THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 7, September 1979
Printed in U.S.A.

Adaptive Echo Cancellation/AGC Structures for Two-Wire, Full-Duplex Data Transmission

By D. D. FALCONER and K. H. MUELLER

(Manuscript received March 2, 1979)

Three different receiver arrangements are studied, all of which incorporate provision for joint adaptive echo cancellation and gain adjustment to provide two-wire full-duplex data communication. In each case, the canceler consists of a data-driven transversal filter, but the architectures differ in the way the gain adjustment is provided. For all architectures, we investigate the properties of a joint adaptive LMS *algorithm based on the receiver's decisions on far-end data symbols and present appropriate computer simulations. We show that an arrangement where the gain control adjusts the reference level after the decision detector output performs significantly better than* AGC *schemes attempting to adjust the level of the analog signal.*

I. INTRODUCTION

High-speed full-duplex data communication on a single channel is of immense practical interest. Data transmission via the DDD telephone network and the possibility of future digital subscriber lines are two of the most challenging applications. Techniques for achieving this goal fall in essentially three categories: Frequency Division Multiplexing (FDM), Time Division Multiplexing (TDM), and echo cancellation. Only echo cancellation allows full-bandwidth continuous use of the channel in each direction. This scheme therefore offers the highest potential bit rates.

The transmitter and receiver are jointly coupled to a two-wire line via a hybrid. In an environment of changing channel characteristics (e.g., switched network), the hybrid balancing, if fixed, will at best provide a compromise match to the channel. In this mode, a vestige of the local transmitted signal, leaking through the hybrid, can be expected to interfere with the incoming signal from the far-end simultaneously operating transmitter. Figure 1 shows the system under discussion, and Fig. 2 models the signals entering and leaving a two-wire

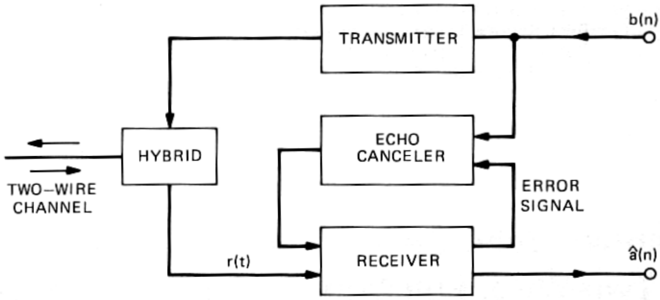


Fig. 1—Basic system configuration.

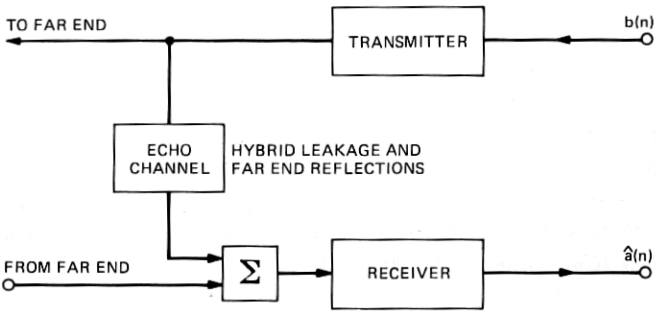


Fig. 2—System model (without canceler).

full-duplex modem. The local transmitter transmits a sequence of data symbols $\{b(n)\}$ at T-second intervals as a PAM data waveform. The received waveform $r(t)$ consists of a PAM data waveform with data symbols $\{a(n)\}$ transmitted also at T-second intervals from the *distant* end, plus noise, plus the interfering vestige of the locally transmitted signal. This interfering signal (which we shall refer to as the *echo signal* or *echo component*) may have power comparable to or even greater than that of the desired far-end signal component.

Decisions on the $\{a(n)\}$ are made by quantizing the sampled receiver output to ± 1 in the case of binary data, or to one of M values in the case of M-level data. A typically encountered echo component arising in a system with a conventional compromise balanced hybrid will cause an unacceptably high error rate.

To remove the interfering echo component, the local receiver must perform echo cancellation; that is, estimate the echo signal and subtract it from the incoming signal prior to making decisions, as shown in Figs. 1 and 3a. The estimate is a transversally filtered version of the local data symbols $\{b(n)\}$* as proposed in Ref. 1. If the $\{b(n)\}$ are

* The $\{b(n)\}$ may be different from the user data since they are defined to include such operations as differential encoding or scrambling.

binary, the implementation is simple, requiring only additions and subtractions. The transversal filter tap coefficients $\{p_m\}$ should approximate the samples of the impulse response of the combination of the local transmitter and the echo path.

Equivalently,[1] the tap coefficients should be chosen to minimize, in a mean-square sense, the measured *receiver error* signal which is the difference between the actual receiver output $y(n)$ and the ideal output. This error is available at each sampling instant of the received data. The subtraction of the (decision-directed) reference $\{\hat{a}(n)\}$ is, of course, what makes it possible to adapt quickly even in the presence of doubletalk. Such adaptation allows tracking time-varying components of the echo channel or coping with larger call-to-call variations in a switched system. However, since the level of the received signal is likely to vary significantly under those conditions, it is essential that proper scaling be done when the error signal is computed. Such scaling involves a gain adjust device which must operate jointly and adaptively with the echo canceler. This paper deals with this joint adaptation

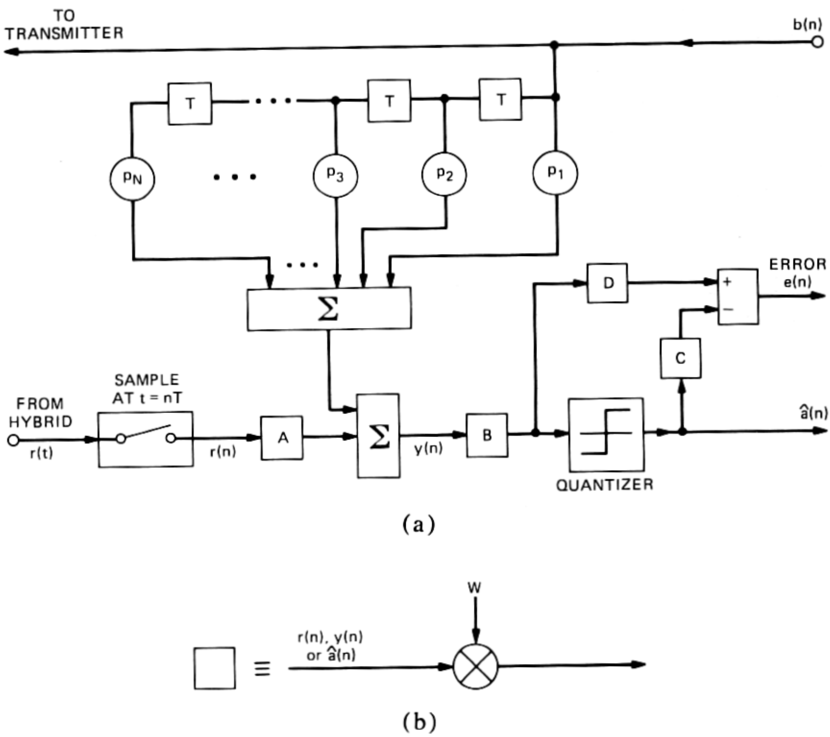


Fig. 3—Receiver structure with echo canceler and automatic gain control. (a) Overall system showing various locations for AGC function. (b) Basic AGC functions in one of above marked locations A, B, C, or D.

problem. Although we perform our study for a system operating at baseband, our results can be applied to passband structure via appropriate redefinitions.

As shown in Fig. 3a, the error $e(n)$ is the difference between the receiver's decision $\hat{a}(n)$ (assumed correct) and the receiver's output $y(n)$. But if the end-to-end channel gain is α, $y(n)$ consists of $\alpha\, a(n)$ plus possibly noise and uncanceled echo. Intersymbol interference is not treated in this study. It is expected to be of secondary concern in typical subscriber cable systems, but we realize that it cannot be neglected for high-speed DDD applications. The error $e(n)$ in the absence of gain control would contain the term $(\alpha - 1)a(n)$, which is relatively large if α differs significantly from unity. Therefore, a small steady-state mean-squared error, with consequent minimal fluctuation of the canceler tap coefficients $\{p_n\}$ and low error rate, is possible *only if the nonunity gain α is compensated for by an* AGC *adjusted to provide a gain approximating* $1/\alpha$. The gain of the AGC, denoted w, will be considered to be adjusted jointly with the echo canceler tap coefficients to minimize the mean squared error.

