adaptation of a finite-tap linear equalizer for combating linear intersymbol interference in the presence of additive white Gaussian noise, under the constraint that decisions are made on a symbol-by-symbol basis by quantizing the equalizer output. We consider only uncoded binary and quadrature amplitude modulation (4-QAM). The most popular design strategy in this setting is the minimum mean-squared-error (MMSE) equalizer. However, as recognized in [1]–[4], a better strategy is to choose the equalizer coefficients so as to minimize the error probability or bit-error rate (BER).

Minimum-BER equalization first appeared in [1], which showed that the continuous-time linear equalizer that minimizes BER can be represented as a matched filter followed by a tapped-delay line. This work was extended to decision-feedback equalization in [2]. Neither work proposed a numerical algorithm for calculating the minimum-BER equalizer coefficients, relying instead on a brute-force computer search that requires knowledge of the channel. More recently, it was shown that, in the limit as the signal-to-noise ratio (SNR) approaches infinity, the minimum-BER equalizer maximizes the eye opening, whereas in the limit of zero SNR, the minimum-BER equalizer approaches an averaged matched filter [3]. The first adaptive algorithm for approximating the minimum-BER equalizer was proposed in [4], where receiver estimates of the channel, noise power, and noiseless channel output were used to approximate a stochastic gradient algorithm. The proposed algorithm is significantly more complex than the least-mean-square (LMS) algorithm, and—even with perfect knowledge of the channel and noise power—would be susceptible to misconvergence. In fact, the BER surface of even a simple channel can be highly irregular with multiple local minima, so that the convergence of any gradient algorithm to the global minimum cannot be guaranteed.

This paper makes two primary contributions. First, we demonstrate that there are circumstances in which the MMSE equalizer is far from optimal. We present an example in which the minimum-BER equalizer outperforms the MMSE equalizer by 16 dB. We propose a recursive algorithm for determining the exact minimum-BER equalizer coefficients, and examine its convergence properties. Second, we derive an adaptive minimum-BER (AMBER) stochastic algorithm for adaptive equalization with the following attributes: it has low complexity (no more complex than the LMS algorithm); it does not require knowledge of the channel or noise power; and it has good convergence properties.

This paper is organized as follows. In Section II, we present models for the channel and equalizer, and we introduce the concepts of signal vectors and the signal cone. We express BER as a simple function of the signal vectors. In Section III, we discuss exact minimum-BER equalization. In Section IV, we propose the AMBER algorithm for adaptive equalization, and propose algorithm modifications for faster convergence and decision-directed adaptation. We also generalize the AMBER algorithm to 4-QAM. In Section V, we present numerical results comparing the AMBER and MMSE equalizers. In Section VI, we summarize our results.

II. Background

A. Problem Statement

Consider the linear discrete-time channel depicted in Fig. 1, whose output is

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

where x_k is a binary input drawn from $\{\pm 1\}$, h_k is the channel impulse response, assumed to be casual with memory M, and

Paper approved by S. B. Gelfand, the Editor for Transmission Systems of the IEEE Communications Society. Manuscript received June 9, 1997; revised January 15, 1999 and September 15, 1999. This paper was presented in part at the IEEE International Conference on Communications, Montreal, Canada, June 8–12, 1997.

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: yeh@ece.gatech.edu; barry@ece.gatech.edu).

Publisher Item Identifier S 0090-6778(00)06164-X.

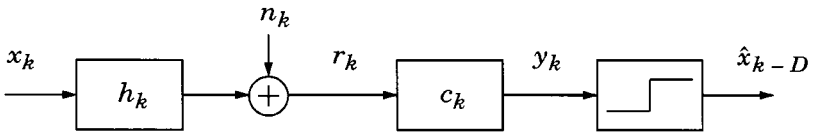


Fig. 1. Block diagram of channel, equalizer, and memoryless decision device.

n_k is white Gaussian noise with power spectral density σ^2. Also shown in Fig. 1 is an N-tap linear equalizer described by the vector $\boldsymbol{c} = [c_0 \cdots c_{N-1}]^T$. The equalizer output at time k can be expressed as the inner product $y_k = \boldsymbol{c}^T \boldsymbol{r}_k$, where the channel output vector $\boldsymbol{r}_k = [r_k \cdots r_{k-N+1}]^T$ is given by

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

where $\boldsymbol{x}_k = [x_k \cdots x_{k-M-N+1}]^T$ is a vector of channel inputs, $\boldsymbol{n}_k = [n_k \cdots n_{k-N+1}]^T$ is a vector of noise samples, and $\boldsymbol{H}$ is the $N \times (M+N)$ Toeplitz convolution matrix

$$\boldsymbol{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_M & 0 & \cdots & & 0 \\ 0 & h_0 & h_1 & \cdots & h_M & 0 & & \vdots \\ & & & & & & \ddots & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & 0 \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & & h_M \end{bmatrix}. \tag{3}$$

This paper explores the design of the equalizer $\boldsymbol{c}$ under the restrictive constraint that decisions are determined by the sign of the equalizer output. (Of course, such a memoryless decision device is suboptimal; better BER performance can be achieved by performing maximum-likelihood sequence detection or maximum *a posteriori* detection on the equalizer output.)

We define a delay parameter D to account for delays in the channel and equalizer, so that the receiver decision regarding x_{k-D} is $\hat{x}_{k-D} = \text{sgn}(y_k)$. It is clear from (2) that the equalizer output at time k depends only on the $M+N$ symbols contained in the vector $\boldsymbol{x}_k$, and hence a meaningful delay parameter will be constrained to the range $D \in \{0, \cdots, M+N-1\}$. Our focus is on designing $\boldsymbol{c}$ once D has been chosen, although we remark that a proper choice of D is also critical to achieving good performance.

By far, the most popular equalization strategy is the minimum mean-squared-error (MMSE) design, in which $\boldsymbol{c}$ is chosen as the unique vector minimizing $\text{MSE} = E[(y_k - x_{k-D})^2]$, namely [5]

$$\boldsymbol{c}_{\text{MMSE}} = (\boldsymbol{H}\boldsymbol{H}^T + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{h}_{D+1} \tag{4}$$

where $\boldsymbol{h}_{D+1}$ is the $(D+1)$th column of $\boldsymbol{H}$. This equalizer is often realized using a stochastic gradient search known as the LMS algorithm [5], [6]

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \mu(y_k - x_{k-D})\boldsymbol{r}_k \tag{5}$$

where μ is a small positive step size. When training data is unavailable, the equalizer can operate in decision-directed mode, whereby $\hat{x}_{k-D}$ is used in place of x_{k-D}.

B. An Expression for BER

The BER after any equalizer $\boldsymbol{c}$ and the decision rule $\hat{x}_{k-D} = \text{sgn}(y_k)$ is

$$\begin{aligned} Pr[\hat{x}_{k-D} \neq x_{k-D}] &= Pr[x_{k-D} y_k < 0] \\ &= Pr[x_{k-D}\boldsymbol{c}^T\boldsymbol{H}\boldsymbol{x}_k + x_{k-D}\boldsymbol{c}^T\boldsymbol{n}_k < 0] \\ &= E\left[Pr[x_{k-D}\boldsymbol{c}^T\boldsymbol{H}\boldsymbol{x}_k + x_{k-D}\boldsymbol{c}^T\boldsymbol{n}_k \right. \\ &\qquad \left. < 0 | \boldsymbol{x}_k]\right] \\ &= E\left[Q\left(\frac{\boldsymbol{c}^T\boldsymbol{H}\boldsymbol{x}_k x_{k-D}}{\|\boldsymbol{c}\|\sigma}\right)\right] \end{aligned} \tag{6}$$

where the expectations are over the 2^{M+N} equally likely $\boldsymbol{x}_k$ binary vectors $\{\pm 1\}^{M+N}$, where $\|\boldsymbol{c}\|^2 = \boldsymbol{c}^T\boldsymbol{c}$, and where Q is the Gaussian error function [6]. Observe that the product $\boldsymbol{x}_k x_{k-D}$ is a binary vector with a one in the $(D+1)$th component, and let $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \cdots, \tilde{\boldsymbol{x}}_L$ denote any ordering of the $L = 2^{M+N-1}$ such distinct vectors. We define the corresponding *signal vectors* by

$$\boldsymbol{s}_i = \boldsymbol{H}\tilde{\boldsymbol{x}}_i, \qquad i = 1 \cdots L. \tag{7}$$

These signal vectors will play a fundamental role in our analysis. From (2), we see that these signal vectors represent the L possible noiseless channel output vectors given that the desired symbol is 1. With this definition, (6) simplifies to

$$\text{BER} = \frac{1}{L}\sum_{i=1}^{L} Q\left(\frac{\boldsymbol{c}^T\boldsymbol{s}_i}{\|\boldsymbol{c}\|\sigma}\right). \tag{8}$$

Because decisions are based on the sign of the equalizer output, the equalizer norm $\|\boldsymbol{c}\|$ is irrelevant; the BER depends on the equalizer direction $\boldsymbol{c}/\|\boldsymbol{c}\|$ only.

