Stochastic Least-Symbol-Error-Rate Adaptive Equalization for Pulse-Amplitude Modulation

S. Chen†, B. Mulgrew‡ and L. Hanzo†
† Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.
‡ Department of Electronics and Electrical Engineering
University of Edinburgh, King's Buildings, Edinburgh EH9 3JL, U.K.

Abstract—**The paper derives a stochastic-gradient minimum symbol-error-rate (MSER) algorithm, called the least symbol error rate (LSER), for training the linear equalizer and linear-combiner decision feedback equalizer (DFE) with M-PAM signalling. This LSER algorithm has some performance advantages, in terms of faster convergence rate and smaller steady-state symbol error rate (SER) misadjustment, over an existing simpler stochastic-gradient adaptive MSER algorithm called the approximate MSER (AMSER).**

I. INTRODUCTION

Design of linear equalizer and DFE is typically based on the minimum mean square error (MMSE) criterion. As the MMSE solution is not optimal [1]-[3], research has been looking into LMS-style adaptive algorithms based on the MSER or minimum bit error rate (MBER) criterion. For binary schemes, two such algorithms have emerged, called the approximate MBER (AMBER) [4],[5] and the least bit error rate (LBER) [6],[7], respectively. The AMBER is simpler than the LBER, although the complexity of LBER is still linear in the equalizer length. The LBER has some performance advantages over the AMBER in terms of faster convergence speed and smaller steady-state BER misadjustment.

This paper presents an adaptive MSER algorithm for linear equalizer and DFE with M-PAM symbols. We adopt the approach used in deriving the LBER algorithm [6],[7], namely using a kernel density estimation [8],[9] to approximate the SER from training data and to derive a stochastic gradient algorithm for sample-by-sample adaptation. The resulting algorithm is therefore called the LSER. The AMBER has already been extended to adaptive linear equalizer with M-PAM symbols by Yeh and Barry [10] and the resulting algorithm is called the AMSER algorithm. Simulation is used to investigate the performance of the LSER algorithm and to compare it with the AMSER algorithm.

II. ADAPTIVE MSER LINEAR EQUALIZER

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

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

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 set $\mathcal{S} = \{s_l = 2l - M - 1,\ 1 \le l \le M\}$. The linear equalizer with a length m has the form:

$$y(k) = \mathbf{w}^T \mathbf{r}(k) \qquad (2)$$

where $\mathbf{r}(k) = [r(k)\ r(k-1) \cdots r(k-m+1)]^T$ is the observation vector and $\mathbf{w} = [w_0\ w_1 \cdots w_{m-1}]^T$ is the equalizer weight vector. The equalizer output $y(k)$ is passed to a threshold detector which provides an estimate $\hat{s}(k-d)$ of $s(k-d)$, with $0 \le d \le m + n_h - 2$ being a decision delay. The received signal vector can be expressed as:

$$\mathbf{r}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{n}(k) \qquad (3)$$

where $\mathbf{H}$ is a Toeplitz matrix, whose (i,j)th elements is h_{j-i} for $0 \le j - i \le n_h - 1$ and 0 otherwise.

Note that $\mathbf{s}(k) = [s(k)\ s(k-1) \cdots s(k-m-n_h+2)]^T$ has $N_s = M^{m+n_h-1}$ combinations, denoted as $\mathbf{s}_j$, $1 \le j \le N_s$. Let $\mathbf{r}_j = \mathbf{H}\mathbf{s}_j$ and express $y(k)$ as

$$y(k) = \mathbf{w}^T(\bar{\mathbf{r}}(k) + \mathbf{n}(k)) = \bar{y}(k) + e(k) \qquad (4)$$

where $e(k)$ is Gaussian with zero mean and variance $\mathbf{w}^T\mathbf{w}\sigma_n^2$, and $\bar{y}(k)$ takes values from the set $\mathcal{Y} = \{y_j = \mathbf{w}^T\mathbf{r}_j,\ 1 \le j \le N_s\}$. $\mathcal{Y}$ can be divided into M subsets

$$\mathcal{Y}_l = \{y_j \in \mathcal{Y} | s(k-d) = s_l\},\ 1 \le l \le M \qquad (5)$$