Shown in Fig. 3a are four possible locations, A, B, C, and D, for placement of the AGC whose functional form is depicted in Fig. 3b. Since D is identical to B as long as the signal is binary and the quantizer is ideal, only three variations need to be examined. The corresponding three receiver architectures, labeled naturally A, B, and C, may differ significantly in their adaption speeds. The main theme of this paper is the convergence of each of the three receiver arrangements, and how it is influenced by channel parameters, such as the relative powers of echo and far-end components.

The study concludes that arrangement C which features an automatic reference control (ARC) at the quantizer output offers the fastest convergence rate, together with the simplest implementation. Arrangements A and B may suffer slower rates of convergence due to coupling interaction between the AGC and echo canceler tap coefficient adaptation. However, for arrangement B, a simple step-size modification is proposed, which on the average decouples the AGC and echo canceler adaptations, thus improving its convergence rate. Readers interested in C, the "best" arrangement, may skip Sections III to V, which deal with A and B.

Mention of earlier related work is in order at this point. Adaptive echo cancellation without a jointly adapting AGC or equalizer for two-wire full duplex data communication has been treated in Refs. 1 to 4. References 1 and 4 treat essentially the echo-cancellation system proposed here, assuming α is known, and thus omit the AGC. This type of scheme, in which the echo canceler's input consists only of local data symbols, offers obvious simplicity of implementation. In Ref. 2, a voice-type canceler is investigated, and Ref. 3 discusses an external data-driven structure which cancels the entire waveform by computing compensation samples at a high enough rate.* Reference 5 summarizes work reported in Ref. 1 and 3, and discusses a further receiver arrangement comprising an adaptive echo canceler and adaptive equalizer for mitigating end-to-end linear distortion. Arrangement A considered here, the so-called "convex canceler," is a special case of this, since an AGC can be considered a one-tap linear equalizer. Reference 6 describes an echo-cancelling receiver structure incorporating decision feedback equalization. Finally, our arrangement C combined with decision feedback equalization has recently been proposed in Ref. 7.

II. SYSTEM MODELING

We shall examine the convergence properties of several adaptation strategies for the three receiver arrangements. Each follows a decision-directed approach; AGC and echo canceler parameters are adjusted once per symbol interval, based on the observed error between the unquantized receiver output and the decision $\hat{a}(n)$ (for arrangements A and B) or $w\hat{a}(n)$ (for arrangement C). For purposes of analysis, the decisions $\hat{a}(n)$ are assumed equal to the actual data symbols $a(n)$. The convergence of decision-directed adaptive receivers making small adjustments at each iteration has been found to be negligibly affected by occasional decision errors.

The analysis is based on a simple linear model of the end-to-end channel and leakage paths: the signal $r(t)$ entering the receiver from the hybrid will be written as

$$r(t) = \sum_n a(n) g_A(t - nT) + \sum_n b(n) g_B(t - nT) + \nu(t). \qquad (1)$$

The first summation represents the signal from the far end, and $g_A(t)$ is the end-to-end channel impulse response, including receiver front-end filtering. The second summation is the echo signal, and $g_B(t)$ is the impulse response of the echo path. The symbol $\nu(t)$ is a waveform of additive white noise. The symbol interval T is equal in both summations. This is tantamount to assuming that the near- and far-end transmitters are synchronized in clock frequency.[1]

Decisions $\hat{a}(n)$ are made by quantizing samples of the receiver's output to ± 1 in the case of binary data. The end-to-end channel impulse response and the phase of the receiver's sampling clock are assumed ideal, so that no intersymbol interference is present in the samples of $r(t)$; i.e. the nth sample is of the form

* In Ref. 2, the canceler input is the sampled transmitted waveform. In Ref. 3, it is the $\{b(n)\}$.

$$r(nT) = a(n)\,\alpha + \sum_k b(k)\beta_{n-k} + \nu(nT), \tag{2}$$

where $\alpha \equiv g_A(0)$, $g_A(nT) = 0$ for $n \neq 0$, and $\beta_n \equiv g_B(nT)$. In subsequent notation, we write $r(nT)$ and $\nu(nT)$ as $r(n)$ and $\nu(n)$, respectively. The binary (± 1) data symbols $\{a(n)\}$ and $\{b(n)\}$ are statistically independent. The noise samples $\nu(n)$ are assumed to be independent with zero mean and variance σ^2.

III. ARRANGEMENT A

We recall from Fig. 3a that arrangement A forms the receiver output as the sum of the echo canceler and AGC outputs. The receiver output is a linear function of the echo canceler tap coefficients and AGC gain. Thus the mean-squared error at the receiver's output is a convex quadratic function of the receiver parameters, and a simple gradient algorithm can be used with confidence to adjust the parameters jointly. As mentioned before, if the AGC is replaced by an adaptive linear equalizer, the arrangement A generalizes to the jointly adaptive echo canceler equalizer structure discussed in Ref. 5.

Given an AGC gain w and a set of N echo canceler tap coefficients $\{p_k\}_{k=1}^{N}$, the receiver's unquantized output sample $y(n)$ is

$$y(n) = w\,r(n) + \sum_{k=1}^{N} p_k b(n-k). \tag{3}$$

Define the N-dimensional vectors

$$\mathbf{p} \equiv \begin{bmatrix} p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_N \end{bmatrix} \text{ and } \mathbf{b}(n) \equiv \begin{bmatrix} b(n-1) \\ \cdot \\ \cdot \\ \cdot \\ b(n-N) \end{bmatrix},$$

and the $(N+1)$-dimensional vectors in partitioned form as

$$\mathbf{c} \equiv \begin{bmatrix} w \\ \hline \mathbf{p} \end{bmatrix} \text{ and } \mathbf{z}(n) \equiv \begin{bmatrix} r(n) \\ \hline \mathbf{b}(n) \end{bmatrix}.$$

Then (3) is written more compactly as

$$y(n) = \mathbf{c}^{\dagger}\mathbf{z}(n), \tag{4}$$

where † denotes transpose. The vector $\mathbf{c}$ is the set of receiver parameters to be adaptively adjusted, and $\mathbf{z}(n)$ is the current set of inputs stored by the receiver.

The ideal output at time nT would be $a(n)$, and the error is

$$e(n) = y(n) - a(n). \tag{5}$$

The expression for the mean-squared error is

$$\langle e(n)^2 \rangle = \mathbf{c}^\dagger A\mathbf{c} - 2\mathbf{c}^\dagger \mathbf{x} + 1, \tag{6}$$

where A is a $(N + 1)$ by $(N + 1)$ covariance matrix

$$A \equiv \langle \mathbf{z}(n)\mathbf{z}(n)^\dagger, \tag{7}$$

and $\mathbf{x}$ is an $(N + 1)$-dimensional vector

$$\mathbf{x} = \langle a(n)\mathbf{z}(n) \rangle. \tag{8}$$

By rewriting (6) as

$$\langle e(n)^2 \rangle = (\mathbf{c} - A^{-1}\mathbf{x})^\dagger A(\mathbf{c} - A^{-1}\mathbf{x}) + 1 - \mathbf{x}^\dagger A^{-1}\mathbf{x}, \tag{9}$$

and recognizing that A by definition is positive semidefinite, it is clear that the mean-squared error has its minimum value

$$e^2_{\min} = 1 - \mathbf{x}^\dagger A^{-1}\mathbf{x}, \tag{10}$$

when

$$\mathbf{c} = \mathbf{c}_{\text{opt}} = A^{-1}\mathbf{x}. \tag{11}$$

Using the independence assumptions for the data symbols and noise, and the expression (2) for $r(n)$, we readily find that A can be written as

$$A = \left[\begin{array}{c|c} A_{00} & \boldsymbol{\beta}^\dagger \\ \hline \boldsymbol{\beta} & I \end{array}\right], \tag{12}$$

where

$$A_{00} \equiv \langle r(n)^2 \rangle = \alpha^2 + \sum_k \beta_k^2 + \sigma^2, \tag{13}$$

$$\boldsymbol{\beta} \equiv \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_N \end{bmatrix} \tag{14}$$

denotes the sampled echo impulse response truncated to N samples, and I is the N-dimensional identity matrix. Note that, because of the truncation involved in defining $\boldsymbol{\beta}$, $|\boldsymbol{\beta}|^2 \le \sum_k \beta_k^2$. Similarly, the vector $\mathbf{x}$ is

$$\mathbf{x} = \left[\begin{array}{c} \alpha \\ \hline \mathbf{0} \end{array}\right], \tag{15}$$

where $\mathbf{0}$ is an N-dimensional all-zero vector.

