Adaptive Minimum Symbol-Error-Rate Decision Feedback Equalization for Multi-Level Pulse-Amplitude Modulation

S. Chen, L. Hanzo and B. Mulgrew

Abstract—**The design of decision feedback equalizers (DFEs) is typically based on the minimum mean square error (MMSE) principle, as this leads to effective adaptive implementation in the form of the least mean square algorithm. It is well-known, however, that in certain situations the MMSE solution can be distinctly inferior to the optimal minimum symbol error rate (MSER) solution. We consider the MSER design for multi-level pulse-amplitude modulation. Block-data adaptive implementation of the theoretical MSER DFE solution is developed based on the Parzen window estimate of probability density function. Furthermore, a sample-by-sample adaptive MSER algorithm, called the least symbol error rate (LSER), is derived for adaptive equalization application. The proposed LSER algorithm has a complexity that increases linearly with the equalizer length. Computer simulation is employed to evaluate the proposed alternative MSER design for equalization application with multi-level signalling schemes.**

Keywords— **Decision feedback equalizer, minimum mean square error, minimum symbol error rate, stochastic gradient adaptive algorithms.**

I. Introduction

Conventional decision feedback equalizers (DFEs) employ a linear combination of channel observations and past detected symbols to combat channel distortion and is a most widely used equalization scheme for multi-level signalling. In practice, it offers adequate performance at an achievable low computational complexity. Classically, the minimum mean square error (MMSE) criterion is adopted for DFE design and standard adaptive algorithms, such as the least mean square (LMS) algorithm, are used for adaptive implementations. Since the MMSE solution is not optimal in equalization application [1]-[4], research has been conducted to find alternative designs that are based on the minimum bit error rate (MBER) criterion for the binary phase shift keying (BPSK) modulation or the minimum symbol error rate (MSER) criterion for the general multi-level signalling scheme. Most of the published works in this area have been focused on the MBER design for the BPSK scheme [2],[3],[5]–[8]. The MBER design principle has also been applied to the linear multiuser detection for downlink code-division multiple-access systems with the BPSK modulation [9],[10] and to the linear beamforming assisted BPSK receiver [11].

This paper considers an MSER approach for the DFE design with the multi-level pulse-amplitude modulation (M-PAM) scheme. We derive the symbol error rate (SER) expression for the generic DFE. Although there exists no closed-form solution for the MSER DFE, a simplified conjugate gradient algorithm [10],[12] can be applied for numerical solutions. An adaptive implementation of the theoretical MSER DFE is then studied. The Parzen window or kernel density technique [13]–[15] is adopted for approximating the probability density function (p.d.f.) of the DFE output, and this naturally leads to a block-data based adaptive MSER algorithm, which iteratively minimizes the estimated SER of the DFE by adjusting the equalizer's weights using the conjugate gradient optimization method. It is shown in a simulation study that this block-data based adaptive MSER algorithm converges rapidly and the length of the data block required for achieving an accurate approximation of the MSER solution is reasonably small. Sample-by-sample adaptive implementation of the MSER DFE solution is also considered and a stochastic gradient adaptive MSER algorithm, referred to as the least symbol error rate (LSER), is derived. This LSER algorithm has a low computational complexity, comparable to that of the simple LMS algorithm. Simulation results suggest that convergence speed of this stochastic gradient algorithm depends on the initial value of the equalizer weight vector, and the MMSE solution is typically not a good initial condition for the algorithm. Simulation study also suggests that the LSER algorithm converges reasonably fast, provided that the initial condition is not set to the MMSE solution. In the literature, we have found a stochastic gradient algorithm that is based on an approximate MSER approach, called the approximate MSER (AMSER) algorithm [16]. We compare our LSER algorithm with this AMSER algorithm using computer simulation.

II. The decision feedback equalizer

It is assumed that the real-valued channel generates the received signal samples of:

$$x(k) = \sum_{i=0}^{n_h-1} h_i s(k-i) + n(k) \tag{1}$$

where h_i are the channel impulse response (CIR) taps, n_h is the CIR length, the Gaussian white noise $n(k)$ has zero mean and variance σ_n^2, and the M-PAM symbol $s(k)$ takes the value from the symbol set:

$$\mathcal{S} \stackrel{\triangle}{=} \{s_l = 2l - M - 1, \ 1 \le l \le M\} . \tag{2}$$

The channel signal to noise ratio (SNR) is defined as

$$\text{SNR} \stackrel{\triangle}{=} \left(\sum_{i=0}^{n_h-1} h_i^2 \right) \frac{\sigma_s^2}{\sigma_n^2} \tag{3}$$

where σ_s^2 is the symbol variance. The reason for concentrating on real-valued channels and signalling schemes is to keep notations and concepts relatively simple. There is no real difficulty to apply the approach to complex-valued channels and

S. Chen and L. Hanzo are with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. B. Mulgrew is with Institute for Digital Communications, School of Engineering and Electronics, University of Edinburgh, Edinburgh EH9 3JL, U.K.

signalling schemes. The DFE considered has a linear-combiner structure and is defined by:

$$y(k) = \mathbf{w}^T\mathbf{x}(k) + \mathbf{b}^T\hat{\mathbf{s}}_b(k) \tag{4}$$

where $\mathbf{x}(k) = [x(k)\ x(k-1)\cdots x(k-m+1)]^T$ is the noisy observation vector with m being the feedforward order, $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1)\cdots\hat{s}(k-d-n_b)]^T$ is the past detected symbol vector with n_b being the feedback order and d the equalizer decision delay, $\mathbf{w} = [w_0\ w_1\cdots w_{m-1}]^T$ and $\mathbf{b} = [b_1\cdots b_{n_b}]^T$ are the feedforward and feedback filter coefficient vectors, respectively. We will choose $d = n_h - 1$, $m = n_h$ and $n_b = n_h - 1$. Although this choice of the DFE structure parameters is not necessarily optimal, it is sufficient to guarantee that the subsets of the noise-free signal states are linearly separable [3],[4] and therefore guarantee an adequate performance. The equalizer output $y(k)$ is passed to a threshold detector which provides an estimate $\hat{s}(k-d)$ of the transmitted symbol $s(k-d)$.