In the remainder of the paper, we will often assume that the channel is *equalizable*.

Definition 1: A channel is said to be *equalizable* by an N-tap equalizer with delay D if and only if there exists an equalizer $\boldsymbol{c}$ having a positive inner product with all signal vectors $\{\boldsymbol{s}_i\}$.

But a positive inner product with all $\{\boldsymbol{s}_i\}$ vectors implies that the noiseless equalizer output is always positive when a one was transmitted ($x_{k-D} = 1$); thus, a channel is equalizable if and only if its noiseless eye diagram can be opened. In terms of the $\{\boldsymbol{s}_i\}$ vectors, a channel is equalizable if and only if there exists a hyperplane passing through the origin such that all $\{\boldsymbol{s}_i\}$ vectors are strictly on one side of the hyperplane.

We define the *signal cone* of an equalizable channel as the span of the signal vectors with positive coefficients:

Definition 2: The *signal cone* of an equalizable channel is the set $S = \{\Sigma_i a_i \boldsymbol{s}_i : a_i > 0\}$.

The set S is a cone because there exists at least one "axis" vector within S that forms a positive inner product with all other elements of S. The zero vector is not an element of the signal

cone of an equalizable channel, because if all $\{\boldsymbol{s}_i\}$ are strictly on one side of the hyperplane, then so will a linear combination with positive coefficients. We remark that if the channel is not equalizable, the set $\{\Sigma_i a_i \boldsymbol{s}_i : a_i > 0\}$ has no such axis vector and is thus not a cone; rather, the set is a linear subspace of $\mathbb{R}^N$, and it includes the zero vector.

III. Exact Minimum BER Equalization

A. Fixed-Point Equation

Let $\boldsymbol{c}_{\mathrm{EMBER}}$ denote an equalizer that achieves exact minimum-BER (EMBER) performance, minimizing (8). Because the BER (8) depends only on the direction of the equalizer, $\boldsymbol{c}_{\mathrm{EMBER}}$ is not unique: if $\boldsymbol{c}$ minimizes BER, then so does $b\boldsymbol{c}$ for any positive constant b. Unlike the coefficient vector $\boldsymbol{c}_{\mathrm{MMSE}}$ (4) that minimizes MSE, there is no closed-form expression for $\boldsymbol{c}_{\mathrm{EMBER}}$. However, as shown below, $\boldsymbol{c}_{\mathrm{EMBER}}$ must satisfy an important fixed-point relationship.

The gradient of the BER (8) is

$$\nabla_{\boldsymbol{c}}\mathrm{BER} = \frac{-1}{\sqrt{2\pi}\sigma\|\boldsymbol{c}\|}\left(\boldsymbol{I} - \frac{\boldsymbol{c}\boldsymbol{c}^T}{\|\boldsymbol{c}\|^2}\right) f(\boldsymbol{c}) \tag{9}$$

where the vector function $f\colon \mathbb{R}^N \to \mathbb{R}^N$ is defined by

$$f(\boldsymbol{c}) = \frac{1}{L}\sum_{i=1}^{L} e^{-z_i^2/2}\boldsymbol{s}_i \tag{10}$$

and where z_i is a normalized inner product between $\boldsymbol{s}_i$ and $\boldsymbol{c}$

$$z_i = \frac{\boldsymbol{c}^T\boldsymbol{s}_i}{\|\boldsymbol{c}\|\sigma}. \tag{11}$$

Recall that the inner products $\{\boldsymbol{c}^T\boldsymbol{s}_i\}$ represent the L possible noiseless equalizer outputs given that $x_{k-D} = 1$. Hence, the numbers $\{z_1, \cdots, z_L\}$ determine (up to an inconsequential positive scaling factor) the values of the noiseless eye diagram at the sampling instant. For example, the zero-forcing equalizer satisfies $z_1 = z_2 = \cdots = z_L$. The eye will be open when $z_i > 0$ for all $i = 1 \cdots L$, in which case the width of the eye opening will be twice the minimum z_i value, or $2 \cdot \min\{z_1, \cdots, z_L\}$.

The function $f(\boldsymbol{c})$ of (10) will play an important role in our analysis. Observe that $f(\boldsymbol{c})$ is an element of the signal cone, since it is a weighted sum of signal vectors with positive weights. The weight $e^{-z_i^2/2}$ is largest when z_i is smallest; *i.e.*, when the eye diagram is most closed. Thus, $f(\boldsymbol{c})$ is approximately a linear combination of the few $\boldsymbol{s}_i$ vectors for which the eye diagram is most closed. For example, if one particular $\boldsymbol{s}_i$ vector closes the eye significantly more than any other signal vector, then $f(\boldsymbol{c})$ will be approximately proportional to that $\boldsymbol{s}_i$.

The importance of $f(\boldsymbol{c})$ stems from the following fixed-point equation.

Lemma 1: An equalizer $\boldsymbol{c}$ that minimizes the BER of an equalizable channel must satisfy

$$\boldsymbol{c} = af(\boldsymbol{c}), \qquad \text{for some } a > 0. \tag{12}$$

Proof: The minimum-BER equalizer must satisfy $\nabla_{\boldsymbol{c}}\mathrm{BER} = \boldsymbol{0}$. Observe that the vector $(\boldsymbol{I} - \boldsymbol{c}\boldsymbol{c}^T/\|\boldsymbol{c}\|^2)f(\boldsymbol{c})$ in (9) is the difference between $f(\boldsymbol{c})$ and its projection onto the subspace spanned by $\boldsymbol{c}$. Thus, the BER gradient will be zero only if this projection error is zero. Recall that $f(\boldsymbol{c})$ is an element of the signal cone, and thus cannot be zero. Thus, the BER gradient will be zero only if $f(\boldsymbol{c})$ and $\boldsymbol{c}$ are colinear: $\boldsymbol{c} = af(\boldsymbol{c})$ for some nonzero constant a. That a must be positive can be proven by contradiction: suppose $\boldsymbol{c} = af(\boldsymbol{c})$ minimizes BER with $a < 0$. Then $\boldsymbol{c}$ is outside the signal cone generated by $\{\boldsymbol{s}_i\}$. Let P be any hyperplane passing through the origin that separates $\boldsymbol{c}$ from the signal cone, and let $\tilde{\boldsymbol{c}}$ denote the reflection of $\boldsymbol{c}$ about P. It is easy to show that compared to $\boldsymbol{c}$, $\tilde{\boldsymbol{c}}$ has larger normalized inner products with all $\boldsymbol{s}_i$ vectors. From (8), it follows that the BER for $\tilde{\boldsymbol{c}}$ is smaller than the BER for $\boldsymbol{c}$, which contradicts the assumption that $\boldsymbol{c}$ minimizes BER. **Q.E.D.**

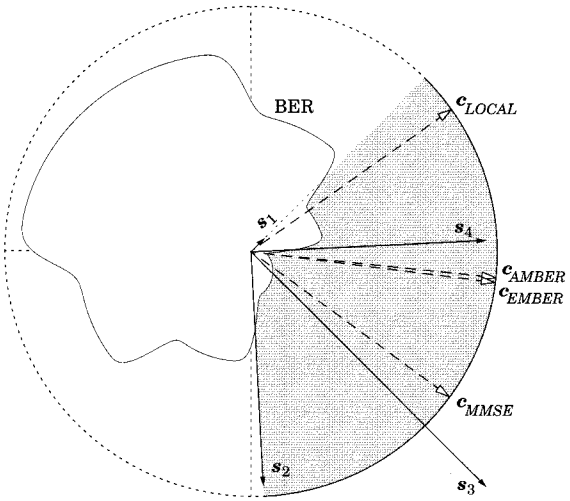


Fig. 2. A polar plot of BER versus θ for $H(z) = -0.9 + z^{-1}$, $N = 2$, $D = 1$. Superimposed are the signals vectors (scaled by a factor of 0.5), and four equalizer vectors (dashed lines).