Adaptive adjustment of the receiver parameter vector $\mathbf{c}$ can be accomplished by employing the Widrow-Hoff LMS algorithm[8] just as in adaptive mean-square equalization.[9] The current value of $\mathbf{c}$ at time nT, $\mathbf{c}(n)$, is then updated according to

$$\mathbf{c}(n+1) = \mathbf{c}(n) - \gamma\, e(n)\mathbf{z}(n), \tag{16}$$

where γ is a constant step size. The average value of the correction term $-\gamma e(n)\mathbf{z}(n)$ is proportional to the negative of the gradient of the mean-squared error with respect to $\mathbf{c}$. Expression (16) portrays the joint updating of the AGC gain and echo canceler taps:

$$w(n+1) = w(n) - \gamma e(n) r(n). \tag{17a}$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma e(n)\mathbf{b}(n). \tag{17b}$$

Adaptation of the echo canceler *alone*, according to (17b) with w fixed at 1, has been analyzed by Mueller.[1] The rate of convergence of $\mathbf{p}(n)$ to $\boldsymbol{\beta}$ for an optimum choice of γ was shown to be determined only by the number of taps, rather than by the detailed characteristics of the echo path.

The joint updating algorithm (16) resembles equalizer adaptation algorithms, whose convergence behavior has been extensively studied.[7–16] The convergence of the more general version of (16), for joint echo cancellation and equalization, was discussed in Ref. 5. These theoretical studies have rested on an untrue assumption of independence of successive equalizer or canceler contents. However, a more rigorous analysis in Ref. 16, in addition to experimental results, suggests that the independence assumption does not cause serious error. The aforementioned studies have revealed that, for a fixed step size coefficient γ, the speed of convergence is largely governed by the spread of the eigenvalues of the matrix A defined by (12). Without elaborating on the details, we can say that a ratio of maximum-to-minimum eigenvalues which is close to unity leads to relatively fast convergence, while slow convergence is associated with a maximum-to-minimum eigenvalue ratio which is much greater than unity. If the step size coefficient γ is chosen to effect a judicious compromise between speed of convergence and noise due to random tap fluctuations, then a system with a small eigenvalue spread would typically converge in a number of iterations equal to a small multiple of the number of adjustable tap coefficients.

The $N + 1$ eigenvalues of matrix A defined by (12), (13), and (14) are readily found to consist of $N - 1$ unit eigenvalues plus λ_{max} and λ_{min}, given by

$$\lambda_{\substack{max\\min}} = {}^1/_2 \Big[1 + \alpha^2 + \sigma^2 + \sum_k \beta_k^2 \pm \sqrt{(1 + \alpha^2 + \sigma^2 + \sum_k \beta_k^2)^2 - 4(\alpha^2 + \sigma^2 + \sum_k \beta_k^2 - |\boldsymbol{\beta}|^2)}\Big], \quad (18)$$

where the plus is associated with λ_{max} and the minus with λ_{min}. It is straightforward to show that

$$\lambda_{min} < 1 < \lambda_{max}.$$

It is interesting to point out that expression (18) for λ_{max} and λ_{min} coincides with the expressions for *bounds* on the maximum and minimum eigenvalues found in Ref. 5 for the more general canceler/equalizer combination. Table I shows values of λ_{max} and λ_{min} for various values of $(\alpha^2 + \sigma^2)$ (the power of the far-end signal component plus noise) and $\sum_k \beta_k^2$ (the power of the echo component), assuming $\sum_k \beta_k^2 = |\beta|^2$, so that perfect echo cancellation is possible.

The ratio $\lambda_{max}/\lambda_{min}$ increases rapidly as $\alpha^2 + \sigma^2$ decreases. Thus a system with a far-end signal component which is much weaker than the near-end echo component would be expected to converge much more slowly than one with a relatively strong far-end component. This sensitivity of the convergence behavior to the relative strengths of far-end and near-end signal components was also noted in Ref. 5 for the canceler/equalizer system.

IV. ARRANGEMENT B

In this arrangement, combining the received signal and echo canceler output is done ahead of the AGC. This may appear as a more natural arrangement, since the intent is to cancel the echo component at the sampling instants before they enter the portion of the receiver devoted to estimating $a(n)$. The local data symbols are processed by the echo canceler and AGC in tandem; the output is not a linear function of the canceler tap coefficients and AGC gain, and the mean-squared error is not a convex function of these receiver parameters. We can thus call this second arrangement a "nonconvex canceler."

The receiver output is

Table I

Far-End Signal Power Plus Noise Power $\alpha^2 + \sigma^2$	Echo Power $\lvert\beta\rvert^2$	Eigenvalue λ_{min}	Eigenvalue λ_{max}	Ratio λ_2/λ_1
0.1	1.0	0.0487	2.015	41.4
0.35	1.0	0.1598	2.190	13.7
0.50	1.0	0.2187	2.281	10.4
1.0	1.0	0.3820	2.618	6.85
2.0	1.0	0.5858	3.414	5.83
1.0	2.0	0.2680	3.732	13.9

$$y(n) = w[r(n) + \mathbf{p}^\dagger \mathbf{b}(n)]. \tag{19}$$

With $r(n)$ modeled as in eq. (2), the mean-squared error for given values of w and $\mathbf{p}$ is

$$\begin{aligned}\langle e(n)^2 \rangle &= \langle (y(n) - a(n))^2 \rangle \\ &= w^2[\alpha^2 + \sum_k \beta_k^2 + \sigma^2 + |\mathbf{p}|^2 + 2\,\mathbf{p}^\dagger \boldsymbol{\beta}] \\ &\quad - 2\,w\alpha + 1,\end{aligned} \tag{20}$$

which can be written as

$$\langle e(n)^2 \rangle = (w\alpha - 1)^2 + |\mathbf{p} + \boldsymbol{\beta}|^2 w^2 + w^2\,[\sigma^2 + \sum_k \beta_k^2 - |\boldsymbol{\beta}|^2]. \tag{21}$$

The final term in brackets represents the effect of additive noise and uncancellable echo, if any. Expression (21) can also be written as

$$\langle e(n)^2 \rangle = (\alpha^2 + \delta^2)(w - w_{\text{opt}})^2 + w^2 |\mathbf{p} - \mathbf{p}_{\text{opt}}|^2 + \frac{\delta^2}{\alpha^2 + \delta^2}, \tag{22}$$

where

$$\delta^2 \equiv \sigma^2 + \sum_k \beta_k^2 - |\boldsymbol{\beta}|^2 \tag{23a}$$

and

$$w_{\text{opt}} \equiv \frac{\alpha}{\alpha^2 + \delta^2} \tag{23b}$$

$$\mathbf{p}_{\text{opt}} = -\boldsymbol{\beta} \tag{23c}$$

are the parameter values which minimize $\langle e(n)^2 \rangle$. In practical systems, δ is small and $w_{\text{opt}} \sim 1/\alpha$.