Using matrix notations with $m = n_h$ and $d = n_b = n_h - 1$, the received signal vector can be expressed as:

$$\mathbf{x}(k) = \mathbf{H}_1\mathbf{s}_f(k) + \mathbf{H}_2\mathbf{s}_b(k) + \mathbf{n}(k) \tag{5}$$

where $\mathbf{n}(k) = [n(k)\ n(k-1)\cdots n(k-m+1)]^T$, $\mathbf{s}_b(k) = [s(k-d-1)\cdots s(k-d-n_b)]^T$, $\mathbf{s}_f(k) = [s(k)\ s(k-1)\cdots s(k-d)]^T$, and the $m\times(d+1)$ and $m\times n_b$ CIR matrices, $\mathbf{H}_1$ and $\mathbf{H}_2$, are given by

$$\mathbf{H}_1 = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} \\ 0 & h_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_1 \\ 0 & \cdots & 0 & h_0 \end{bmatrix} \tag{6}$$

and

$$\mathbf{H}_2 = \begin{bmatrix} 0 & \cdots & 0 \\ h_{n_h-1} & \ddots & \vdots \\ \vdots & \ddots & 0 \\ h_1 & \cdots & h_{n_h-1} \end{bmatrix}, \tag{7}$$

respectively. Under the assumption that the past decisions are correct, $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$ and the received signal vector can be expressed as $\mathbf{x}(k) = \mathbf{H}_1\mathbf{s}_f(k) + \mathbf{H}_2\hat{\mathbf{s}}_b(k) + \mathbf{n}(k)$. Thus, the decision feedback can be viewed to translate the original observation space $\mathbf{x}(k)$ into a new space $\mathbf{r}(k)$ [3],[17],[18]:

$$\mathbf{r}(k) \stackrel{\triangle}{=} \mathbf{x}(k) - \mathbf{H}_2\hat{\mathbf{s}}_b(k) = \mathbf{H}_1\mathbf{s}_f(k) + \mathbf{n}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k). \tag{8}$$

In this translated observation space, a "linear equalizer" can be formulated as:

$$y(k) = \mathbf{w}^T\mathbf{r}(k) = \mathbf{w}^T(\bar{\mathbf{r}}(k) + \mathbf{n}(k)) = \bar{y}(k) + e(k) \tag{9}$$

where $e(k)$ is Gaussian with zero mean and variance $\mathbf{w}^T\mathbf{w}\sigma_n^2$. The elements of $\mathbf{r}(k)$ can be computed recursively according to [3],[4]:

$$\left.\begin{aligned} &r(k-i) = z^{-1}r(k-i+1) - h_{n_h-i}\hat{s}(k-d-1), \\ &\qquad \text{for } i = m-1, m-2, \cdots, 1 \\ &r(k) = x(k) \end{aligned}\right\} \tag{10}$$

where z^{-1} is interpreted as the unit delay operator. Thus, in an adaptive implementation, one needs to estimate the CIR, rather than to estimate $\mathbf{b}$, when adopting the equalizer structure (9) and (10). If the weight vectors $\mathbf{w}$ in both (4) and (9) are set to the MMSE solution $\mathbf{w}_{\text{MMSE}}$, then the equalizer (9) is identical to the DFE (4). In this case, (10) can be obtained by setting the feedback filter coefficient vector to $\mathbf{b} = -\mathbf{H}_2^T\mathbf{w}_{\text{MMSE}}$.

Define the combined impulse response of the equalizer and channel as $\mathbf{c}$, which is given by

$$\mathbf{c}^T = \mathbf{w}^T\mathbf{H}_1 = \mathbf{w}^T[\mathbf{h}_0\ \mathbf{h}_1\cdots\mathbf{h}_d] = [c_0\ c_1\cdots c_d] \tag{11}$$

where the last column of $\mathbf{H}_1$ is $\mathbf{h}_d = [h_{n_h-1}\ h_{n_h-2}\cdots h_1\ h_0]^T$. Then $y(k)$ can be expressed as

$$y(k) = c_d s(k-d) + \sum_{i=0}^{d-1} c_i s(k-i) + e(k). \tag{12}$$

The first term in (12) is the desired signal, and the second term represents the residual intersymbol interference. Thus the optimal decision is made according to

$$\hat{s}(k-d) = \begin{cases} s_1, & \text{if } y(k) \le (s_1+1)c_d\,, \\ s_l, & \text{if } (s_l-1)c_d < y(k) \le (s_l+1)c_d \\ & \quad \text{for } l = 2,\cdots M-1\,, \\ s_M, & \text{if } y(k) > (s_M-1)c_d\,. \end{cases} \tag{13}$$

The scaling factor $c_d > 0$ and is often ignored. Ignoring this scaling factor, however, will introduce bias [19]. Only in the case of $M = 2$ (BPSK), can c_d be ignored, as in this case the only decision boundary $y = 0$ is not affected by the value of c_d.

Notice that $\mathbf{s}_f(k)$ has $N_f = M^{d+1}$ possible combinations. Denote these N_f sequence values of $\mathbf{s}_f(k)$ as $\mathbf{s}_j$, $1 \le j \le N_f$. Then, $\bar{\mathbf{r}}(k)$ only takes values from the signal state set defined by

$$\mathcal{R} \stackrel{\triangle}{=} \{\bar{\mathbf{r}}_j = \mathbf{H}_1\mathbf{s}_j,\ 1 \le j \le N_f\} \tag{14}$$

which can be divided into M subsets conditioned on $s(k-d)$

$$\mathcal{R}_l \stackrel{\triangle}{=} \{\bar{\mathbf{r}}_j \in \mathcal{R} | s(k-d) = s_l\},\ 1 \le l \le M. \tag{15}$$

Similarly, $\bar{y}(k)$ can only take values from the scalar set

$$\mathcal{Y} \stackrel{\triangle}{=} \{\bar{y}_j = \mathbf{w}^T\bar{\mathbf{r}}_j,\ 1 \le j \le N_f\}. \tag{16}$$

$\mathcal{Y}$ can be divided into M subsets, depending on the value of $s(k-d)$

$$\mathcal{Y}_l \stackrel{\triangle}{=} \{\bar{y}_j \in \mathcal{Y} | s(k-d) = s_l\},\ 1 \le l \le M. \tag{17}$$

Note that $\mathcal{R}_l$, $1 \le l \le M$, are always linearly separable [3]. That is, there always exists an equalizer weight vector $\mathbf{w}$ such that the resulting $\mathcal{Y}_l$, $1 \le l \le M$, are completely separated by the decision thresholds $(s_l \pm 1)c_d$.

III. The MSER decision feedback equalizer

Due to the symmetric distribution of the symbol constellation (2), for $l = 1, 2, \cdots, M-1$, $\mathcal{R}_{l+1}$ is the shifted version of $\mathcal{R}_l$ by the amount $2\mathbf{h}_d$. This can easily be verified. From

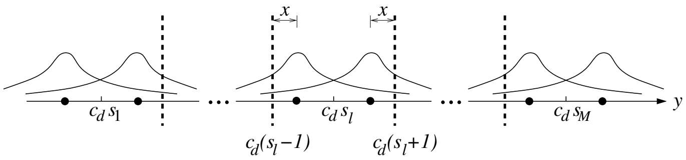


Fig. 1. Illustration of the symmetric distribution of $\mathcal{Y}_l$ around $c_d s_l$, $1 \leq l \leq M$, and the related decision boundaries. The conditional probability density function of $y(k)$ given $s(k-d) = s_l$ is a Gaussian mixture, and each Gaussian component centers at a $\bar{y}_i^{(l)} \in \mathcal{Y}_l$.

the definitions of $\mathbf{H}_1$ and $\mathbf{s}_j$, for each $\bar{\mathbf{r}}_j^{(l)} \in \mathcal{R}_l$, there exists a $\bar{\mathbf{r}}_i^{(l+1)} \in \mathcal{R}_{l+1}$ such that

$$\bar{\mathbf{r}}_i^{(l+1)} = \bar{\mathbf{r}}_j^{(l)} + (s_{l+1} - s_l)\mathbf{h}_d = \bar{\mathbf{r}}_j^{(l)} + 2\mathbf{h}_d \tag{18}$$

that is, $\mathcal{R}_{l+1} = \mathcal{R}_l + 2\mathbf{h}_d$. Thus, $\mathcal{Y}_{l+1}$ is the shifted version of $\mathcal{Y}_l$ by the amount $2\mathbf{w}^T\mathbf{h}_d = 2c_d$, namely $\mathcal{Y}_{l+1} = \mathcal{Y}_l + 2c_d$. This shift property allows us to consider only one subset $\mathcal{Y}_l$ in evaluating the SER of the DFE (9) with the decision rule (13). We further point out a symmetric structure of $\mathcal{Y}_l$, namely the points of $\mathcal{Y}_l$ are distributed symmetrically around the symbol point $c_d s_l$. In particular, given a point in $\mathcal{Y}_l$ with a distance x to the decision boundary $c_d(s_l - 1)$, there is another point in $\mathcal{Y}_l$ which has the same distance to the other decision boundary $c_d(s_l + 1)$. This symmetry property, as illustrated in Fig. 1, is useful in simplifying the SER expression.