The fixed-point relationship of (12) is a necessary but not sufficient condition for minimum-BER equalization, as illustrated in the following example.

Example 1: Consider the simple two-tap channel $H(z) = -0.9 + z^{-1}$ with $E_b/N_0 = \Sigma_k h_k^2/(2\sigma^2) = 17$ dB and a two-tap equalizer ($N = 2$) with delay $D = 1$. (This delay yields the MMSE equalizer with the smallest BER.) In Fig. 2, we present a polar plot of BER versus θ for the equalizer $\boldsymbol{c} = [\cos\theta, \sin\theta]^T$. Superimposed on this plot are the $L = 4$ signal vectors $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_4\}$, depicted by solid lines. Also superimposed are three unit-length equalizer vectors (depicted by dashed lines): the exact minimum-BER (EMBER) equalizer with $\theta = -7.01°$; the MMSE equalizer with $\theta = -36.21°$; and a local-minimum equalizer ($\boldsymbol{c}_{\mathrm{LOCAL}}$), satisfying $\nabla_{\boldsymbol{c}}\mathrm{BER} = \boldsymbol{0}$, with $\theta = 35.63°$. (A fourth equalizer with $\theta = -5.84°$ is also depicted for future reference—it is the AMBER equalizer of Lemma 3 in Section IV-A.) The shaded region denotes the signal cone. Although both $\boldsymbol{c}_{\mathrm{EMBER}}$ and $\boldsymbol{c}_{\mathrm{LOCAL}}$ satisfy (12) with $a > 0$, the equalizer $\boldsymbol{c}_{\mathrm{LOCAL}}$ does not minimize BER, and it does not open the eye diagram.

The previous example illustrates that the fixed-point condition of Lemma 1 is not sufficient to minimize BER; both $\boldsymbol{c}_{\mathrm{EMBER}}$ and $\boldsymbol{c}_{\mathrm{LOCAL}}$ satisfy $\boldsymbol{c} = af(\boldsymbol{c})$ with $a > 0$, but only $\boldsymbol{c}_{\mathrm{EMBER}}$ minimizes BER. One general method for finding an EMBER equalizer is to find all solutions to the fixed-point

relationship $\boldsymbol{c} = af(\boldsymbol{c})$ with $a > 0$, and choose the solution that yields the smallest BER. Fortunately, this brute-force method can be avoided in certain cases by taking advantage of the following sufficiency test.

Theorem 1: If $\boldsymbol{c} = af(\boldsymbol{c})$ and $\mathrm{BER} \leq 1/(2L)$, then $\boldsymbol{c}$ minimizes BER.

Proof: See Appendix I.

This is a sufficient but not necessary condition for minimizing BER, because even the minimum-BER may exceed $1/2L$ when the E_b/N_0 is low. Note that the condition $\mathrm{BER} \leq 1/(2L)$ implies that the equalizer opens the eye diagram.

B. Deterministic EMBER Algorithm

Example 1 in the previous section illustrates that the BER cost function may not be convex. Nevertheless, a gradient algorithm may still be used to search for a local minimum. In particular, a normalized gradient algorithm based on (9) yields

$$\begin{aligned}\boldsymbol{c}_{k+1} &= \boldsymbol{c}_k - \mu_1\sqrt{2\pi}\sigma\|\boldsymbol{c}_k\|\nabla_{\boldsymbol{c}_k}\mathrm{BER} \\ &= \boldsymbol{c}_k + \mu_1\left(\boldsymbol{I} - \frac{\boldsymbol{c}_k\boldsymbol{c}_k^T}{\|\boldsymbol{c}_k\|^2}\right)f(\boldsymbol{c}_k) \qquad (13)\\ &= \left(1 - \mu_1\boldsymbol{c}_k^T f(\boldsymbol{c}_k)/\|\boldsymbol{c}_k\|^2\right)(\boldsymbol{c}_k + \mu f(\boldsymbol{c}_k)) \qquad (14)\end{aligned}$$

where we have introduced $\mu = \mu_1/(1 - \mu_1\boldsymbol{c}_k^T f(\boldsymbol{c}_k)/\|\boldsymbol{c}_k\|^2)$. Recall that the norm of $\boldsymbol{c}$ has no impact on BER, and observe that the factor $(1 - \mu_1\boldsymbol{c}_k^T f(\boldsymbol{c}_k)/\|\boldsymbol{c}_k\|^2)$ in (14) represents an adjustment of the norm of $\boldsymbol{c}_{k+1}$. Eliminating this factor leads to the following recursion, which we refer to as the *deterministic EMBER algorithm*:

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k + \mu f(\boldsymbol{c}_k). \qquad (15)$$

The update equation (15) can be viewed as an iterative system designed to recover the solution to the fixed-point equation of (12). The transformation from (13) to (15) affects the convergence rate, the steady-state norm $\|\boldsymbol{c}_\infty\|$, and the steady-state direction $\boldsymbol{c}_\infty/\|\boldsymbol{c}_\infty\|$, so it is no longer appropriate to call (15) a gradient search algorithm. Nevertheless, convergence to a local extremum can be guaranteed:

Lemma 2: Given an equalizable channel, the deterministic EMBER algorithm of (15) converges to a local extremum solution satisfying $\boldsymbol{c} = af(\boldsymbol{c})$ with $a > 0$.

Proof: See Appendix II.

Taken together, Theorem 1 and Lemma 2 suggest the following strategy for finding the exact minimum-BER equalizer. First, iterate the deterministic EMBER algorithm of (15) until it converges. If the resulting $\mathrm{BER} \leq 1/(2L)$, stop. Otherwise, initialize the deterministic EMBER algorithm somewhere else and repeat the process. This is an effective strategy when the initial condition of the EMBER algorithm is chosen carefully (for example, chosen to be the MMSE equalizer) and when the E_b/N_0 is sufficiently large that $\mathrm{BER} \leq 1/(2L)$ is possible.

IV. Adaptive Minimum BER Equalization

Although the deterministic EMBER algorithm (15) of the previous section is useful for finding the minimum-BER equalizer of known channels, it is poorly suited for adaptive equalization. We now propose a modified version of the algorithm that leads to a stochastic update equation with extremely low complexity.

A. The Deterministic AMBER Algorithm

The error function $Q(z)$ is upper bounded and approximated by $(e^{-z^2/2})/(z\sqrt{2\pi})$ [6], so that $f(\boldsymbol{c})$ of (10) can be approximated by

$$\begin{aligned}f(\boldsymbol{c}) &\approx \frac{\sqrt{2\pi}}{L}\sum_{i=1}^{L} z_i Q(z_i)\boldsymbol{s}_i \qquad (16)\\ &\approx z_{\min}\frac{\sqrt{2\pi}}{L}\sum_{i=1}^{L} Q(z_i)\boldsymbol{s}_i \qquad (17)\\ &= z_{\min}\sqrt{2\pi}\,g(\boldsymbol{c}) \qquad (18)\end{aligned}$$

where $z_i = \boldsymbol{c}^T\boldsymbol{s}_i/(\|\boldsymbol{c}\|\sigma)$ as in (11), $z_{\min} = \min\{z_i\}$, and where we have introduced the vector function $\mathrm{g}: \mathbb{R}^N \to \mathbb{R}^N$

$$\mathrm{g}(\boldsymbol{c}) = \frac{1}{L}\sum_{i=1}^{L} Q(z_i)\boldsymbol{s}_i = E\left[Q\left(\frac{\boldsymbol{c}^T\boldsymbol{s}}{\|\boldsymbol{c}\|\sigma}\right)\boldsymbol{s}\right]. \qquad (19)$$

Comparing (10) and (19), we see that the vector function $\mathrm{g}(\boldsymbol{c})$ has the same form as $f(\boldsymbol{c})$, but with $Q(z)$ replacing $e^{-z^2/2}$. The approximation in (17) is valid because only the terms in (17) for which $z_i \approx z_{\min}$ are relevant; the other terms have negligible impact. Using $\mathrm{g}(\boldsymbol{c})$ to approximate $f(\boldsymbol{c})$ in (15) leads to the following *approximate* minimum-BER update equation, which we refer to as the *deterministic AMBER algorithm*:

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k + \mu\mathrm{g}(\boldsymbol{c}_k). \qquad (20)$$

Observe the similarity between (20) and the deterministic EMBER algorithm of (15). Because of the approximations in (16) and (17), the deterministic AMBER algorithm no longer minimizes BER exactly. However, the approximations are well-justified, and the numerical results of Section V verify that (20) converges to a solution that very nearly minimizes BER.