It is instructive to plot contours of constant mean-squared error in the plane whose coordinates are the AGC error

$$e_w \equiv w - w_{\text{opt}} \tag{24}$$

and canceler error magnitude

$$|\mathbf{e}_p| = |\mathbf{p} - \mathbf{p}_{\text{opt}}|, \tag{25}$$

respectively; i.e., curves satisfying

$$\alpha^2 e_w^2 + \left(e_w + \frac{1}{\alpha}\right)^2 |\mathbf{e}_p|^2 = \text{MSE} \tag{26}$$

(assuming δ is negligible), for various positive values of MSE. Such contour plots are shown in Figs. 4 and 5 for values of the end-to-end gain α of 0.5 and 1. Each plot shows 10 contours for MSE ranging from

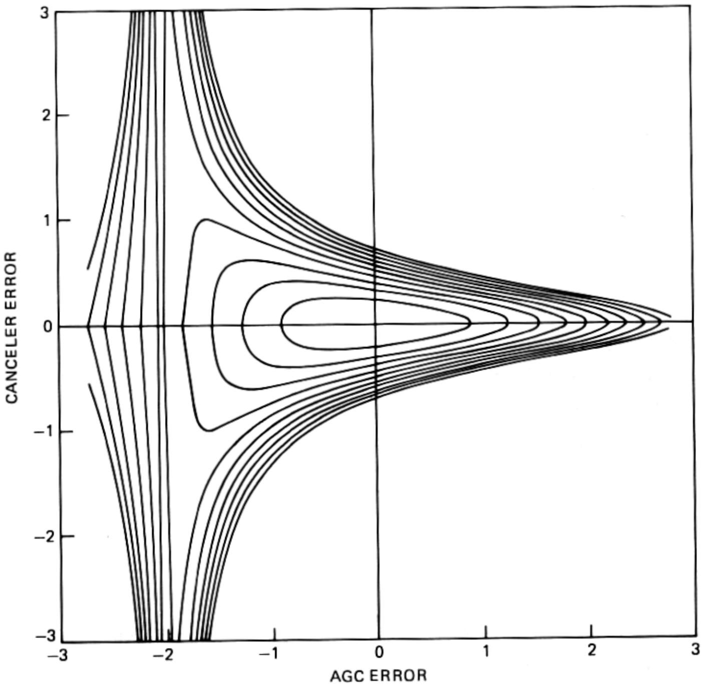


Fig. 4—Contours of constant mean-squared error for arrangement B ($\alpha = 0.5$).

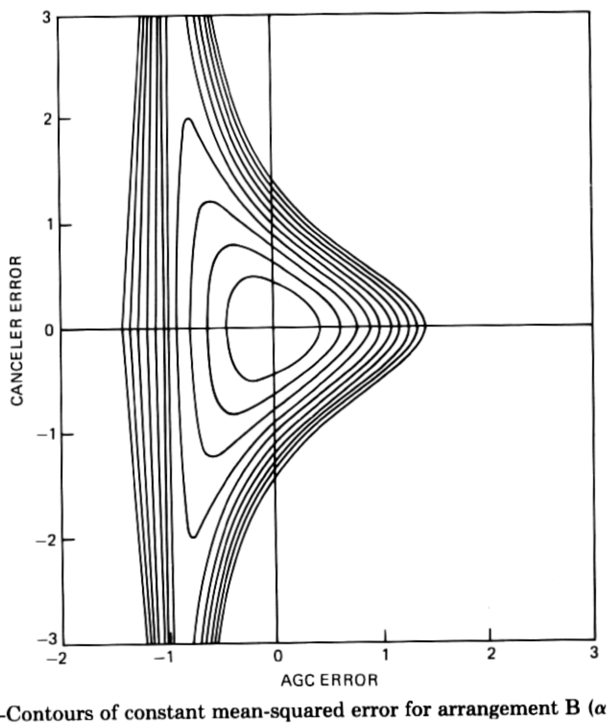


Fig. 5—Contours of constant mean-squared error for arrangement B ($\alpha = 1.0$).

0.2 to 2.0 in steps of 0.2. In each case, there is a unique minimum ($e_w = |\mathbf{e}_p| = 0$), but the MSE surface is not convex. Moreover, it exhibits a kind of trough running along the line $e_w = -1/\alpha$. The floor of this trough slopes very gradually toward the origin for large values of $|\mathbf{e}_p|$. The shape of the contours for this system differs radically from the ellipsoidal contours that are characteristic of the convex system. Those contours are plotted in Figs. 6 and 7 for the same values of MSE and of α and $|\boldsymbol{\beta}| = 1$, as for the nonconvex system. The ellipses are plotted for convenience for error values mapped onto the two eigenvectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ of the matrix A. The eccentricity of the ellipse is the ratio $\lambda_{max}/\lambda_{min}$. Similarly, the convex system would converge slowly, starting from zero-valued parameters if α is small (implying $\lambda_{max}/\lambda_{min} \gg 1$).

A gradient procedure (LMS algorithm) for jointly adjusting w and $\mathbf{p}$ should, with proper choice of step size, converge to the optimum parameters. It would, on the average, follow a path perpendicular to the contours that it crosses. Thus very slow convergence of the nonconvex system would be expected if the initial value of all parameters is zero; i.e., starting at $e_w = -(1/\alpha)$, $|\mathbf{e}_p| = |\boldsymbol{\beta}|$.

An LMS algorithm for the nonconvex system is obtained by making the correction terms proportional to the negative gradients of the

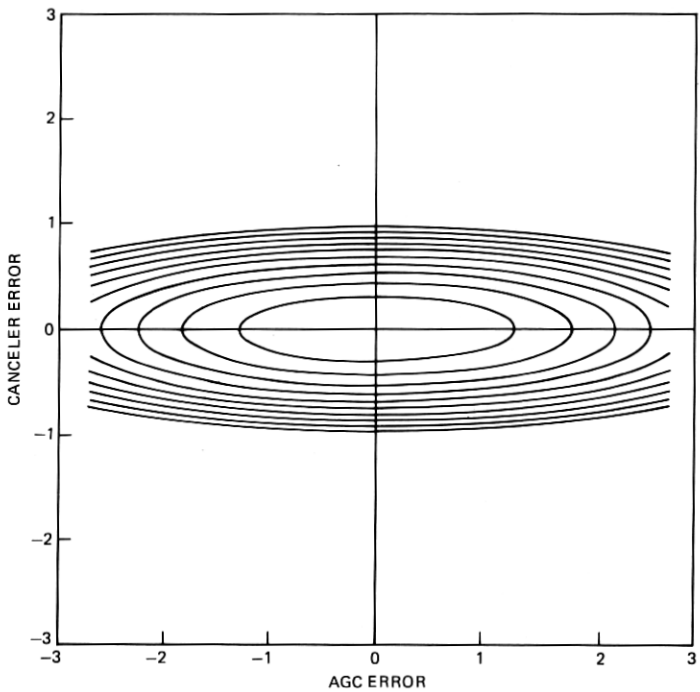


Fig. 6—Contours of constant mean-squared error for arrangement A ($\alpha = 0.5$, $|\beta| = 1$).