A. The expression of symbol error rate

The conditional p.d.f. of $y(k)$ given $s(k-d) = s_l$ is a Gaussian mixture given by

$$p_y(y|s_l) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \sum_{i=1}^{N_{sb}} \exp\left(-\frac{\left(y - \bar{y}_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{19}$$

where N_{sb} is the number of points in $\mathcal{Y}_l$. Referring to Fig. 1, for $\mathcal{Y}_l, l = 2, \cdots, M-1$, an error occurs if $y(k)$ is outside the region $[c_d(s_l-1),\ c_d(s_l+1)]$. For $\mathcal{Y}_1$ and $\mathcal{Y}_M$, however, an error only occurs at one side. The conditional SER, for $l = 2, \cdots, M-1$, is

$$P_{E,l}(\mathbf{w}) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \times \sum_{i=1}^{N_{sb}} \left\{ \int_{-\infty}^{c_d s_l - c_d} + \int_{c_d s_l + c_d}^{\infty} \right\} \exp\left(-\frac{\left(y - \bar{y}_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) dy\,. \tag{20}$$

Taking into account the symmetric distribution of $\mathcal{Y}_l$ as illustrated in Fig. 1,

$$P_{E,l}(\mathbf{w}) = \frac{2}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \times \sum_{i=1}^{N_{sb}} \int_{-\infty}^{c_d s_l - c_d} \exp\left(-\frac{\left(y - \bar{y}_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) dy = \frac{2}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})) \tag{21}$$

where

$$g_{l,i}(\mathbf{w}) = \frac{\bar{y}_i^{(l)} - c_d s_l + c_d}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \tag{22}$$

and

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{y^2}{2}\right) dy\,. \tag{23}$$

Using the shift property and noticing the fact that an error only occurs at one side for $\mathcal{Y}_1$ and $\mathcal{Y}_M$, the SER of the DFE (9) can be expressed as

$$P_E(\mathbf{w}) = \frac{M-1}{M}\frac{2}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})) = \frac{\gamma}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})) \tag{24}$$

where $\gamma = \frac{2M-2}{M}$. It is seen that the SER can be evaluated using any single subset $\mathcal{Y}_l$.

B. The minimum symbol-error-rate solution

The MSER solution is defined as

$$\mathbf{w}_{\text{MSER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w}) \tag{25}$$

Note that, since the SER is invariant to a positive scaling of $\mathbf{w}$, there is an infinite number of MSER solutions. In fact, if $\mathbf{w}_{\text{MSER}}$ is an MSER solution, then $\alpha\mathbf{w}_{\text{MSER}}$ are all MSER solutions for any $\alpha > 0$. Unlike the MMSE, there exists no closed-form solution for $\mathbf{w}_{\text{MSER}}$, and a numerical solution has to be sought. The gradient of $P_E(\mathbf{w})$ with respect to $\mathbf{w}$ is

$$\nabla P_E(\mathbf{w}) = \frac{\gamma}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \sum_{i=1}^{N_{sb}} \exp\left(-\frac{(\bar{y}_i^{(l)} - c_d(s_l-1))^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \times \left(\frac{(\bar{y}_i^{(l)} - c_d(s_l-1))}{\mathbf{w}^T\mathbf{w}}\mathbf{w} - \bar{\mathbf{r}}_i^{(l)} + (s_l-1)\mathbf{h}_d\right). \tag{26}$$

With the gradient, the optimization problem (25) can be solved iteratively using a gradient-based optimization algorithm. It is also advantageous to normalize $\mathbf{w}$ to a unit-length after every iteration so that the gradient can be simplified as

$$\nabla P_E(\mathbf{w}) = \frac{\gamma}{N_{sb}\sqrt{2\pi}\sigma_n} \sum_{i=1}^{N_{sb}} \exp\left(-\frac{(\bar{y}_i^{(l)} - c_d(s_l-1))^2}{2\sigma_n^2}\right) \times \left((\bar{y}_i^{(l)} - c_d(s_l-1))\mathbf{w} - \bar{\mathbf{r}}_i^{(l)} + (s_l-1)\mathbf{h}_d\right). \tag{27}$$

Computational requirements are further simplified by considering the subset $\mathcal{Y}_l$ with $l = \frac{M}{2} + 1$, which results in $s_l - 1 = 0$. We point out that the simplified conjugate gradient algorithm [10],[12] offers an efficient means to find an MSER solution, and this algorithm is summarized in the following.

Initialization. Choose step size $\mu > 0$ and termination scalar $\beta > 0$; given $\mathbf{w}(0)$ and $\mathbf{d}(0) = -\nabla P_E(\mathbf{w}(0))$; set iteration index $\iota = 0$.

Loop. If $\|\nabla P_E(\mathbf{w}(\iota))\| < \beta$: goto *Stop*.

$$\mathbf{w}(\iota+1) = \mathbf{w}(\iota) + \mu\mathbf{d}(\iota)$$

$$\mathbf{w}(\iota+1) = \frac{\mathbf{w}(\iota+1)}{\|\mathbf{w}(\iota+1)\|}$$

$$\phi_\iota = \frac{\|\nabla P_E(\mathbf{w}(\iota+1))\|^2}{\|\nabla P_E(\mathbf{w}(\iota))\|^2}$$
$$\mathbf{d}(\iota+1) = \phi_\iota \mathbf{d}(\iota) - \nabla P_E(\mathbf{w}(\iota+1))$$

$\iota = \iota + 1$, goto *Loop*.
Stop. $\mathbf{w}(\iota)$ is the solution.

At a minimum, $\|\nabla P_E(\mathbf{w})\| = 0$. So the termination scalar β determines the accuracy of the solution obtained. The step size μ controls the rate of convergence. Typically, a much larger value of μ can be used, compared with the steepest descent gradient algorithm. As the SER surface $P_E(\mathbf{w})$ is highly complicated, occasionally the search direction $\mathbf{d}$ may no longer be a good approximation to the conjugate gradient direction or may even point to the uphill direction when the iteration number becomes large. It is thus advisable to periodically re-set $\mathbf{d}$ to the negative gradient in the above conjugate gradient algorithm.