We assume that $\mathcal{Y}_l$ are linearly separable, which is necessary for a linear equalizer to work. Let the combined impulse response of the equalizer and channel be $\mathbf{c} = [c_0\ c_1 \cdots c_{m+n_h-2}]^T$, which is given by $\mathbf{c}^T = \mathbf{w}^T\mathbf{H} = \mathbf{w}^T[\mathbf{h}_0\ \mathbf{h}_1 \cdots \mathbf{h}_{m+n_h-2}]$, where $\mathbf{h}_i$ denotes the ith column of $\mathbf{H}$. Then $y(k)$ can be expressed as

$$y(k) = c_d s(k-d) + \sum_{i \ne d} c_i s(k-i) + e(k) \qquad (6)$$

The first term in (6) is the desired signal, and the second term the residual ISI. Thus the decision is made according to

$$\hat{s}(k-d) = \begin{cases} s_1, & \text{if } y(k) \le (s_1+1)c_d \\ s_l, & \text{if } (s_l-1)c_d < y(k) \le (s_l+1)c_d \\ & \text{for } l = 2, \cdots M-1 \\ s_M, & \text{if } y(k) > (s_M-1)c_d \end{cases} \tag{7}$$

A. Expression of symbol error rate

The conditional probability density function (p.d.f.) of $y(k)$ given $s(k-d) = s_l$ is

$$p_y(y_s|s_l) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\sum_{i=1}^{N_{sb}}\exp\left(-\frac{\left(y_s - y_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{8}$$

Where $N_{sb} = N_s/M$ and $y_i^{(l)} = \mathbf{w}^T\mathbf{r}_i^{(l)} \in \mathcal{Y}_l$. It can easily be verified that $\mathcal{Y}_{l+1} = \mathcal{Y}_l + 2c_d$. Also points in $\mathcal{Y}_l$ are symmetrically distributed around its mean. Taking into account these two properties, the SER can be expressed as

$$P_E(\mathbf{w}) = \gamma\frac{1}{N_{sb}}\sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})) \tag{9}$$

where $\gamma = \frac{2M-2}{M}$, $g_{l,i}(\mathbf{w}) = (y_i^{(l)} - c_d s_l + c_d)/\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}$ and

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-\frac{y^2}{2}\right)dy \tag{10}$$

As the SER is invariant to a scaling of $\mathbf{w}$, with a re-scaling $\mathbf{w} = \mathbf{w}/\sqrt{\mathbf{w}^T\mathbf{w}}$, the gradient of $P_E(\mathbf{w})$ is given by

$$\nabla P_E(\mathbf{w}) = \frac{\gamma}{N_{sb}\sqrt{2\pi}\sigma_n}\sum_{i=1}^{N_{sb}}\exp\left(-\frac{(y_i^{(l)} - c_d s_l + c_d)^2}{2\sigma_n^2}\right) \times \left((y_i^{(l)} - c_d s_l + c_d)\mathbf{w} - \mathbf{r}_i^{(l)} + \mathbf{h}_d s_l - \mathbf{h}_d\right) \tag{11}$$

Steepest-descent or conjugate gradient algorithms can be used to find the MSER solution that minimizes $P_E(\mathbf{w})$. Note that calculation is based on a single subset $\mathcal{Y}_l$. Computational requirements can further be simplified by considering the subset $\mathcal{Y}_l$ with $l = \frac{M}{2}+1$, which results in $s_l - 1 = 0$.

B. The least symbol-error-rate algorithm

The p.d.f. of $y(k)$ is:

$$p_y(y_s) = \frac{1}{\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{N_s}\sum_{l=1}^{M}\sum_{i=1}^{N_{sb}}\exp\left(-\frac{\left(y_s - y_i^{(l)}\right)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{12}$$

and the SER can alternatively be expressed as:

$$P_E(\mathbf{w}) = \gamma\frac{1}{N_s}\sum_{l=1}^{M}\sum_{i=1}^{N_{sb}} Q(g_{l,i}(\mathbf{w})) \tag{13}$$