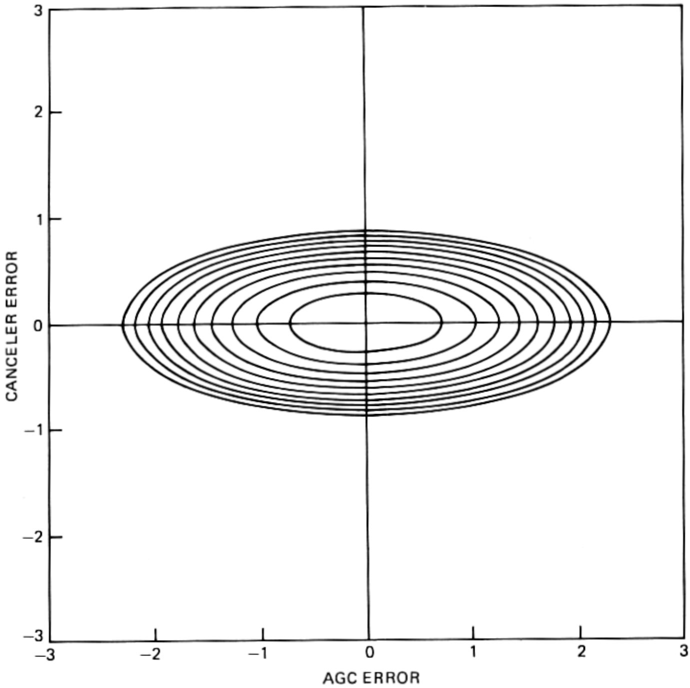


Fig. 7—Contours of constant mean-squared error for arrangement A ($\alpha = 1$, $|\beta| = 1$).

squared error. The resulting joint AGC/canceler updating algorithm would then be

$$w(n+1) = w(n) - \gamma e(n)(r(n) + \mathbf{b}(n)^{\dagger}\mathbf{p}(n)) \tag{27a}$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma e(n) w(n) \mathbf{b}(n), \tag{27b}$$

where

$$e(n) = w(n)(r(n) + \mathbf{b}(n)^{\dagger}\mathbf{p}(n)) - a(n). \tag{27c}$$

V. REDUCING COUPLING EFFECTS ON THE NONCONVEX SYSTEM'S ADAPTATION

Combining (27b) and (27c) results in the following algorithm for updating $\mathbf{p}(n)$:

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma w(n)^2 \mathbf{b}(n)[r(n) + \mathbf{b}(n)^{\dagger}\mathbf{p}(n)] + \gamma w(n) a(n) \mathbf{b}(n). \tag{28}$$

It would be desirable to reduce or eliminate the coupling effect of $w(n)$ in this algorithm. *Consider modifying the algorithm by replacing the* *constant step size* γ *with* $\gamma/w(n)^2$, where the second γ is a constant. Then (28) becomes

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma\mathbf{b}(n)[r(n) + \mathbf{b}(n)^\dagger\mathbf{p}(n)] + \frac{\gamma}{w(n)} a(n)\mathbf{b}(n). \quad (29)$$

The *average* value of the correction term for fixed $\mathbf{p}(n)$ is then $-\gamma[\beta + \mathbf{p}(n)]$, *which is the same as the average correction term for an adaptive echo canceler with no* AGC. Thus by the simple choice of step size $\gamma/w(n)^2$, we can *on the average* remove the influence of AGC adaptation on echo canceler adaptation. The only remaining coupling stems from the term $[\gamma/w(n)]a(n)\mathbf{b}(n)$, whose mean value is zero. The mean of the correction term in (27a) for updating the AGC gain is

$$-\gamma\langle e(n)(r(n) + \mathbf{b}(n)^\dagger\mathbf{p}(n))\rangle$$
$$= -\gamma\alpha(w(n)\alpha - 1) - \gamma w(n)\,|\,\mathbf{p}(n) + \boldsymbol{\beta}\,|^2$$
$$-\gamma w(n)(\sigma^2 + \sum_k \beta_k^2 - |\,\boldsymbol{\beta}\,|^2), \quad (30)$$

in which coupling from the echo canceler adaptation is evident in the middle term $\gamma w(n)\,|\,\mathbf{p}(n) + \boldsymbol{\beta}\,|^2$. There is no simple way to eliminate this coupling, apart from observing that $\mathbf{p}(n) + \boldsymbol{\beta}$ is the error in the echo canceler's tap coefficients, which eventually dies away.

In summary, we propose the following algorithm for jointly updating the AGC gain $w(n)$ and the echo canceler tap coefficient vector $\mathbf{p}(n)$:

$$w(n+1) = w(n) - \gamma_1 e(n)(r(n) + \mathbf{b}(n)^\dagger\mathbf{p}(n)) \quad (31a)$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \frac{\gamma_2}{w(n)} e(n)\mathbf{b}(n), \quad (31b)$$

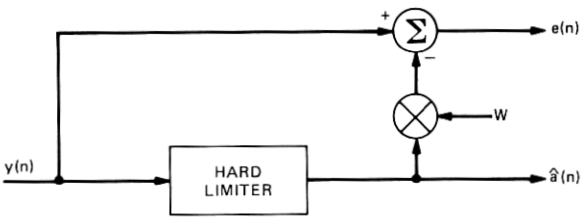
where γ_1 and γ_2 may be different constants. Algorithm (31b) is, on the average, uncoupled from (31a) and equivalent to the echo canceler algorithm operating in solitude with step size γ_2. The latter algorithm has been examined by Mueller,[1] who determined an optimum step size γ equal to the reciprocal of the number of taps, and demonstrated favorable convergence characteristics, independent of echo path characteristics. The choice of γ_1 would best be made by experiment. While the above decoupling modification was only heuristically motivated, the simulations reported in Section VI confirm its usefulness.

VI. ARRANGEMENT C

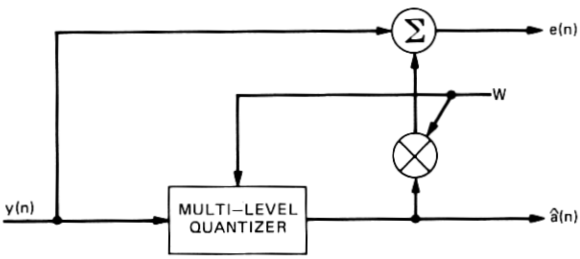
Arrangement C, shown in Figs. 3a and 8a, simply omits the AGC for purposes of making a decision on the binary symbol $a(n)$; the quantizer input, which is the algebraic sum of the channel output sample and the echo canceler output, is hardlimited to ± 1. Note that, if the $\{a(n)\}$ are binary symmetric ($a(n) = \pm 1$) data symbols, then the attenuation of the end-to-end channel is irrelevant and no explicit AGC is necessary for making a decision. This comment also applies to other baseband data symbol formats such as diphase, and also to phase-modulated signals, in which case the data symbols $a(n)$ are numbers lying on the unit circle in the complex plane.

To enable adaptive adjustment of the echo canceler in arrangement C, the *receiver's decision* $\hat{a}(n)$ is scaled by an adjustable coefficient w before being subtracted from the unquantized output to form the error which is used to update the tap coefficients. The coefficient w thus has the role of an automatic reference control (ARC), rather than an AGC, since it adjusts the receiver's reference signal to a level commensurate with the attenuation of the end-to-end channel. It is adjusted jointly with the echo canceler tap coefficients. Moreover, it multiplies a discrete-valued data symbol, not a continuous-valued channel output, and so digital implementation of the receiver is simplified.

This receiver arrangement can also be applied to multiamplitude data formats as shown in Fig. 8b. In the multiamplitude case, the quantizer compares the analog signal with reference levels which require proper scaling in relation to its amplitude. The quantity w provides this information and can thus directly serve as a reference input to the quantizer as shown in Fig. 8b. A less attractive alternative



(a)



(b)

Fig. 8—Decision-making and automatic reference control in arrangement C. (a) Case of binary data symbols. (b) Case of multilevel data symbols.

would be to multiply the input signal by $1/w$ to bring it to a constant level (as with the AGC schemes) and then apply it to a quantizer with a fixed reference.

Given a set of N echo canceler tap coefficients $\{p_k\}^N$, the receiver's output sample $y(n)$, which will subsequently be quantized to form the output data symbol $\hat{a}(n)$, is

$$y(n) = r(n) + \sum_{k=1}^{N} p_k b(n-k). \tag{32}$$

Ideally, the coefficients $\{p_k\}_{k=1}^{N}$ should approximate the negatives of the echo channel's impulse response samples $\{\beta_k\}_{k=1}^{N}$, so that the echo is canceled and $y(n)$ consists of $\alpha a(n)$ plus noise.

The desired value of $y(n)$ *is* $wa(n)$, where w is an estimate of the end-to-end channel gain α. The error between the actual and desired receiver outputs, measured at the nth symbol intervals, is then

$$e(n) \equiv r(n) + \sum_{k=1}^{N} p_k b(n-k) - wa(n). \tag{33}$$

The ARC parameter w implicitly assumes the role of an AGC, multiplying the desired receiver output data symbol rather than the receiver input.* The parameters $\{p_k\}$ and w are to be jointly adjusted with the aim of minimizing the mean-squared value of $e(n)$.