In theory, there is no guarantee that the above algorithm can always find a global minimum point of the SER surface $P_E(\mathbf{w})$. In practice, we have found that the algorithm works well and we have never observed any case of the algorithm being trapped at some local minimum solutions. This is perhaps because of the special shape of the SER surface. As the SER is invariant to a positive scaling of $\mathbf{w}$, the size of $\mathbf{w}$ does not matter (except zero size). Thus, the SER surface has an infinitely long valley, and any point at the bottom of this valley is a true global MSER solution. Once a weight vector $\mathbf{w}$ is near the edge of this infinitely long valley, convergence to the bottom is assured at a very fast rate, as the slope or gradient is very large. For an illustration of the SER surface, consider the simple example in which the transfer function of the CIR is $H(z) = 1.0 + 0.5z^{-1}$ and the signalling scheme is 4-PAM. A two-tap equalizer with $d = 0$ and without decision feedback is used. At the SNR of 35 dB, the SER surface $P_E(\mathbf{w})$ is depicted in Fig. 2.

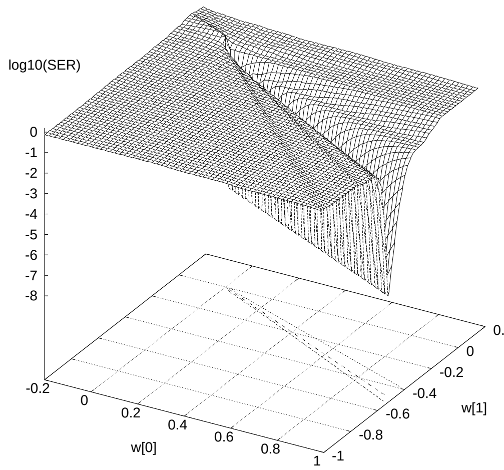


Fig. 2. Symbol error rate surface for the simple illustrative example. The central contour line: $\log_{10}\mathrm{SER} = -7.16$, the outer contour: $\log_{10}\mathrm{SER} = -2.76$, $\log_{10} P_E(\mathbf{w}_{\mathrm{MMSE}}) = -2.76$ and $\log_{10} P_E(\mathbf{w}_{\mathrm{MSER}}) = -7.16$.

IV. The adaptive MSER DFE

To derive an adaptive version of the MSER algorithm, it is more convenient to write down the p.d.f. of $y(k)$ explicitly:

$$p_y(y) = \frac{1}{\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{N_f}\sum_{l=1}^{M}\sum_{i=1}^{N_{sb}}\exp\left(-\frac{\left(y-\bar{y}_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \quad (28)$$

and express the SER alternatively as:

$$P_E(\mathbf{w}) = \gamma\frac{1}{N_f}\sum_{l=1}^{M}\sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})). \quad (29)$$

In reality, the p.d.f. of $y(k)$ is unknown, and the key to adaptive implementation of the MSER solution is an effective estimate of the p.d.f. (28). Parzen window estimate, also known as kernel density estimate, [13]–[15] is a well known method for estimating a probability distribution. Parzen window method estimates a p.d.f. using a window or block of $y(k)$ by placing a symmetric unimodal kernel function on each $y(k)$. Kernel density estimation is capable of producing reliable p.d.f. estimates with short data records and in particular is natural when dealing with Gaussian mixtures. In our particular application, it is obvious and natural to choose a Gaussian kernel function with a kernel width $\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}$ that is similar in form to the noise standard deviation $\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}$ for $e(k)$.

A. The block-data gradient adaptive MSER algorithm

Given a block of K training samples $\{\mathbf{r}(k), s(k-d)\}$, a Parzen window estimate of the p.d.f. (28) is given by:

$$\hat{p}_y(y) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{K}\sum_{k=1}^{K}\exp\left(-\frac{(y-y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right). \quad (30)$$

The p.d.f. estimate (30) is known to possess a mean integrated square error convergence rate at order of K^{-1} [13]. The radius parameter ρ_n is related to the noise standard deviation σ_n for $n(k)$. In [14], a lower bound of $\rho_n = \left(\frac{4}{3K}\right)^{\frac{1}{5}}\sigma_n$ is suggested. In practice, ρ_n can often be chosen from a large range of values.

From the estimated p.d.f. (30), the estimated SER is given by

$$\hat{P}_E(\mathbf{w}) = \gamma\frac{1}{K}\sum_{k=1}^{K} Q(\hat{g}_k(\mathbf{w})) \quad (31)$$

where

$$\hat{g}_k(\mathbf{w}) = \frac{y(k) - \hat{c}_d s(k-d) + \hat{c}_d}{\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}, \quad (32)$$

$\hat{c}_d = \mathbf{w}^T\hat{\mathbf{h}}_d$, and $\hat{\mathbf{h}}_d$ an estimate of $\mathbf{h}_d$. The gradient of $\hat{P}_E(\mathbf{w})$ is given by

$$\nabla\hat{P}_E(\mathbf{w}) = \frac{\gamma}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{K}\times$$
$$\sum_{k=1}^{K}\exp\left(-\frac{(y(k)-\hat{c}_d(s(k-d)-1))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right)\times$$
$$\left(\frac{(y(k)-\hat{c}_d(s(k-d)-1))}{\mathbf{w}^T\mathbf{w}}\mathbf{w} - \mathbf{r}(k) + (s(k-d)-1)\hat{\mathbf{h}}_d\right). \quad (33)$$

By substituting $\nabla P_E(\mathbf{w})$ with $\nabla\hat{P}_E(\mathbf{w})$ in the conjugate gradient updating mechanism, a block-data gradient adaptive algorithm is readily obtained. Again the step size μ controls the rate of convergence. The estimate $\hat{\mathbf{h}}_d$ is the MMSE estimate of the CIR. In adaptive implementation, the LMS algorithm can be used to update $\hat{\mathbf{h}}_d$. Note that this is not an "extra" work required by the MSER approach. The estimate of c_d and therefore the CIR $\mathbf{h}_d$ is always required to implement the optimal decision rule (13). Furthermore, as pointed out in Section II, the estimation of the CIR $\mathbf{h}_d$ substitutes the estimation of the feedback coefficient vector $\mathbf{b}$. In fact, estimating $\mathbf{h}_d$ is better than estimating $\mathbf{b}$. This is because the former problem has a unit eigenvalue spread, while the latter problem is tied with the estimation of $\mathbf{w}$ which has a larger eigenvalue spread.

B. The least symbol error rate algorithm

In the Parzen window estimate (30), the kernel width $\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}$ depends on the equalizer weight vector $\mathbf{w}$. In a general density estimate, there is no reason why the kernel width should be chosen in such a way except that we notice the dependency of the width to $\mathbf{w}$ in the true density (28). However, the SER is invariant to $\mathbf{w}^T\mathbf{w}$. To fully take advantage of this fact, we propose to use a constant width ρ_n in density estimate. One advantage of using a constant width ρ_n, rather than a variable one $\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}$, in the density estimate is that the gradient of the resulting estimated SER has a much simpler form, which leads to considerable reduction in computational complexity. This is particular relevant in the derivation of stochastic gradient updating mechanisms. Adopting this approach, an alternative Parzen window estimate of the true p.d.f. (28) is given by

$$\tilde{p}_y(y) = \frac{1}{\sqrt{2\pi}\rho_n}\frac{1}{K}\sum_{k=1}^{K}\exp\left(-\frac{(y-y(k))^2}{2\rho_n^2}\right) \tag{34}$$

and an approximation of the SER is

$$\tilde{P}_E(\mathbf{w}) = \gamma\frac{1}{K}\sum_{k=1}^{K}Q(\tilde{g}_k(\mathbf{w})) \tag{35}$$

with

$$\tilde{g}_k(\mathbf{w}) = \frac{y(k)-\hat{c}_d s(k-d)+\hat{c}_d}{\rho_n}. \tag{36}$$

This approximation is valid provided that the width ρ_n is chosen appropriately.