Given a block of K training samples $\{\mathbf{r}(k), s(k-d)\}$, a kernel density estimate of the p.d.f. (12) is given by:

$$\hat{p}_y(y_s) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\frac{1}{K}\sum_{k=1}^{K}\exp\left(-\frac{(y_s - y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{14}$$

where the radius parameter ρ_n is related to σ_n. From this estimated p.d.f., the estimated SER expression is given by

$$\hat{P}_E(\mathbf{w}) = \gamma\frac{1}{K}\sum_{k=1}^{K} Q(\hat{g}_k(\mathbf{w})) \tag{15}$$

where $\hat{g}_k(\mathbf{w}) = (y(k) - \hat{c}_d s(k-d) + \hat{c}_d)/\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}$, $\hat{c}_d = \mathbf{w}^T\hat{\mathbf{h}}_d$, and $\hat{\mathbf{h}}_d$ an estimate of $\mathbf{h}_d$. Thus block adaptive gradient algorithms can readily be obtained by substituting $\nabla P_E(\mathbf{w})$ with $\nabla\hat{P}_E(\mathbf{w})$.

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

$$\hat{p}_y(y_s, k) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\exp\left(-\frac{(y_s - y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{16}$$

Using the instantaneous stochastic gradient with a re-scaling to ensure $\mathbf{w}^T\mathbf{w} = 1$ gives rise to the LSER algorithm:

LSER. At sample k, update the channel estimate $\hat{\mathbf{h}}_d(k)$ using the normalized LMS algorithm with an adaptive gain $\bar{\mu}$, compute $\hat{c}_d(k) = \mathbf{w}^T(k)\hat{\mathbf{h}}_d(k)$ and update the weight vector

$$\begin{aligned} \mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu\gamma}{\sqrt{2\pi}\rho_n}\times \\ \exp\left(-\frac{(y(k) - \hat{c}_d(k)s(k-d) + \hat{c}_d(k))^2}{2\rho_n^2}\right)\times \\ (\mathbf{r}(k) - (s(k-d) - 1)\hat{\mathbf{h}}_d(k) \\ - (y(k) - \hat{c}_d(k)s(k-d) + \hat{c}_d(k))\mathbf{w}(k)) \end{aligned} \tag{17}$$

where the adaptive gain μ and width ρ_n are the two algorithm parameters that need to be set appropriately. The weight vector is then re-scaled to ensure $\mathbf{w}^T(k+1)\mathbf{w}(k+1) = 1$.

III. Adaptive MSER linear-combiner DFE

The linear-combiner DFE is defined by:

$$y(k) = \mathbf{w}^T\mathbf{r}(k) + \mathbf{b}^T\hat{\mathbf{s}}_b(k) \tag{18}$$

where $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1)\cdots\hat{s}(k-d-n_b)]^T$ is the past detected symbol vector with n_b being the feedback order, and $\mathbf{b} = [b_1\cdots b_{n_b}]^T$ the feedback filter tap vector. We will choose $d = n_h - 1$, $m = n_h$ and $n_b = n_h - 1$, as this choice is sufficient to guarantee that the subsets of noise-free channel states are linearly separable [2],[3].

Define $\mathbf{s}_f(k) = [s(k)\ s(k-1) \cdots s(k-d)]^T$ and partition the CIR matrix $\mathbf{H}$ into $\mathbf{H} = [\mathbf{H}_1 \mid \mathbf{H}_2]$, where where $\mathbf{H}_1$ has a dimension of $m \times (d+1)$ and $\mathbf{H}_2$ a dimension of $m \times n_b$. The last column of $\mathbf{H}_1$ is $\mathbf{h}_d = [h_{n_h-1} \cdots h_1\ h_0]^T$. Under the assumption that the past decisions are correct, that is,

$$\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n_b)]^T \quad (19)$$

the received signal vector can be expressed as

$$\mathbf{r}(k) = \mathbf{H}_1 \mathbf{s}_f(k) + \mathbf{H}_2 \hat{\mathbf{s}}_b(k) + \mathbf{n}(k) \quad (20)$$

Thus, in the translated space defined by

$$\mathbf{r}^{'}(k) = \mathbf{r}(k) - \mathbf{H}_2 \hat{\mathbf{s}}_b(k) \quad (21)$$

the DFE (18) becomes a "linear equalizer" [2],[3]:

$$y(k) = \mathbf{w}^T \mathbf{r}^{'}(k) = \tilde{y}(k) + e(k) \quad (22)$$

The elements of $\mathbf{r}^{'}(k)$ can be computed recursively [2]:

$$\begin{cases} r^{'}(k-i) = z^{-1} r^{'}(k-i+1) - h_{n_h-i}\hat{s}(k-d-1), \\ \qquad\qquad i = m-1, \cdots, 2, 1 \\ r^{'}(k) = r(k) \end{cases} \quad (23)$$