For notational compactness, define the N-dimensional vectors

$$\mathbf{p} \equiv [p_1, p_2, \cdots, p_N]^\dagger \tag{34a}$$

and

$$\mathbf{b}(n) \equiv [b(n-1), b(n-2), \cdots, b(n-N)]^\dagger, \tag{34b}$$

and the $(N+1)$-dimensional partitioned vectors

$$\mathbf{c} \equiv [\mathbf{p} \,\vdots\, -w]^\dagger \tag{35a}$$

and

$$\mathbf{u}(n) \equiv [\mathbf{b}(n) \,\vdots\, a(n)]^\dagger. \tag{35b}$$

Then the unquantized receiver output is written

$$y(n) = r(n) + \mathbf{p}^\dagger \mathbf{b}(n), \tag{36}$$

and the error is

$$e(n) = r(n) - \mathbf{c}^\dagger \mathbf{u}(n). \tag{37}$$

Squaring both sides of (37) and taking the expectation, we find that the mean-squared error is

---

* In arrangements A and B, the ideal value of w equals $1/\alpha$; in arrangement C, this value equals α.

$$\langle e(n)^2 \rangle = \sigma^2 + |\boldsymbol{\epsilon}|^2, \tag{38}$$

where $\boldsymbol{\epsilon} \equiv \mathbf{c} - \mathbf{s}$ and

$$\mathbf{s} \equiv \langle r(n)\mathbf{u}(n) \rangle = \begin{pmatrix} \boldsymbol{\beta} \\ \text{---} \\ \alpha \end{pmatrix}. \tag{39}$$

The minimum mean-squared error is σ^2, and the *excess* mean-squared error at time nT is defined as $|\boldsymbol{\epsilon}(n)|^2$, where

$$\boldsymbol{\epsilon}(n) \equiv \mathbf{c}(n) - \mathbf{s} \tag{40}$$

is the difference between the tap coefficient vector $\mathbf{c}(n)$ at time nT and its optimum value $\mathbf{s}$.

A simple gradient algorithm for updating $\mathbf{c}(n)$ is

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \gamma e(n)\mathbf{u}(n), \tag{41}$$

where γ is a constant. (Note from expression (37) that $e(n)\mathbf{u}(n)$ is proportional to the gradient of $e(n)^2$ with respect to $\mathbf{c}$.) We shall examine the convergence of the excess mean-squared error $\langle |\boldsymbol{\epsilon}(n)|^2 \rangle$, as in earlier studies, assuming that successive input vectors $\mathbf{u}(n)$ are statistically independent. Subtracting $\mathbf{s}$ from both sides of (41), we have

$$\boldsymbol{\epsilon}(n+1) = \boldsymbol{\epsilon}(n) + \gamma e(n)\mathbf{u}(n). \tag{42}$$

Since the $N+1$ components of $\mathbf{u}(n)$ are ± 1,

$$|\mathbf{u}(n)|^2 = N + 1. \tag{43}$$

Also the average of the inner product

$$\begin{aligned} \langle \boldsymbol{\epsilon}(n)^\dagger \mathbf{u}(n) e(n) \rangle &= \langle \boldsymbol{\epsilon}(n)^\dagger \mathbf{u}(n)(r(n) - \mathbf{u}(n)^\dagger \mathbf{c}(n)) \rangle \\ &= \langle \boldsymbol{\epsilon}(n)^\dagger (\mathbf{s} - \mathbf{c}(n)) \rangle \\ &= -\langle |\boldsymbol{\epsilon}(n)|^2 \rangle, \end{aligned} \tag{44}$$

where we have used the assumption that $\mathbf{u}(n)$ is independent of $\mathbf{u}(n-1)$, and therefore of $\mathbf{c}(n)$. Using (38), (43), and (44), after squaring and averaging both sides of (42) we get an equation describing the evolution of the excess mean-squared error

$$\langle |\boldsymbol{\epsilon}(n+1)|^2 \rangle = \langle |\boldsymbol{\epsilon}(n)|^2 \rangle [1 - 2\gamma + \gamma^2(N+1)] + \gamma^2(N+1)\sigma^2. \tag{45}$$

This expression is readily iterated to yield

$$\langle |\boldsymbol{\epsilon}(n)|^2 \rangle = [1 - 2\gamma + \gamma^2(N+1)]^n \langle |\boldsymbol{\epsilon}(0)|^2 \rangle + \frac{\gamma(N+1)\sigma^2[1 - (1 - 2\gamma + \gamma^2(N+1))^n]}{2 - \gamma(N+1)}. \tag{46}$$

To guarantee convergence, the constant γ must be such that

$$0 < \gamma < \frac{2}{N + 1}.$$

Then the first term in (46) is a transient, which eventually decays to zero and the steady-state value $\lim_{n \to \infty} \langle |\epsilon(n)|^2 \rangle$ is

$$\frac{\gamma(N + 1)\sigma^2}{2 - \gamma(N + 1)}. \tag{47}$$

The steady-state total mean-squared error is this plus σ^2, or

$$\lim_{n \to \infty} \langle e(n)^2 \rangle = \frac{2\sigma^2}{2 - \gamma(N + 1)}. \tag{48}$$

Expressions (46) and (47), describing the evolution of the excess mean-squared error of arrangement C, are of the same form as those derived in Ref. 1 for the echo canceler alone, with no AGC. The only difference is that $(N + 1)$ replaces N (as a result of the extra AGC coefficient w). Note that the convergence rate of $\langle e(n)^2 \rangle$ is independent of all channel parameters, in contrast to the dependence of the con-

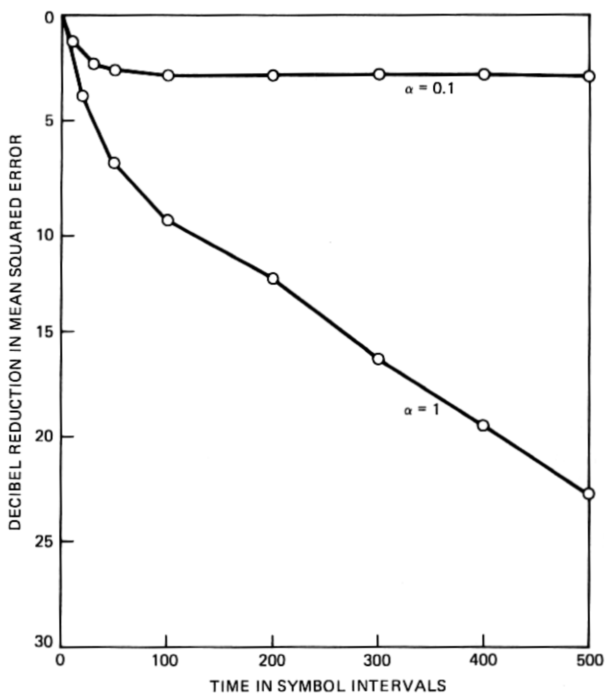


Fig. 9—Reduction of mean-squared error. Arrangement A: $\omega_0 = 1$, $\gamma = 0.01$, 16 taps.

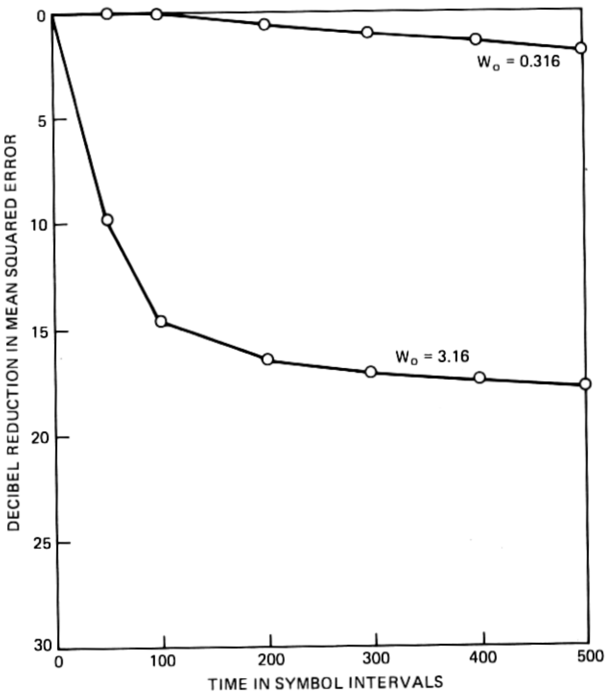


Fig. 10—Reduction of mean-squared error. Arrangement A: $\alpha = 0.316$, $\gamma = 0.01$, 16 taps.

vergence rates of arrangements A and B on the relative echo and far-end signal powers.