Although there can exist multiple solutions to the EMBER fixed-point equation $\boldsymbol{c} = af(\boldsymbol{c})$ with $a > 0$, there exists only one fixed-point solution when $f(\boldsymbol{c})$ is replaced by $\mathrm{g}(\boldsymbol{c})$.

Lemma 3: For any equalizable channel there is a unique unit-length vector, namely $\boldsymbol{c}_{\mathrm{AMBER}}$, satisfying the following fixed-point relationship:

$$\boldsymbol{c} = a\mathrm{g}(\boldsymbol{c}), \qquad \text{for some } a > 0. \qquad (21)$$

Proof: See Appendix III.

Moreover, the deterministic AMBER algorithm converges to $\boldsymbol{c}_{\mathrm{AMBER}}$:

Theorem 2: For any equalizable channel, the deterministic AMBER algorithm of (20) is guaranteed to converge to the direction of $\boldsymbol{c}_{\mathrm{AMBER}}$ satisfying (21), regardless of its initial condition, and for any $\mu > 0$.

Proof: See Appendix IV.

Observe the similarity between (12) and (21). At steady state, as shown in (19) and (21), $\boldsymbol{c}_{\mathrm{AMBER}}$ is a weighted combination of $\boldsymbol{s}_i$ vectors, with weights proportional to the conditional error probability. Since these conditional error probabilities are dominated by the few $\boldsymbol{s}_i$ vectors that close the eye the most, $\boldsymbol{c}_{\mathrm{AMBER}}$

is approximately a linear combination of the few $\boldsymbol{s}_i$ vectors for which the eye diagram is most closed.

B. The Stochastic AMBER Algorithm

The deterministic AMBER algorithm of (20) has a second advantage over the deterministic EMBER algorithm of (15): besides being globally convergent to a unique fixed-point solution, we will now show that there also exists a low-complexity stochastic approximation of the deterministic AMBER algorithm.

We begin by defining an error indicator function I_k that is zero or one, depending on whether a decision error occurs at time k

$$I_k = \begin{cases} 0, & \text{if sgn } \{\boldsymbol{c}^T\boldsymbol{r}_k\} = x_{k-D} \\ 1, & \text{if sgn } \{\boldsymbol{c}^T\boldsymbol{r}_k\} \neq x_{k-D}. \end{cases} \tag{22}$$

In other words, $I_k = (1 - \text{sgn}[x_{k-D}\boldsymbol{c}^T\boldsymbol{r}_k])/2$. It follows that $E[I_k] = \text{BER}$. As exploited in (6), the conditional expectation of I_k given the signal vector $\boldsymbol{s} = \boldsymbol{H}\boldsymbol{x}_k x_{k-D}$ is

$$E[I_k|\boldsymbol{s}] = Q\left(\frac{\boldsymbol{c}^T\boldsymbol{s}}{||\boldsymbol{c}||\sigma}\right). \tag{23}$$

Therefore, the AMBER function of (19) is a simple function of the error indicator function

$$\text{g}(\boldsymbol{c}) = E[E[I_k|\boldsymbol{s}]\boldsymbol{s}] = E[I_k\boldsymbol{s}]. \tag{24}$$

We can use this indicator function to simplify the deterministic AMBER algorithm of (20), through the following set of straightforward equalities:

$$\begin{aligned} \boldsymbol{c}_{k+1} &= \boldsymbol{c}_k + \mu E[I_k\boldsymbol{s}] \\ &= \boldsymbol{c}_k + \mu E[I_k x_{k-D}\boldsymbol{H}\boldsymbol{x}_k] \\ &= \boldsymbol{c}_k + \mu E[I_k x_{k-D}(\boldsymbol{r}_k - \boldsymbol{n}_k)] \\ &\approx \boldsymbol{c}_k + \mu E[I_k x_{k-D}\boldsymbol{r}_k]. \end{aligned} \tag{25}$$

The last approximation is valid at high SNR, and is best justified by the good performance of the resulting algorithm, as demonstrated in the numerical results to follow.

A simple and unbiased stochastic gradient update algorithm can be formed by removing the expectation in (25)

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k + \mu I_k x_{k-D}\boldsymbol{r}_k. \tag{26}$$

We refer to (26) as the *adaptive minimum-BER (AMBER) algorithm*. If the last approximation of (25) is valid, then it can be argued that, for sufficiently small step size, this algorithm converges in the mean-square sense to a solution satisfying the fixed-point relationship of (21).

A closer look at (26) leads to some insightful geometric interpretations of the AMBER algorithm.

- The indicator function I_k ensures that the equalizer is updated only when a decision error is made. The AMBER algorithm is thus more passive in nature than more aggressive adaptive algorithms—such as the LMS algorithm—that update with every iteration. The stop-and-go property of AMBER bears some resemblance to the blind equalization algorithm of [7], but with an important distinction: the algorithm of [7] updates only when the equalizer output is reliable, whereas AMBER updates only when the output is *not* reliable.
- The update term $x_{k-D}\boldsymbol{r}_k$ in (26) is a noisy estimate of the signal vector $\boldsymbol{s}$

$$x_{k-D}\boldsymbol{r}_k = x_{k-D}(\boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{n}_k) = \boldsymbol{s} + x_{k-D}\boldsymbol{n}_k \approx \boldsymbol{s}. \tag{27}$$

  Hence, when a decision error is made, $\boldsymbol{c}$ takes a small step in the general direction of the signal vector $\boldsymbol{s}$ that caused the error.
- Averaged over many iterations, $\boldsymbol{c}$ will move toward each $\boldsymbol{s}_i$ with a frequency proportional to the probability $Q(z_i)$ that $\boldsymbol{s}_i$ causes an error, where $z_i = \boldsymbol{c}^T\boldsymbol{s}_i/(||\boldsymbol{c}||\boldsymbol{\sigma})$. Therefore, at steady state we expect $\boldsymbol{c} \propto \sum_i Q(z_i)\,\boldsymbol{s}_i$, which is precisely the AMBER fixed-point relationship of (21), and which closely approximates the minimum-BER fixed-point relationship of (12), namely $\boldsymbol{c} \propto \sum_i e^{-z_i^2/2}\,\boldsymbol{s}_i$.

The last observation is especially illuminating because it provides a direct connection between the AMBER algorithm (26) and the relationship $\boldsymbol{c} \propto \sum_i e^{-z_i^2/2}\,\boldsymbol{s}_i$ of Lemma 1 that the minimum-BER equalizer is known to satisfy. Thus, the gist of the AMBER algorithm can be partially justified on intuitive grounds from Lemma 1 alone.

We can gain additional insight into the AMBER algorithm by comparing it to two other well-known adaptive algorithms: the LMS algorithm, which implements the MMSE equalizer, and the sign-LMS algorithm [8], a lower-complexity version of the LMS algorithm which approximates the MMSE equalizer. All three algorithms can be expressed in a similar form

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \mu e_k\boldsymbol{r}_k \qquad (\text{LMS}) \tag{28}$$

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \mu\text{sgn}\{e_k\}\boldsymbol{r}_k \qquad (\text{sign} - \text{LMS}) \tag{29}$$

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \mu I_k\text{sgn}\{e_k\}\boldsymbol{r}_k \qquad (\text{AMBER}) \tag{30}$$

where $e_k = \boldsymbol{c}^T\boldsymbol{r}_k - x_{k-D}$ is the error signal of the MMSE detector, and where we have made use of the identity $I_k x_{k-D} = -I_k\text{sgn}\{e_k\}$ to transform (26) to (30). Observe the remarkable similarity between the AMBER and sign-LMS algorithms. Simply stated, the AMBER algorithm can be viewed as the sign-LMS algorithm *modified to update only when a detection error is made*. The sign-LMS was motivated by its low complexity compared to the LMS algorithm, despite its poorer performance. The simple modification for AMBER, on the other hand, provides a dramatic improvement in performance, while retaining a complexity that is comparable to that of the sign-LMS algorithm.