To derive a sample-by-sample adaptive algorithm, consider a single-sample estimate of $p_y(y)$:

$$\tilde{p}_y(y,k) = \frac{1}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{(y-y(k))^2}{2\rho_n^2}\right) \tag{37}$$

and the corresponding single-sample SER "estimate" $\tilde{P}_E(\mathbf{w},k)$. Using the instantaneous stochastic gradient

$$\nabla\tilde{P}_E(\mathbf{w},k) = \frac{\gamma}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{(y(k)-\hat{c}_d(s(k-d)-1))^2}{2\rho_n^2}\right) \times\Big((s(k-d)-1)\hat{\mathbf{h}}_d-\mathbf{r}(k)\Big) \tag{38}$$

gives rise to the following stochastic gradient adaptive algorithm which we refer to as the LSER:

$$\mathbf{w}(k+1) = \mathbf{w}(k)+\frac{\mu\gamma}{\sqrt{2\pi}\rho_n}\times \exp\left(-\frac{(y(k)-\hat{c}_d(k)(s(k-d)-1))^2}{2\rho_n^2}\right)\times \Big(\mathbf{r}(k)-(s(k-d)-1)\hat{\mathbf{h}}_d(k)\Big) \tag{39}$$

where the adaptive gain μ and width ρ_n are the two algorithm parameters that need to be set appropriately. Note that there is no need to normalize the weight vector after each updating.

Convergence analysis of this LSER algorithm is yet to be worked out. But the previous empirical results using this algorithm in some other applications with BPSK signalling [11],[20] have suggested that the algorithm behaves well and has a reasonably fast convergence rate. There also exists some theoretical analysis to support this empirical observation. The terms involving $s(k-d)$ in the algorithm (39) can be seen to perform some kind of shifting operation which is needed due to the multi-level signalling nature. For simplicity, let us consider the BPSK scheme. In this case, it can easily be shown that the LSER algorithm (39) becomes [11],[20]:

$$\mathbf{w}(k+1) = \mathbf{w}(k)+\mu\frac{\mathrm{sgn}(s(k-d))}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{y^2(k)}{2\rho_n^2}\right)\mathbf{r}(k). \tag{40}$$

This belongs to the general stochastic gradient-based adaptive algorithm investigated in [21]. Therefore, the results of convergence analysis presented in [21] can readily be applied.

C. The approximate minimum symbol-error-rate algorithm

The AMSER algorithm [16], originally derived for linear equalizer, is also a stochastic gradient algorithm based on the considerations of minimizing the SER. For a comparison purpose, the AMSER algorithm is summarized here. The algorithm updates the weight vector using

$$\mathbf{w}(k+1) = \mathbf{w}(k)+\mu I_k(\tau)\mathrm{sgn}(\hat{e}(k))\mathbf{r}(k) \tag{41}$$

where the error function $\hat{e}(k)$ is given by

$$\hat{e}(k) = s(k-d)-\frac{y(k)}{\hat{c}_d(k)}, \tag{42}$$

$\hat{c}_d(k) = \mathbf{w}^T(k)\hat{\mathbf{h}}_d(k)$ and the indication function $I_k(\tau)$ is defined by

$$I_k(\tau) = \begin{cases} 1, & \text{if } \frac{y(k)}{\hat{c}_d(k)} < s(k-d)-1+\tau \text{ and} \\ & \quad s(k-d)\neq -M+1 \\ & \text{or } \frac{y(k)}{\hat{c}_d(k)} > s(k-d)+1-\tau \text{ and} \\ & \quad s(k-d)\neq M-1\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{43}$$

The adaptive gain μ and the nonnegative threshold τ are the two algorithm parameters that need to be set appropriately. The AMSER has the same number of algorithm parameters that require tuning as the LSER.

V. Simulation study

Two examples were used in a simulation study to investigate the proposed MSER approach for the DFE design.

Example 1. The transfer function of the CIR was $H(z) = 0.15+0.6z^{-1}+1.0z^{-2}-0.6z^{-3}$ and the 4-PAM signalling

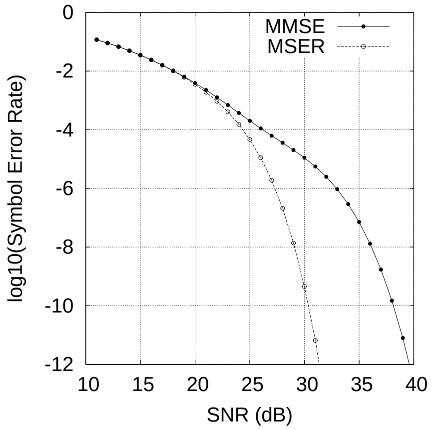


Fig. 3. Theoretical symbol error rate comparison for the MMSE and MSER decision feedback equalizers of Example 1.

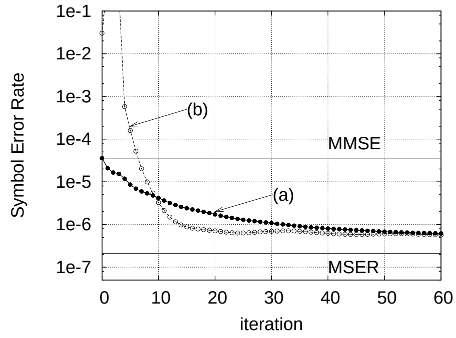


Fig. 5. Convergence rates of the block-data gradient adaptive MSER algorithm for Example 1 given SNR$=$ 28 dB, and with a block size $K = 600$, step size $\mu = 0.9$ and squared width $\rho_n^2 = 2\sigma_n^2 \approx 0.03$. (a) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, and (b) $\mathbf{w}(0) = [-0.01\ 0.01\ 0.01\ 0.01]^T$.

scheme was employed. The structure parameters of the DFE were accordingly set to $m = 4$, $d = 3$ and $n_b = 3$. The theoretical SERs of the MMSE and MSER DFEs, computed using the expression (24) on the single subset $\mathcal{Y}_3$, are given in Fig. 3, where the MSER solutions were computed numerically using the simplified conjugated gradient algorithm of Subsection III-B. The theoretical SERs shown in Fig. 3 represent lower-bound performance, as they were obtained assuming correct symbols being fed back. To investigate the effects of error propagation, the SERs of the MMSE and MSER DFEs were also calculated using simulation with detected symbols being fed back, and the results obtained are depicted in Fig. 4, in comparison with the theoretical SERs. It is interesting to see that for this example the degradation due to error propagation was negligible for the MSER DFE while the effects of error propagation were serious for the MMSE DFE.