The choice of the adaptation coefficient γ reflects a compromise between fast adaptation and small steady-state mean-squared error. A suitable choice, proposed in Refs. 1 and 11, is

$$\gamma = \frac{1}{N+1},$$

which yields a steady-state mean-squared error of twice the minimum mean-squared error σ^2. With this choice of γ substituted in (46), we find that the excess mean-squared error decreases toward its minimum value (47) at a rate of $4.34/(N+1)$ dB per adjustment.

VII. SIMULATION

Computer simulations afforded a comparison of the actual convergence behavior of the three arrangements. Results of these simulations are shown in Figs. 9 through 15. In each case, the same echo channel is used in combination with a 16-tap canceler and $\gamma = 0.01$. The echo power is normalized to unity, and the sampled echo response is

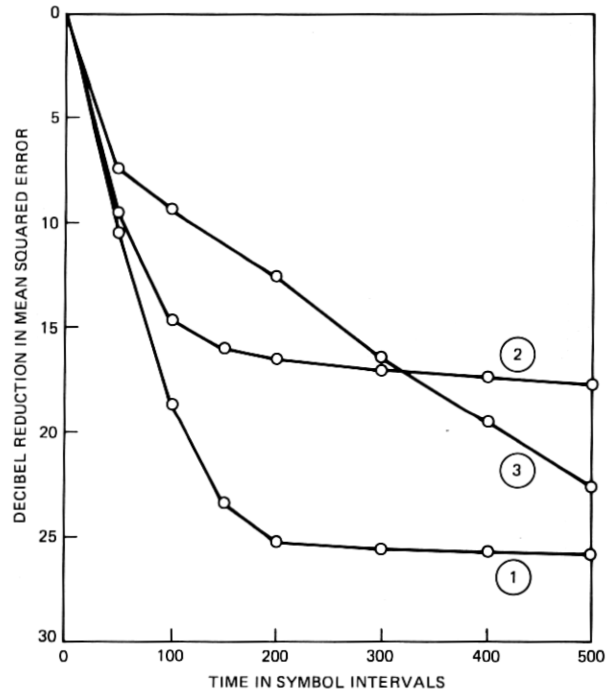


Fig. 11—Reduction of mean-squared error. Arrangement A:

① $\alpha = 0.1, \omega_0 = 10$
② $\alpha = 0.316, \omega_0 = 3.16$
③ $\alpha = 1, \omega_0 = 1$

Initial value of AGC is $1/\alpha$; $\gamma = 0.01$, 16 taps.

truncated after 16 samples, so that perfect cancellation could be obtained via a set of proper coefficients. For the nonconvex arrangement B, only the decoupled updating algorithm (31) has been used since some initial runs without this modification showed convergence problems for a variety of parameter choices.

In both arrangements A and B, a uniform, nice exponential convergence seems to be the exception rather than the rule. For weak received signals, the mean-squared error often initially reduces rapidly up to a certain point, after which convergence can become very slow. This problem appears to exist even in the case where the AGC is preset to the reciprocal of the received signal. One explanation for this behavior is that, in those cases where the initial mean-squared error is dominantly caused by a misadjustment in only one loop, the overall behavior at first approximates that of a system where either only w *or* $\mathbf{p}$ is the only parameter to be adjusted. This provides a relatively fast reduction of the initial error to the point where a joint improvement

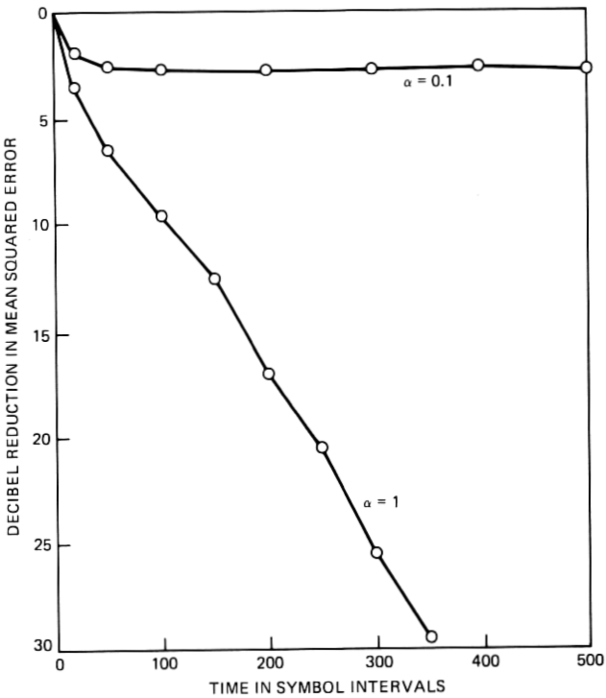


Fig. 12—Reduction of mean-squared error. Arrangement B: $\omega_0 = 1$, $\alpha = 0.01$, 16 taps.

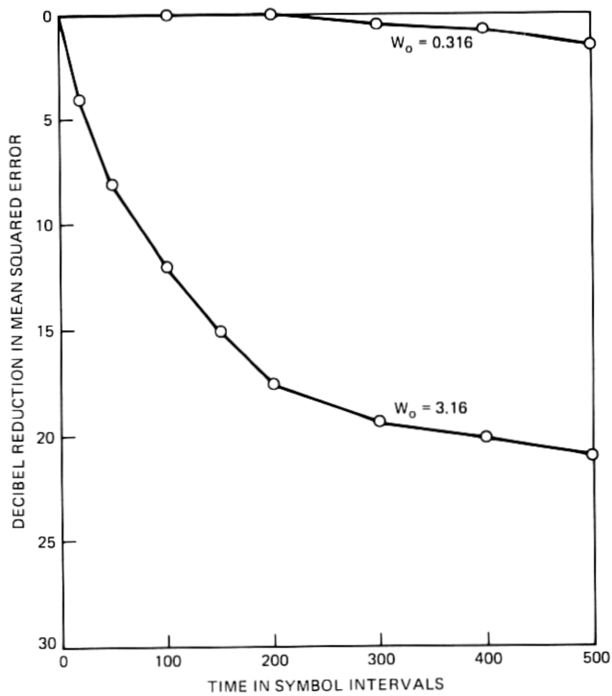


Fig. 13–Reduction of mean-squared error. Arrangement B: $\alpha = 0.316$, $\gamma = 0.01$, 16 taps.

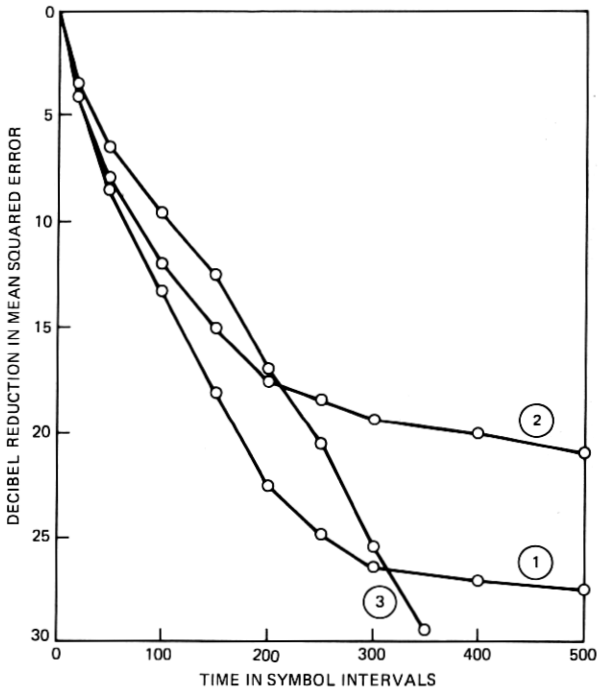


Fig. 14—Reduction of mean-squared error. Arrangement B:

① $\alpha = 0.1,\ \omega_0 = 10$
② $\alpha = 0.316,\ \omega_0 = 3.16$
③ $\alpha = 1,\ \omega_0 = 1$
} Initial value of AGC is $1/\alpha$; $\gamma = 0.01$, 16 taps.

of w *and* **p** is required. Once this situation is realized, a much slower convergence rate is expected (as we have pointed out in earlier sections), in particular with weak received signals. In the latter case, the value of w must become large and any inaccuracies in **p** are magnified.