C. Maintaining a Fast Convergence Speed

Because the AMBER algorithm updates only when a detection error is made, the speed of convergence can decrease with time as the BER itself decreases. One method for maintaining a fast convergence speed at low BER is to modify the AMBER algorithm so that it not only updates when an error is made, but also when an error is almost made. Specifically, we can modify the error indicator function of (22) by introducing a nonnegative threshold $\tau \geq 0$ as follows:

$$I_k = \frac{1}{2}(1 - \text{sgn}[x_{k-D}y_k - \tau]). \tag{31}$$

In other words, the modified indicator function is $I_k = 1$ if $x_{k-D}y_k \leq \tau$ and $I_k = 0$ otherwise. This indicator function reverts back to the original (22) when the threshold τ is zero. Although the threshold modification can increase the speed of convergence, it also changes the steady-state performance of the equalizer. Additional methods for speeding convergence are described in [9].

The original AMBER algorithm requires a training sequence, because knowledge of x_{k-D} is required to evaluate the error indicator function I_k. Besides speeding convergence, the threshold modification of (31) also allows the AMBER algorithm to be operated in a decision-directed manner, using $\hat{x}_{k-D}$ in place of x_{k-D} in (26) and (31). The modified algorithm in decision-directed mode will then update only when an error is almost made.

D. The AMBER Algorithm for QAM

Although we have derived the AMBER algorithm assuming binary signaling, it can be generalized to 4-QAM by allowing the signals in (1) to be complex valued. Using superscripts R and I to denote real and imaginary parts, respectively, of a complex number, the BER (not symbol-error rate) of a 4-QAM system with $x_k \in \{\pm 1 \pm j\}$ and Gray coding is [9]

$$\begin{aligned}\mathrm{BER} &= \frac{1}{2}\Pr[\hat{x}^R_{k-D} \neq x^R_{k-D}] + \frac{1}{2}\Pr[\hat{x}^I_{k-D} \neq x^I_{k-D}] \\ &= \frac{1}{2}E\left[Q\left(\frac{x^R_{k-D}(\boldsymbol{c}^T\boldsymbol{H}\boldsymbol{x}_k)^R}{\|\boldsymbol{c}\|\sigma}\right)\right] \\ &\quad + \frac{1}{2}E\left[Q\left(\frac{x^I_{k-D}(\boldsymbol{c}^T\boldsymbol{H}\boldsymbol{x}_k)^I}{\|\boldsymbol{c}\|\sigma}\right)\right] \qquad (32)\end{aligned}$$

where n^R_k and n^I_k are assumed to be white, Gaussian, and independent with power spectral density σ^2. A straightforward extension of the AMBER derivation leads to the following 4-QAM extension [9]

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k + \mu\tilde{I}_k\boldsymbol{r}^*_k \qquad (33)$$

where $\boldsymbol{r}^*$ denotes the conjugate of $\boldsymbol{r}$, and where

$$\tilde{I}_k = x^R_{k-D}F(x^R_{k-D}y^R_k) + jx^I_{k-D}F(x^I_{k-D}y^I_k)$$

with $F(t) = (1 - \mathrm{sgn}(t-\tau))/2$, as in (31).

V. Numerical Results

In this section we present simulation results for two systems that demonstrate the effectiveness of the AMBER algorithm.

A. Channel A: Binary Signaling

Here we consider a binary alphabet $x_k \in \{\pm 1\}$ for a channel with transfer function $H(z) = 1.2 + 1.1z^{-1} - 0.2z^{-2}$. In Fig. 3 we plot BER versus $E_b/N_0 = \Sigma_k h_k^2/(2\sigma^2)$ using (8), considering both MMSE and AMBER equalizers of length three and five. The MMSE and minimum-BER equalizers were calculated deterministically using (4) and (15), respectively, whereas the AMBER equalizer was calculated after 2×10^6 iterations of the stochastic AMBER algorithm of (26) with the following parameters: the step size and threshold were initialized to $\mu = 0.02$ and $\tau = 0.8$, respectively, and both decreased exponentially with a half-life of 10^6 iterations. The delay parameter was chosen to minimize the BER of the MMSE equalizer; $D = 2$ for the 3-tap equalizer and $D = 4$ for the 5-tap equalizer.

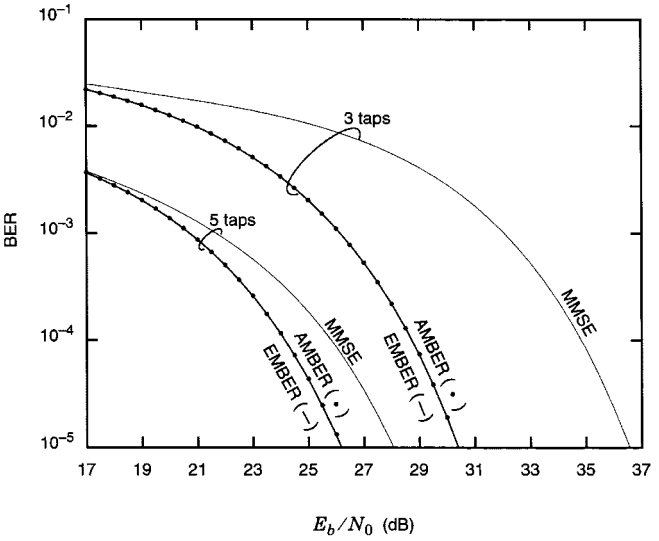


Fig. 3. Steady-state BER performance comparison for channel A.

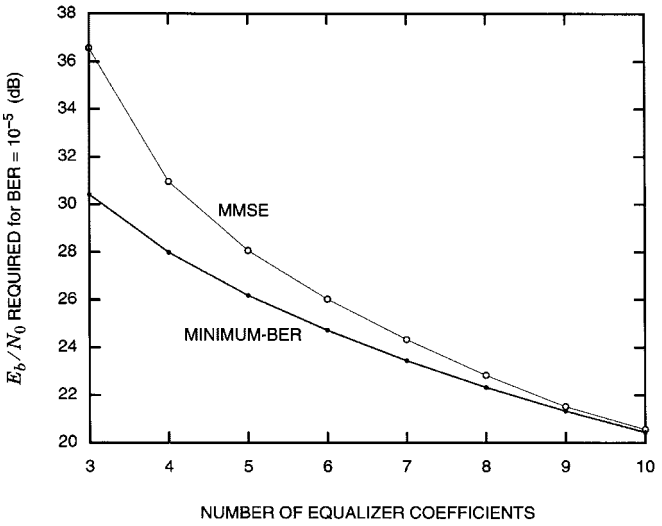


Fig. 4. E_b/N_0 requirement vs. equalizer length for channel A.

The figure shows that the 3-tap AMBER equalizer outperforms the 3-tap MMSE equalizer by more than 6.5 dB. The 5-tap AMBER equalizer outperforms the 5-tap MMSE equalizer by nearly 2 dB. Furthermore, the figure shows that there is no observable difference between the performance of the AMBER equalizer (marked by circles) and the performance of the exact minimum-BER (EMBER) equalizer (marked by the solid curves). The closeness with which the AMBER algorithm matches the minimum-BER equalizer indicates that the two approximations made in deriving the AMBER algorithm have a negligible impact for this example.