Note that the feedback filter coefficients do not disappear. They in fact have been set to their optimal values. Denote the $N_f = M^{d+1}$ combinations of $\mathbf{s}_f(k)$ as $\mathbf{s}_{f,j}$, $1 \le j \le N_f$, and let $\tilde{\mathbf{r}}_j = \mathbf{H}_1 \mathbf{s}_{f,j}$. Obviously, $\tilde{y}(k)$ takes values from the set $\tilde{\mathcal{Y}} = \{\tilde{y}_j = \mathbf{w}^T \tilde{\mathbf{r}}_j,\ 1 \le j \le N_f\}$, which can be partitioned into M subsets

$$\tilde{\mathcal{Y}}_l = \{\tilde{y}_j \in \tilde{\mathcal{Y}} | s(k-d) = s_l\},\ \ 1 \le l \le M \quad (24)$$

Note that $\tilde{\mathcal{Y}}_l$ are always linearly separable [2],[3]. All the results for the linear equalizer case can readily be applied to the translated DFE.

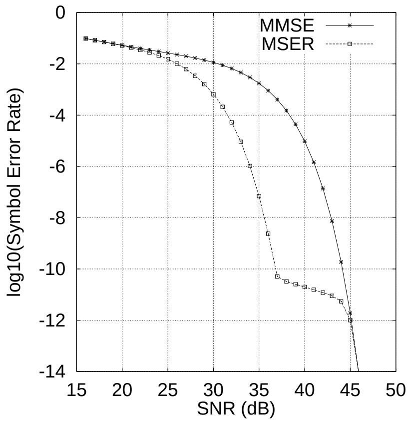
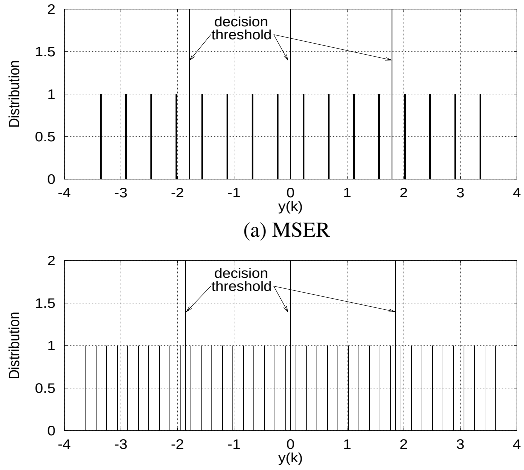


Fig. 1. Symbol error rate comparison for Example 1.



(a) MSER

(b) MMSE

Fig. 2. Distributions of the four subsets, $\mathcal{Y}_1$ to $\mathcal{Y}_4$, and the three corresponding decision thresholds for Example 1 with SNR$=$ 35 dB. A scalar point is plotted as an impulse of unit height.

IV. SIMULATION STUDY

Example 1. The transfer function of the CIR was $H(z) = 1.0 + 0.5z^{-1}$ with 4-PAM symbols. The linear equalizer had $m = 2$ and $d = 0$. The theoretical SERs of the MMSE and MSER linear equalizers are depicted in Fig. 1. The distributions of the four scalar subsets, $\mathcal{Y}_1$ to $\mathcal{Y}_4$, are shown in Fig. 2 together with the three corresponding decision thresholds, given SNR$=$ 35 dB. Since the SER is invariant to a positive scaling of the weight vector, the weight vector was scaled to a unit length, and the SER was mainly determined by the minimum distance of the scalars in a subset $\mathcal{Y}_l$ to its related decision threshold $c_d(s_l - 1)$. The MSER solution has a much larger minimum distance, as can be seen in Fig. 2. The LSER and AMSER were investigated in simulation with SNR$=$ 35 dB and the initial weight vector set to the MMSE solution. The step size μ and width ρ_n for the LSER, and the step size μ and threshold τ for the AMSER [10] were set to produce a good combined performance of convergence speed and steady-state error for the respective algorithm. For

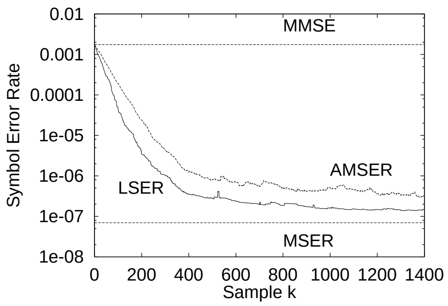


Fig. 3. Learning curves of the two stochastic gradient adaptive MSER algorithms for Example 1 with SNR$=$ 35 dB. The results were averaged over 100 runs.