Although we do not fully understand the dynamics of these systems, we feel it is worthwhile to present these results to point out the inherent problems to communication system designers. Further work would be required to provide a more thorough insight into these systems, but it is our feeling that this may be of more academic interest since our analysis has already shown that arrangement C provides as attractive a solution to the problem as one could possibly wish. This is demonstrated in Fig. 15. Only a single curve is presented, but actual simulations have shown that channel attenuation and initial ARC value can be varied by orders of magnitude and all such resulting curves would be essentially identical to the one shown in Fig. 15. The fact that simulations tend to deliver more rapid convergence than theoret-

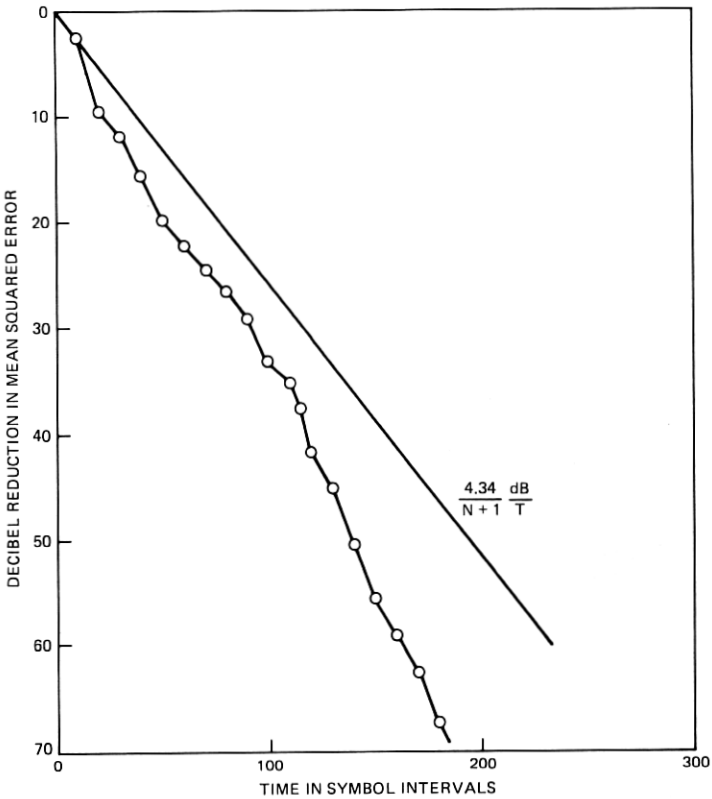


Fig. 15—Reduction of mean-squared error. Arrangement C: 16 taps, $\gamma = 1/(N+1)$.

ically predicted is, of course, due to the programmed "ideal randomness" of the data sequence used in these simulations, whereas the analytic result is an average which includes many sequences that will not provide convergence at all (e.g., constant zeros, ones, or a dotting pattern).

VIII. DISCUSSION AND SUMMARY

Three arrangements of joint adaptive echo cancellation and gain control for full duplex data transmission have been examined. Each corresponds to a different AGC location. The third, arrangement C, has proven superior in two respects:

(*i*) Its convergence rate depends only on the adaptation coefficient and on the number of adjustable tap coefficients. On the other hand, arrangement A suffers slower convergence as the ratio of echo power to distant signal power increases. Arrangement B's echo canceler can, by proper choice of adaptation coefficient, on the average be made to converge at a rate independent of channel parameters, but the convergence of its gain control still depends on the gain of the forward channel.

(*ii*) Arrangement C offers simpler digital implementation; the gain w multiplies a data symbol, instead of a finely quantized channel output or receiver output. Multiplication in arrangement C becomes addition or subtraction in the case of binary symbols.

Although the foregoing analyses presuppose binary data symbols, each receiver arrangement accommodates multilevel symbols.

Severe intersymbol interference may necessitate forward equalization. If so, arrangement A with the single coefficient $w(n)$ replaced by an adaptive transversal filter is necessary. This case is discussed in Ref. 5. The favorable convergence properties and hardware simplicity of arrangement C may be retained if adaptive decision feedback equalization with a fixed forward equalizer is used.[6,7]

REFERENCES

1. K. H. Mueller, "A New Digital Echo Canceller for Two-Wire Full-Duplex Data Transmission," IEEE Trans. Commun., *COM-24,* No. 9 (September 1976), pp. 956–962.
2. V. G. Koll and S. B. Weinstein, "Simultaneous Two-Way Data Transmission over a Two-Wire Circuit," IEEE Trans. Commun., *COM-21,* No. 2 (February 1973), pp. 143–147.
3. S. B. Weinstein, "A Passband Data-Driven Echo Canceller for Full-Duplex Transmission on Two-Wire Circuits," IEEE Trans. Commun., *COM-25*, No. 7 (July 1977), pp. 654–666.
4. H. C. Van den Elzen, P. J. Van Gerwen, and W. A. M. Snijders, "Echo Cancellation in Two-Wire Full Duplex Data Transmission with Bipolar Encoding," Proceedings of National Telecommunications Conference, Dallas, December 1976.
5. D. D. Falconer, K. H. Mueller, and S. B. Weinstein, "Echo Cancellation Techniques for Full-Duplex Data Transmission on Two-Wire Lines," Proceedings of National Telecommunications Conference, Dallas, December 1976.
6. K. H. Mueller, "Combining Echo Cancellation and Decision Feedback Equalization," B.S.T.J., *58,* No. 2 (February 1979), pp. 491–500.
7. P. T. Nielsen and M. W. Gram, "A Digital Hybrid for Two-Wire Digital Subscriber Loops," Proceedings of National Telecommunications Conference, Birmingham, December 1978.
8. B. Widrow and M. Hoff, Jr., "Adaptive Switching Circuits," in IRE WESCON Conv. Rec., Pt. 4, 1960, pp. 96–104.
9. A. Gersho, "Adaptive Equalization of Highly Dispersive Channels for Data Transmission," B.S.T.J. *48,* No. 1 (January 1969), pp. 55–70.
10. R. W. Chang, "A New Equalizer for Fast Start-Up Digital Communication," B.S.T.J., *50,* No. 6 (July–August 1971), pp. 1969–2014.
11. R. D. Gitlin, J. Mazo, and M. G. Taylor, "On the Design of Gradient Algorithms for Digitally-Implemented Adaptive Filters," IEEE Trans. on Circuit Theory, *CT-20,* No. 2 (March 1973), pp. 125–136.
12. G. Ungerboeck, "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communications," IBM J. Res. Devel., November 1972, pp. 546–555.
13. K. H. Mueller and D. A. Spaulding, "Cyclic Equalization—A New Rapidly Converging Equalization Technique for Synchronous Data Communication," B.S.T.J., *54,* No. 2 (February 1975), pp. 369–406.
14. J. G. Proakis, "Advances in Equalization for Intersymbol Interference," in *Advances in Communication Systems*, A. J. Viterbi, ed., New York: Academic Press, 1975.
15. K. H. Mueller, "A New, Fast-Converging Mean-Square Algorithm for Adaptive Equalizers with Partial-Response Signaling," B.S.T.J., *54,* No. 1 (January 1975), pp. 143–153.
16. J. E. Mazo, "On the Independence Theory of Equalizer Convergence," B.S.T.J., *58,* No. 5 (May–June 1979), pp. 963–993.