In Fig. 3 we observe that the improvement of EMBER over MMSE drops from 6.5 to 2 dB as the equalizer length increases from three to five. This trend extends to longer equalizers as well, as demonstrated in Fig. 4, where we plot the E_b/N_0 required to achieve $\mathrm{BER} = 10^{-5}$ versus equalizer length for both

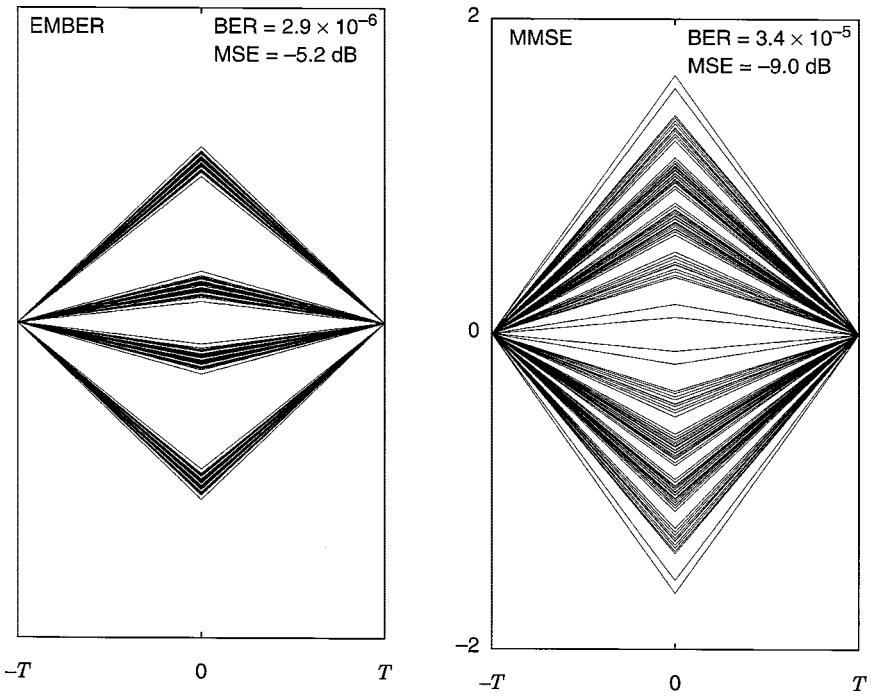


Fig. 5. Equalized noiseless eye patterns for (a) EMBER and (b) MMSE for channel A with 5 taps and $E_b/N_0 = 27$ dB.

EMBER and MMSE. We see that, at least for this particular channel, MMSE performance approaches minimum-BER performance as the length of the equalizers increases.

In Fig. 5, still for channel A, we present "artificial" noiseless eye patterns for the EMBER and MMSE equalizers, assuming five equalizer taps and $E_b/N_0 = 27$ dB. These patterns were obtained by interpolating all possible noiseless equalizer outputs with a triangular pulse shape. Both equalizers were normalized to have identical norm (and thus identical noise enhancement). The striking difference between the MMSE and EMBER eye diagrams results from MMSE's effort to force all possible equalizer outputs to $\{\pm 1\}$, despite the benefits of sparing the outputs with large noise immunity.

We compare the convergence rate of the AMBER and LMS 3-tap equalizers for channel A in Fig. 6, where we plot BER versus time assuming $E_b/N_0 = 27$ dB. The AMBER step size was $\mu = 0.2$ and the threshold was $\tau = 0.5$, whereas the LMS step size was 0.01. To test their ability to escape from a closed-eye condition, both equalizers were initialized to the negative of the MMSE equalizer, $\boldsymbol{c}_0 = -\boldsymbol{c}_{\mathrm{MMSE}}$. Observe that AMBER offers both faster initial convergence (outperforming the closed-form MMSE equalizer in fewer than 50 iterations) and better steady-state performance than LMS.

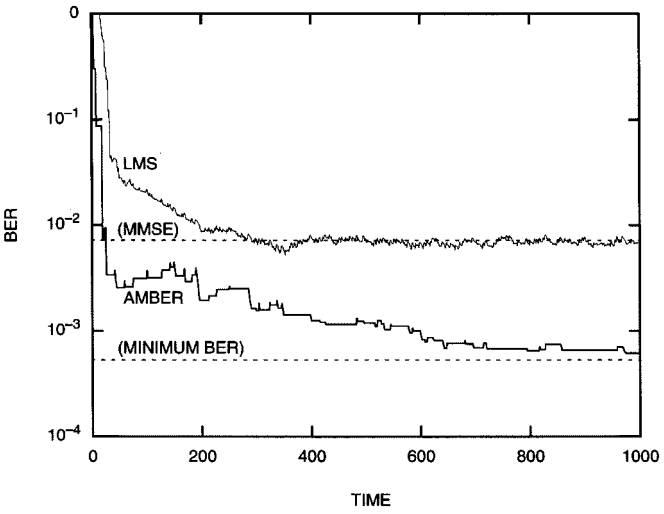


Fig. 6. BER versus time for the LMS and AMBER algorithms for channel A.

B. Channel B: 4-QAM

Here we consider the 4-QAM alphabet $\{\pm 1 \pm j\}$ for the channel $H(z) = (0.7 - 0.2j) + (0.4 - 0.5j)z^{-1} + (-0.2 + 0.3j)z^{-2}$, and $E_b/N_0 = \Sigma_k |h_k|^2/(2\sigma^2)$. In Fig. 7 we compare the BER performance of the MMSE equalizer to the AMBER equalizer, which was calculated after 2×10^6 iterations of the stochastic algorithm of (33) with the following parameters: the step size and threshold were initialized to $\mu = 0.02$ and $\tau = 0.8$, respectively, and both decreased exponentially with a half-life of 10^6 iterations. The delay parameter was chosen to minimize the BER of the MMSE equalizer; $D = 3$ for the 4-tap equalizer and $D = 4$ for the 5-tap equalizer. We see that the 4-tap AMBER equalizer outperforms the MMSE equalizer by more than 16 dB. With five taps, the gain drops to slightly more than 2 dB.

In Fig. 8 we present the noiseless constellation diagrams for the 4-tap AMBER and MMSE equalizers. Observe the interesting structure of the AMBER constellation clouds; although they result in a higher MSE than the MMSE clouds (which appear roughly Gaussian), they result in a lower BER, in part because the edges of the AMBER clouds are further apart.

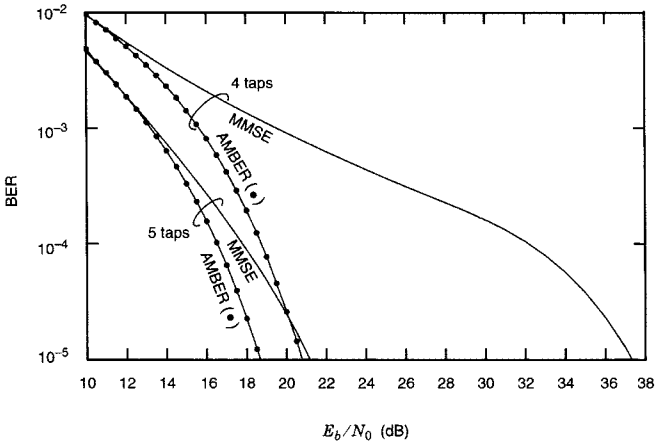
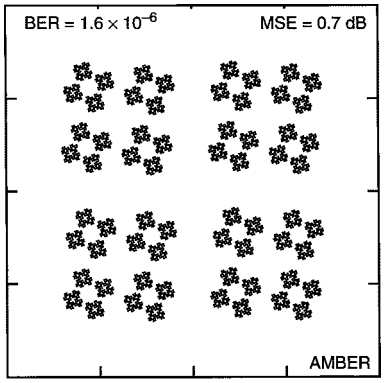
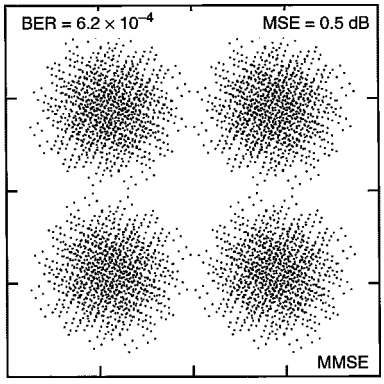


Fig. 7. Steady-state BER performance comparison for channel B.



(a)



(b)

Fig. 8. Noiseless equalized constellations for (a) AMBER and (b) MMSE; on channel B with 4 taps and $E_b/N_0 = 22$ dB.

VI. Conclusions

The MMSE linear equalizer is widely believed to provide good BER performance. Indeed, our numerical results support the conjecture that, when the number of equalizer coefficients is sufficiently large, the BER performance of the MMSE equalizer approaches that of the minimum-BER equalizer. In contrast, however, when the number of equalizer coefficients is insufficient to accurately realize the channel inverse, we have shown that the MMSE equalizer can be far from optimal.