The performance of the block-data gradient adaptive MSER algorithm employing the conjugate gradient updating mechanism, as described in Subsection IV-A, was next investigated. Fig. 5 illustrates the convergence rates of the algorithm under SNR$=$ 28 dB, the block size $K = 600$, and two different initial weight vector conditions: (a) $\mathbf{w}(0)$ set to the MMSE solution, and (b) $\mathbf{w}(0)$ arbitrarily set to $[-0.01\ 0.01\ 0.01\ 0.01]^T$.

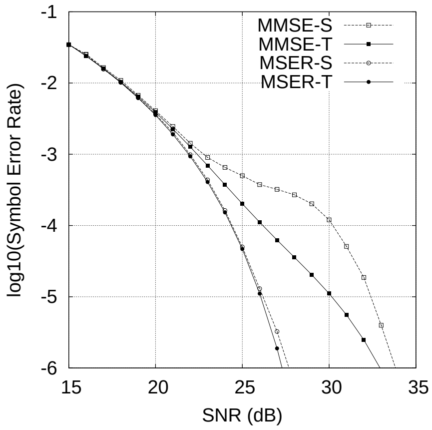


Fig. 4. Effects of error propagation on symbol error rate for the MMSE and MSER decision feedback equalizers of Example 1, where T: theoretical SER, and S: simulated SER with detected symbols being fed back.

From Fig. 5, it can be seen that this block-data gradient adaptive MSER algorithm converges rapidly. The effect of the block size to the performance of this block-data gradient adaptive MSER algorithm was also investigated, and the results obtained are summarized in Fig. 6, where it can be seen that the block size K required for the algorithm to closely approximate the MSER solution is reasonably small considering the fact that the signal set $\mathcal{Y}$ contains 256 points. The performance of the two stochastic gradient adaptive MSER algorithms, the LSER and AMSER discussed respectively in Subsections IV-B and IV-C, were then studied. Figs. 7 and 8 show the learning curves of the two algorithms averaged over 50 runs, given SNR$=$ 28 dB and two different initial weight vector conditions, respectively.

In the investigation, each stochastic gradient MSER algorithm operated in the two modes, namely the training mode in which $s(k-d)$ was known and the decision-directed (DD) mode in which the detected symbol $\hat{s}(k-d)$ was used to substitute for $s(k-d)$. In Fig. 7 (a) and (b), and in Fig. 8 (b), the learning curves corresponding to the two operation modes are indistinguishable. For a comparison purpose, we carefully tuned the algorithm parameters, μ and ρ_n^2 for the LSER and μ and τ for the AMSER, so that each algorithm achieved its best performance both in terms of convergence speed and steady-state SER misadjustment. Comparing Fig. 7 with Fig. 8, it is interesting to see that the convergence rate of a stochastic gradient adaptive MSER algorithm was much slower with the initial

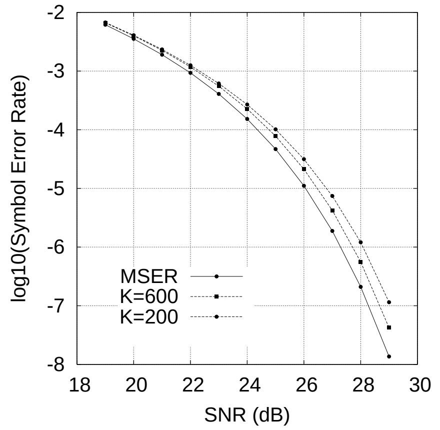
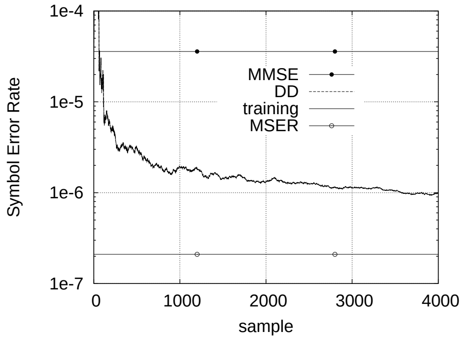
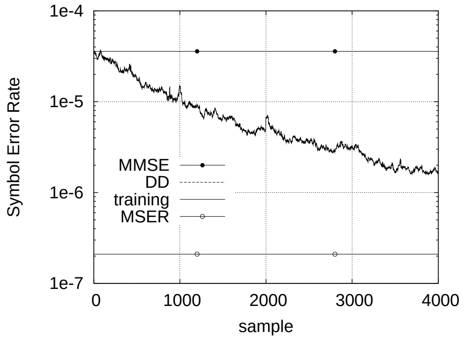


Fig. 6. Effect of block size on the performance of the block-data gradient adaptive MSER algorithm for Example 1.

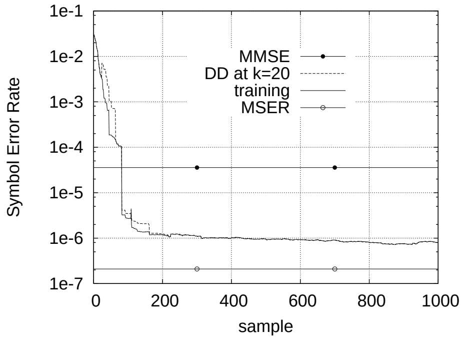


(a) The LSER algorithm with $\mu = 0.01$ and $\rho_n^2 = 30\sigma_n^2 \approx 0.4$

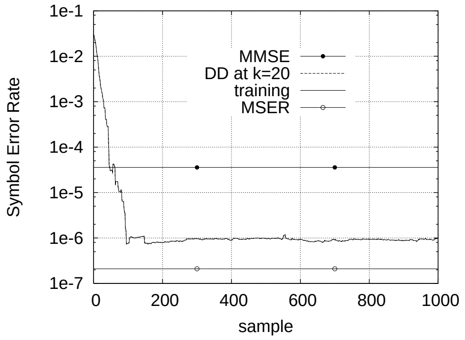


(b) The AMSER algorithm with $\mu = 0.005$ and $\tau = 0.5$

Fig. 7. Learning curves of the two stochastic gradient adaptive MSER algorithms averaged over 50 runs for Example 1 given SNR= 28 dB and with $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, where DD denotes decision directed adaptation with $\hat{s}(k-d)$ substituting $s(k-d)$. DD (dashed) and training (solid) curves are indistinguishable.



(a) The LSER algorithm with $\mu = 0.01$ and $\rho_n^2 = 3\sigma_n^2 \approx 0.04$



(b) The AMSER algorithm with $\mu = 0.005$ and $\tau = 0.4$

Fig. 8. Learning curves of the two stochastic gradient adaptive MSER algorithms averaged over 50 runs for Example 1 given SNR= 28 dB and with $\mathbf{w}(0) = [-0.01\ 0.01\ 0.01\ 0.01]^T$, where DD denotes decision directed adaptation with $\hat{s}(k-d)$ substituting $s(k-d)$.

weight condition set to the MMSE solution than with an arbitrarily set initial weight condition of $[-0.01\ 0.01\ 0.01\ 0.01]^T$. It appears that the MMSE solution is a bad choice as the initial condition for the both stochastic gradient adaptive MSER algorithms. This was also observed at another application involving the BPSK signalling scheme [20]. For this example, with $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, the LSER had a much faster convergence speed than the AMSER, while for the other arbitrarily set initial condition both algorithms had a similar convergence rate.