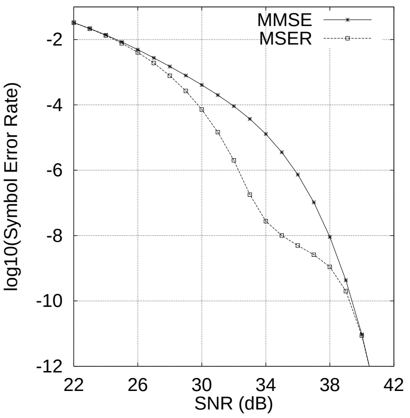
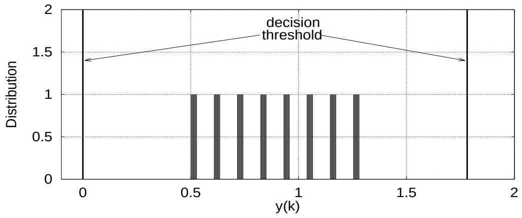


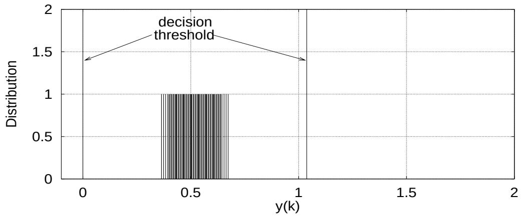
Fig. 4. Lower-bound symbol error rate comparison for Example 4, assuming correct symbols being fed back.

a fair comparison, the resulting two learning curves had similar "roughness", as shown in Fig. 3. The results in Fig. 3 were obtained by averaging over 100 runs.

Example 2. The transfer function of the CIR was $H(z) = 0.3 + 1.0z^{-1} - 0.3z^{-2}$ with 8-PAM symbols. 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. 4. The distributions of the scalar subset, $\tilde{\mathcal{Y}}_5$, together with the two related decision thresholds are shown in Fig. 5, given SNR$=$ 34 dB. In the simulation using the two stochastic gradient adaptive MSER algorithms, SNR$=$ 34 dB and the initial weight vector was set to the MMSE solution. The learning curves of these two stochastic gradient algorithms are depicted in Fig. 6, where the results were averaged over 300 runs.



(a) MSER



(b) MMSE

Fig. 5. Distributions of the scalar subset, $\tilde{\mathcal{Y}}_5$, and the two corresponding decision thresholds for Example 2 with SNR$=$ 34 dB. A scalar point is plotted as an impulse of unit height.

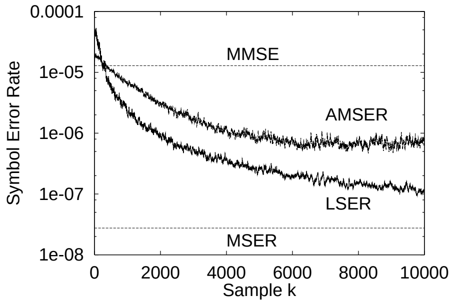


Fig. 6. Learning curves of the two stochastic gradient adaptive MSER algorithms for Example 2 with SNR$=$ 34 dB. The results were averaged over 300 runs.

V. Conclusions

The theoretical MSER solution for the linear equalizer and DFE with M-PAM signalling has been derived with the aim of deriving a stochastic gradient adaptive implementation. The resulting adaptive MSER algorithm, called the LSER, is based on an approach of kernel density estimation from the training data, which has been shown to be a natural and convenient way of approximating the density distribution of the equalizer decision variable. The LSER algorithm has a computational complexity that is linear with the equalizer length. Simulation results have shown that the proposed LSER algorithm has better performance than an existing simpler adaptive MSER algorithm called the AMSER, in terms of convergence rate and steady-state SER misadjustment.

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.

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

[4] 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.

[5] 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.

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

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

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

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

[10] 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.