The minimum-BER equalizer is related to the set of signal vectors $\{\boldsymbol{s}_i\}$ by the simple fixed-point relationship $\boldsymbol{c} \propto \sum_i e^{-z_i^2/2}\boldsymbol{s}_i$, where $z_i = \boldsymbol{c}^T\boldsymbol{s}_i/(\|\boldsymbol{c}\|\sigma)$. The stochastic AMBER algorithm closely approximates this relationship by using $Q(z_i)$ to approximate $e^{-z_i^2/2}$, and by neglecting the noise term $I_k x_{k-D}\boldsymbol{n}_k$ in (25). Despite these two approximations, we observed no appreciable difference in our examples between the steady-state BER performance of the stochastic AMBER algorithm and the true minimum-BER equalizer. (Whether or not there exists a channel for which AMBER does not closely approximate the minimum-BER equalizer is an open question.) Furthermore, we conjecture without proof that the stochastic AMBER algorithm will be globally convergent for a sufficiently small step size and an equalizable channel, regardless of its initial condition; this conjecture is justified in part by Lemma 3, by the negligible impact of the noise term $I_k x_{k-D}\boldsymbol{n}_k$ in (25) that separates the stochastic AMBER algorithm from the deterministic algorithm of (20), and by extensive simulation results [9]. The complexity of the AMBER algorithm is no greater than that of the LMS algorithm.

The AMBER algorithm has recently been extended to decision-feedback equalization [9], to multiple-level pulse-amplitude and quadrature-amplitude modulation schemes such as 64-QAM [10], and to multiuser detection for synchronous code-division-multiple access systems [11]. Areas for future work include its application to fractionally spaced equalization, nonlinear channels, and asynchronous multiuser systems.

Appendix I
Proof of Theorem 1

Let $\mathbb{E} \subseteq \mathbb{R}^N$ denote the set of all eye-opening equalizers having unit length, *i.e.*, $\mathbb{E}$ is the set of unit-length vectors having positive inner product with all $\boldsymbol{s}^i$ signal vectors. This set is not empty when the channel is equalizable, by definition. We can write $\mathbb{E} = \cap_{i=1}^{L}\mathbb{E}_i$, where $\mathbb{E}_i = \{\boldsymbol{e}\colon \boldsymbol{e}^T\boldsymbol{s}_i > 0, \|\boldsymbol{e}\| = 1\}$. Observe from (8) that the condition $\mathrm{BER} \leq 1/2L$ implies that $\boldsymbol{c} \in \mathbb{E}$. We must show that if $\boldsymbol{c} \in \mathbb{E}$ and $\boldsymbol{c} = af(\boldsymbol{c})$ then $\boldsymbol{c}$ globally minimizes BER. First, observe from (8) that any equalizer not in $\mathbb{E}$ will have a BER of $1/2L$ or greater, whereas at least one equalizer within $\mathbb{E}$ (namely $\boldsymbol{c}$) has $\mathrm{BER} \leq 1/2L$, so that the global minimum must be in the eye-opening region $\mathbb{E}$. However, as shown below, BER has only one local extremum over $\mathbb{E}$; thus, the local extremum $\boldsymbol{c} = af(\boldsymbol{c})$ must be the global minimum.

That the BER has only one local minimum over the eye-opening region $\mathbb{E}$ can be proven by contradiction: suppose both $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are distinct local minima in $\mathbb{E}$, so that $\boldsymbol{c}_1 = a_1 f(\boldsymbol{c}_1) \neq \boldsymbol{c}_2 = a_2 f(\boldsymbol{c}_2)$, where $a_1 > 0$, $a_2 > 0$, and $\|\boldsymbol{c}_1\| = \|\boldsymbol{c}_2\| = 1$. Let P denote the plane containing the origin and the perpendicular bisector of $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$, as shown in Fig. 9 for a three-tap equalizer. This plane bisects the signal cone $S = \{\Sigma_i a_i \boldsymbol{s}_i\colon a_i > 0\}$ into the disjoint subcones S_1 and S_2 satisfying $S_1 \cap S_2 = \varnothing$ and $S = S_1 \cup S_2$, where S_1 is the intersection of S with the set of vectors on the $\boldsymbol{c}_1$ side of P, excluding P, and S_2 is the intersection of S with the set of vectors on the $\boldsymbol{c}_2$ side of P, including P. Observe that $\boldsymbol{c}_1 \in S_1$ and $\boldsymbol{c}_2 \in S_2$.

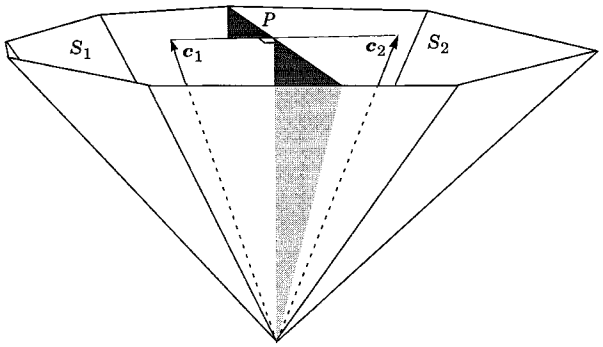


Fig. 9. Dividing the signal cone into two subcones S_1 and S_2 with the plane P.

From (10), $f(\boldsymbol{c}_1)$ and $f(\boldsymbol{c}_2)$ can be decomposed as

$$f(\boldsymbol{c}_m) = \frac{1}{L}\sum_{\boldsymbol{s}_i \in S_1} e^{-z_{m,i}^2/2}\boldsymbol{s}_i + \frac{1}{L}\sum_{s_j \in S_2} e^{-z_{m,j}^2/2}\boldsymbol{s}_j, \quad \text{for } m = 1 \text{ or } 2 \tag{34}$$

where $z_{m,i} = \boldsymbol{c}_m^T\boldsymbol{s}_i/\sigma$ is a normalized inner product. Observe that $z_{1,i} > z_{2,i}$ for signal vectors $\boldsymbol{s}_i \in S_1$, whereas $z_{1,j} \le z_{2,j}$ for signal vectors $\boldsymbol{s}_j \in S_2$. Moreover, because $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are in the eye-opening region and thus the inner products $\{z_{m,i}\}$ are all positive, $e^{-z_{1,i}^2/2} < e^{-z_{2,i}^2/2}$ for signal vectors $\boldsymbol{s}_i \in S_1$, whereas $e^{-z_{1,j}^2/2} \ge e^{-z_{2,j}^2/2}$ for signal vectors $\boldsymbol{s}_j \in S_2$. In other words, $f(\boldsymbol{c}_2)$ can be decomposed into the same two summations (34) as can $f(\boldsymbol{c}_1)$, except with larger weights for signal vectors in S_1 and smaller weights for signal vectors in S_2. Since $f(\boldsymbol{c}_1) = \boldsymbol{c}_1/a_1 \in S_1$, it must also be that $f(\boldsymbol{c}_2) \in S_1$, which contradicts $f(\boldsymbol{c}_2) = \boldsymbol{c}_2/a_2 \in S_2$. **Q.E.D.**

Appendix II
Proof of Lemma 2

Since the $\boldsymbol{s}_i$ vectors generate a signal cone, we can find a hyperplane P, containing the origin, such that all $\boldsymbol{s}_i$ vectors are strictly on one side of P. Every $\boldsymbol{s}_i$ makes an angle $\theta_i \in [0, 90°)$ with the normal to P and consists of two components: one (with norm $||\boldsymbol{s}_i|| \sin\theta_i$) parallel to P and the other (with norm $||\boldsymbol{s}_i|| \cos\theta_i$) perpendicular to P. At each update, the correction vector $\mu f(\boldsymbol{c}_k)$ is strictly inside the cone and its norm is lower bounded by $\mu \exp(-||\boldsymbol{s}||_{\max}^2/2\sigma^2)||\boldsymbol{s}||_{\min}\cos\theta_{\max}$, where $||\boldsymbol{s}||_{\min} = \min_i\{||\boldsymbol{s}_i||\}$, $||\boldsymbol{s}||_{\max} = \max_i\{||\boldsymbol{s}_i||\}$, and $\theta_{\max} = \max_i\{\theta_i\}$. At iteration $M+1$, the sum of the past M correction vectors is a vector strictly inside the cone and has a norm of at least $M\mu \exp(-||\boldsymbol{s}||_{\max}^2/2\sigma^2)||\boldsymbol{s}||_{\min}\cos\theta_{\max}$. We conclude that, for any initial $\boldsymbol{c}_0$ with a finite norm, there exists a finite M such that $\boldsymbol{c}_{M+1}$ is strictly inside the cone. In addition, we conclude that the equalizer norm $||\boldsymbol{c}_k||$ grows without bound as k increases.