Example 2. The transfer function of the CIR was $H(z) = 0.3 + 1.0z^{-1} - 0.3z^{-2}$ with 8-PAM symbols. The structure of the DFE was specified by $m = 3$, $d = 2$ and $n_b = 2$. The theoretical lower-bound SERs of the MMSE and MSER DFEs, assuming correct symbols being fed back, are plotted in Fig. 9. Fig. 10 shows the effects of error propagation on the SER performance of the MMSE and MSER DFEs. The convergence rates of the block-data gradient adaptive MSER algorithm are illustrated in Fig. 11 with SNR= 33 dB and two different initial weight vector conditions. The effect of block size on the performance of this block-data gradient adaptive MSER algorithm is shown in Fig. 12. Figs. 13 and 14 compare the convergence rates of the two stochastic gradient algorithms, the LSER and AMSER, given SNR= 33 dB and with two different initial weight vector conditions, respectively. The adaptive algorithm parameters, μ and ρ_n^2 for the LSER and μ and τ for the AMSER, were found empirically to ensure good performance in terms of convergence speed and steady-state SER misadjustment.

The results obtained again demonstrate that the MSER design can outperform the MMSE design in terms of a smaller SER, and the MSER DFE appears to be more robust to error propagation than the MMSE DFE. The proposed block-data gradient adaptive MSER algorithm has a rapid convergence rate and it requires a relatively small block size to accurately approximate the theoretical MSER solution. For this example, the LSER algorithm has a faster convergence rate than the AMSER algorithm. It is also seen that the MMSE solution is not a good choice as the initial weight vector condition for the both stochastic gradient adaptive MSER algorithms.

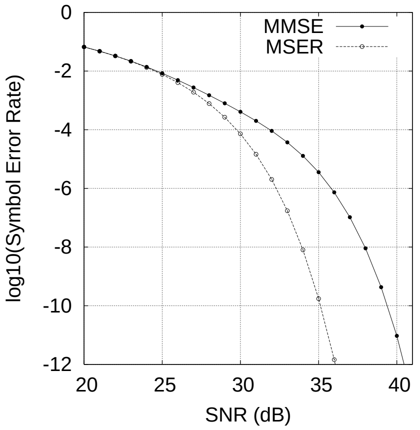


Fig. 9. Theoretical symbol error rate comparison for the MMSE and MSER decision feedback equalizers of Example 2.

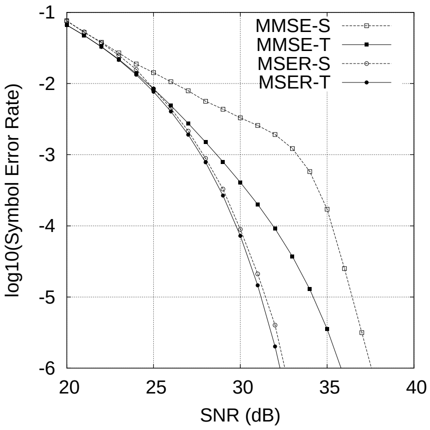


Fig. 10. Effects of error propagation on symbol error rate for the MMSE and MSER decision feedback equalizers of Example 2, where T: theoretical SER, and S: simulated SER with detected symbols being fed back.

VI. Conclusions

An approach based on direct minimization of SER has been proposed for the DFE design with M-PAM signalling scheme. It has been demonstrated that the MSER solution is capable of achieving significant performance gains in terms of reduced SER over the traditional MMSE solution. Adaptive implementation of the proposed MSER design has been developed based on a classical Parzen window estimate of the p.d.f. of the DFE output. A block-data based conjugate gradient adaptive MSER algorithm has been shown to converge rapidly and requires a reasonable small data block size to accurately approximate the theoretical MSER solution. An LMS-style stochastic gradient adaptive MSER algorithm has been shown to perform well, provided that the initial condition is set appropriately. Other interesting results obtained in this study include that the MSER design seems to be less sensitive to error propagation than the MMSE design and the proposed LSER algorithm has faster convergence rate than an existing LMS-style stochastic gradient adaptive MSER algorithm known as the AMSER algorithm.

References

[1] 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. ICC'74*, 1974, pp.25F1–25F5.

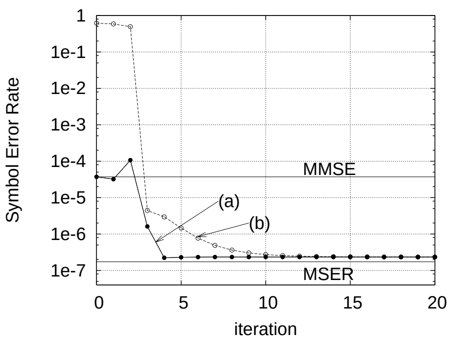


Fig. 11. Convergence rates of the block-data gradient adaptive MSER algorithm for Example 2 given SNR= 33 dB and with a block size $K = 600$. (a) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, step size $\mu = 0.2$ and squared width $\rho_n^2 = 2\sigma_n^2 \approx 0.025$; and (b) $\mathbf{w}(0) = [-0.01\ 0.01\ 0.01]^T$, step size $\mu = 0.05$ and squared width $\rho_n^2 = 2\sigma_n^2 \approx 0.025$.

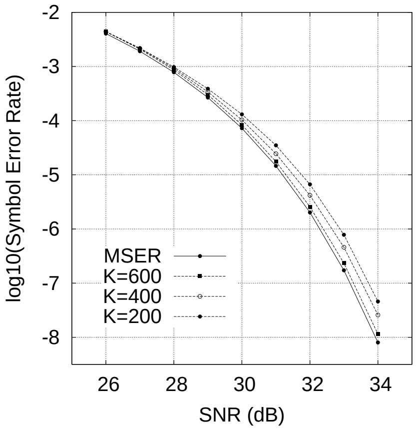
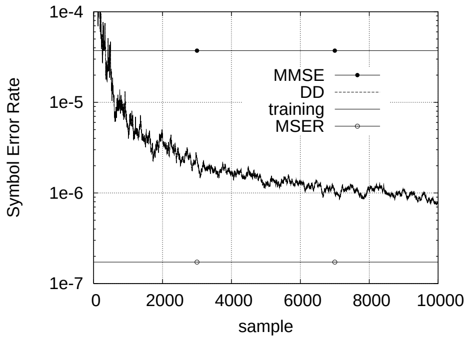
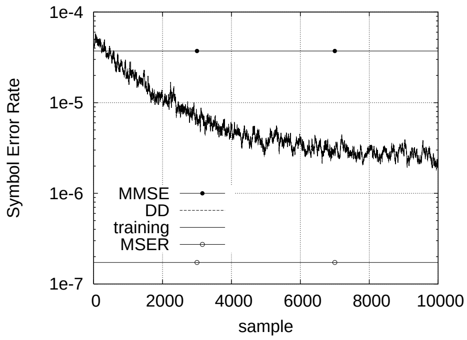


Fig. 12. Effect of block size on the performance of the block-data gradient adaptive MSER algorithm for Example 2.