Showing that $\boldsymbol{c}_k$ converges to the direction of an extremum solution satisfying $\boldsymbol{c} = af(\boldsymbol{c})$ for $a > 0$ is equivalent to showing that the angle between $\boldsymbol{c}_k$ and $f(\boldsymbol{c}_k)$ approaches zero. First we observe that $\boldsymbol{c}_k/||\boldsymbol{c}_k||$ must converge to some fixed vector $\tilde{\boldsymbol{c}}_\infty$, since $||\boldsymbol{c}_k||$ becomes arbitrarily large while the norm of the update, $||\mu f(\boldsymbol{c}_k)||$, is upper-bounded by $\mu||\boldsymbol{s}||_{\max}$. It follows that $f(\boldsymbol{c}_k)$ converges to $f(\tilde{\boldsymbol{c}}_\infty)$, and thus, for any $\varepsilon > 0$, there exists a finite $k(\varepsilon)$ such that for all $k > k(\varepsilon)$, $|\,||f(\boldsymbol{c}_k)|| - ||f(\tilde{\boldsymbol{c}}_\infty)||\,| \le ||f(\boldsymbol{c}_k) - f(\tilde{\boldsymbol{c}}_\infty)|| < \varepsilon$. Manipulating the inequalities yields that the angle between $f(\boldsymbol{c}_k)$ and $f(\tilde{\boldsymbol{c}}_\infty)$ is less than some $\theta(\varepsilon)$, where

$$\theta(\varepsilon) = \cos^{-1}\left[\frac{1 - \varepsilon/||f(\tilde{\boldsymbol{c}}_\infty)||}{1 + \varepsilon/||f(\tilde{\boldsymbol{c}}_\infty)||}\right]. \tag{35}$$

For any $M > 0$, $\sum_{j=0}^{M-1} f(\boldsymbol{c}_{k(\varepsilon)+j})$ is a vector strictly within the cone $W[f(\tilde{\boldsymbol{c}}_\infty); \theta(\varepsilon)]$ consisting of all vectors less than $\theta(\varepsilon)$ away from $f(\tilde{\boldsymbol{c}}_\infty)$. For a $\boldsymbol{c}_{k(\varepsilon)}$ with a finite norm, we can find a finite M such that $\boldsymbol{c}_{k(\varepsilon)+M} = \boldsymbol{c}_{k(\varepsilon)} + \mu\sum_{j=0}^{M-1} f(\boldsymbol{c}_{k(\varepsilon)+j})$ is strictly inside $W[f(\tilde{\boldsymbol{c}}_\infty); \theta(\varepsilon)]$. As ε approaches 0, $\theta(\varepsilon)$ approaches 0 and thus the angle between $\boldsymbol{c}_{k(\varepsilon)+M}$ and $f(\boldsymbol{c}_{k(\varepsilon)+M})$ approaches 0 as well. **Q.E.D.**

Appendix III
Proof of Lemma 3

This proof closely parallels the proof in Appendix I. Suppose $\boldsymbol{c}_1 = a_1 g(\boldsymbol{c}_1) \ne \boldsymbol{c}_2 = a_2 g(\boldsymbol{c}_2)$, where $a_1 > 0$ $a_2 > 0$, and $||\boldsymbol{c}_1|| = ||\boldsymbol{c}_2|| = 1$. Define the subcones S_1 and S_2 as in Appendix I. From (19), we can write

$$g(\tilde{\boldsymbol{c}}_m) = \frac{1}{L}\sum_{s_i \in S_1} Q(z_{m,i})\boldsymbol{s}_i + \frac{1}{L}\sum_{s_j \in S_2} Q(z_{m,j})\boldsymbol{s}_j, \quad \text{for } m = 1 \text{ or } 2 \tag{36}$$

where $z_{m,i} = \boldsymbol{c}_m^T\boldsymbol{s}_i/\sigma$. As in Appendix I, we have $z_{1,i} > z_{2,i}$ and thus $Q(z_{1,i}) < Q(z_{2,i})$ for signal vectors $\boldsymbol{s}_i \in S_1$, whereas $z_{1,j} \le z_{2,j}$ and thus $Q(z_{1,j}) \ge Q(z_{2,j})$ for signal vectors $\boldsymbol{s}_j \in S_2$. Since $g(\boldsymbol{c}_1) = \boldsymbol{c}_1/a_1 \in A$, it must also be that $g(\boldsymbol{c}_2) \in S_1$, which contradicts $g(\boldsymbol{c}_2) = \boldsymbol{c}_2/a_2 \in S_2$. **Q.E.D.**

Appendix IV
Proof of Theorem 2

The proof is similar to the proof for Lemma 2 of Appendix II, and is only sketched here. Because the update term of (20) is lower bounded by $\mu Q(||\boldsymbol{s}||_{\max}/\sigma)||\boldsymbol{s}||_{\min}\sin\theta_{\min}$, we can show that the equalizer norm $||\boldsymbol{c}_k||$ grows without bound as k increases, and that there exists a finite M such that $\boldsymbol{c}_{M+1}$ is strictly inside the cone. Because the equalizer norm grows without bound while the update $\mu g(\boldsymbol{c}_k)$ is bounded, it follows that $\boldsymbol{c}_k/||\boldsymbol{c}_k||$ converges to some fixed vector $\tilde{\boldsymbol{c}}_\infty$, and that $g(c_k)$ converges to $g(\tilde{\boldsymbol{c}}_\infty)$. Thus, just as in Appendix II, the update converges to $\mu g(\tilde{\boldsymbol{c}}_\infty)$, which can be lower bounded, so that eventually the equalizer converges to the direction $\tilde{\boldsymbol{c}}_\infty = a g(\tilde{\boldsymbol{c}}_\infty)$. The constant a is positive since each update $\mu g(\boldsymbol{c}_k)$ is within the signal cone. But from Lemma 3 we know that there is only one equalizer with this property, namely $\boldsymbol{c}_{\mathrm{AMBER}}$. **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. Int. Conf. on Communications*, 1974, pp. 25F1–25F5.

[3] P. Galko and S. Pasupathy, "Optimal linear receiver filters for binary digital signals," in *Proc. Int. Conf. on Communications*, 1982, pp. 1H.6.1–1H.6.5.
[4] S. Chen, E. Chng, B. Mulgrew, and G. Gibson, "Minimum-BER linear-combiner DFE," in *Proc. Int. Conf. on Communications*, 1996, pp. 1173–1177.
[5] S. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sept. 1985.
[6] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Boston, MA: Kluwer, 1994.
[7] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a stop-and-go decision-directed algorithm," *IEEE Trans. Commun.*, vol. IT-35, pp. 877–887, Sept. 1987.
[8] N. A. M. Verhoeckx, H. van den Elzen, F. A. M. Snijders, and P. J. van Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 768–781, Dec. 1979.
[9] C.-C. Yeh, "Minimum error probability equalization and multiuser detection," Ph.D. dissertation, Georgia Inst. of Technol., 1998.
[10] C.-C. Yeh and J. R. Barry, "Approximate minimum bit-error rate equalization for pulse-amplitude and quadrature amplitude modulation," in *Proc. IEEE Int. Conf. on Communications*, vol. 1, Atlanta, GA, June 7–11, 1998, pp. 16–20.
[11] C.-C. Yeh, R. R. Lopes, and J. R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. IEEE Global Communications Conf.*, Sydney, Australia, Nov. 1998, pp. 3590–3595.

Chen-Chu Yeh was born in Taipei, Taiwan, in 1969. He 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 and computer engineering from the Georgia Institute of Technology, Atlanta, in 1994 and 1998, respectively.

During the summers of 1992–1995, he worked at the NCR and AT&T Corporations on character recognition, signature verification neural networks, and Reed–Solomon coding for 2-D bar codes. In June 1998, he joined Harris Corporation, Palm Bay, FL, where he is a Systems Analyst in the Communications and Signal Processing Department, working on various aspects of spaceborne and airborne communications systems. His main research interests are in signal processing in digital communications and error correction coding.

John R. Barry 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 at the School of Electrical and Computer Engineering. He is the author of the book, *Wireless Infrared Communications* (Norwell, MA: Kluwer, 1994). His current research interests include wireless communications, equalization, and multiuser communications.