(a) The LSER algorithm with $\mu = 0.001$ and $\rho_n^2 = 20\sigma_n^2 \approx 0.25$



(b) The AMSER algorithm with $\mu = 0.0005$ and $\tau = 0.8$

Fig. 13. Learning curves of the two stochastic gradient adaptive MSER algorithms averaged over 50 runs for Example 2 given SNR= 33 dB and with $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, where DD denotes decision directed adaptation with $\hat{s}(k-d)$ substituting $s(k-d)$.

[2] S. Chen, E.S. Chng, B. Mulgrew and G. Gibson, "Minimum-BER linear-combiner DFE," in *Proc. ICC'96* (Dallas, Texas), 1996, Vol.2, pp.1173–1177.

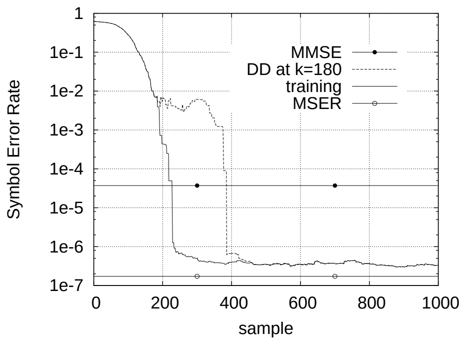
[3] S. Chen, B. Mulgrew, E.S. Chng and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, Vol.145, No.5, pp.316–322, 1998.

[4] S. Chen and B. Mulgrew, "The minimum-SER linear-combiner decision feedback equalizer," *IEE Proc. Communications*, Vol.146, No.6, pp.347–353, 1999.

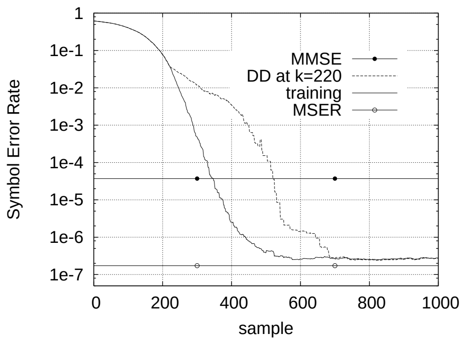
[5] C.C. Yeh and J.R. Barry, "Approximate minimum bit-error rate equalization for binary signaling," in *Proc. ICC'97* (Montreal, Canada), 1997, Vol.2, pp.1095–1099.

[6] C.C. Yeh and J.R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Communications*, Vol.48, No.7, pp.1226–1235, 2000.

[7] B. Mulgrew and S. Chen, "Stochastic gradient minimum-BER decision



(a) The LSER algorithm with $\mu = 0.002$ and $\rho_n^2 = 3\sigma_n^2 \approx 0.04$



(b) The AMSER algorithm with $\mu = 0.0005$ and $\tau = 0.4$

Fig. 14. Learning curves of the two stochastic gradient adaptive MSER algorithms averaged over 50 runs for Example 2 given SNR$=$ 33 dB and with $\mathbf{w}(0) = [-0.01\ 0.01\ 0.01]^T$, where DD denotes decision directed adaptation with $\hat{s}(k-d)$ substituting $s(k-d)$.

feedback equalisers," in *Proc. IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control* (Lake Louise, Alberta, Canada), Oct.1-4, 2000, pp.93–98.

[8] B. Mulgrew and S. Chen, "Adaptive minimum-BER decision feedback equalisers for binary signalling," *Signal Processing*, Vol.81, No.7, pp.1479–1489, 2001.

[9] C.C. Yeh, R.R. Lopes and J.R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. Globecom'98* (Sydney, Australia), Nov. 1998, pp.3590–3595.

[10] S. Chen, A.K. 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, No.6, pp.1240–1247, 2001.

[11] S. Chen, L. Hanzo and N.N. Ahmad, "Adaptive minimum bit error rate beamforming assisted receiver for wireless communications," in *Proc. ICASSP 2003* (Hong Kong, China), April 6-10, 2003, pp.640–643.

[12] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley, 1993.

[13] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, Vol.33, pp.1066–1076, 1962.

[14] B.W. Silverman, *Density Estimation*. London: Chapman Hall, 1996.

[15] A.W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997.

[16] C.C. Yeh and J.R. Barry, "Approximate minimum bit-error rate equalization for pulse-amplitude and quadrature-amplitude modulation," in *Proc. ICC'98* (Atlanta, USA), 1998, Vol.1, pp.16–20.

[17] A.P. Clark, L.H. Lee and R.S. Marshall, "Developments of the conventional non-linear equaliser," *IEE Proc. Part F*, Vol.129, No.2, pp.85–94, 1982.

[18] D. Williamson, R.A. Kennedy and G.W. Pulford, "Block decision feedback equalization," *IEEE Trans. Communications*, Vol.40, No.2, pp.255-264, 1992.

[19] 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. Communications*, Vol.43, No.10, pp.2582-2594, 1995.

[20] S. Chen, "Least bit error rate adaptive multiuser detection," in: L.P. Wang, eds. *Soft Computing for Communications*, Physica Verlag, 2003, to appear.

[21] R. Sharma, W.A. Sethares and J.A. Bucklew, "Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions," *IEEE Trans. Signal Processing*, Vol.44, No.9, pp.2186–2194, 1996.

Sheng Chen obtained a BEng degree in control engineering from the East China Petroleum Institute in 1982, and a PhD degree in control engineering from the City University at London in 1986. He joined the Department of Electronics and Computer Science at the University of Southampton in September 1999. He previously held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth. Dr Chen is a Senior Member of IEEE in the USA. His recent research works include adaptive nonlinear signal processing, modeling and identification of nonlinear systems, neural network research, finite-precision digital controller design, evolutionary computation methods and optimization. He has published over 200 research papers.

Lajos Hanzo received his degree in electronics in 1976 and his doctorate in 1983. During his career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the Department of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He co-authored 10 books totalling 8000 pages on mobile radio communications, published about 450 research papers, organised and chaired conference sessions, presented overview lectures and has been awarded a number of distinctions. Currently he heads an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. Lajos is also an IEEE Distinguished Lecturer of both the Communications as well as the Vehicular Technology Society and a Fellow of the IEE. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk

Bernard Mulgrew received the B.Sc. degree in electrical and electronic engineering in 1979 from Queen's University, Belfast, U.K. He received the Ph.D. degree in 1987 from the University of Edinburgh, U.K. After graduation, he worked for four years as a development engineer with the Radar System Department, GEC-Marconi Avionics, Edinburgh, U.K. From 1983 to 1986, he was a research associate with the Department of Electrical Engineering, University of Edinburgh. He became a lecturer in 1986 and was promoted to senior lectureship and readership in 1994 and 1996, respectively. He was elected to the Personal Chair in Signal and Systems in 1999. Currently he is the head of Institute for Digital Communications, School of Engineering and Electronics, University of Edinburgh. Professor Mulgrew's research interests are in adaptive signal processing and estimation theory and in their application to radar, audio and communications systems. He is a co-author of three books on signal processing, and has published over 300 research papers.