Copyright © 1975 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 1, January 1975
Printed in U.S.A.

A New, Fast-Converging Mean-Square Algorithm for Adaptive Equalizers With Partial-Response Signaling

By K. H. MUELLER

(Manuscript received December 5, 1972)

A new, generalized mean-square algorithm is presented to adjust the taps of a transversal adaptive equalizer. Any knowledge of the channel or the signaling format can be taken into account and will speed up the convergence process. The main applications are seen in partial-response signaling, where the new algorithm eliminates the interaction between the individual tap increments. This is achieved by decorrelating the components of the gradient in a fixed weighting matrix prior to the adjustments. Convergence is then extremely fast. If the channel has only phase distortion, a single iteration is sufficient to obtain the optimum tap vector (for any timing and carrier phase). This is verified by computer simulations. Finally, the new equalizer is compared with another structure recently proposed, and some advantages of the new system with regard to implementation and flexibility are pointed out.

I. INTRODUCTION

Together with sophisticated modulation techniques and bandwidth-conserving signal designs, automatic transversal equalizers are the prime characteristics of the new generation of high-speed data modems that have appeared on the market in the past few years. These modems allow data rates in the range of 3600 to 9600 b/s to be transmitted over voice-band channels. In addition to point-to-point transmission between computers and data processing centers, a large demand for high-speed data sets exists in multiparty polling systems. Because a message in such a system may consist of only a few hundred bits, it is essential that the start-up time of the modem be very short. Preferably, the modem start-up time should waste fewer bits than are contained in an average message to guarantee a reasonable system throughput. This would basically imply that start-up time should be inversely proportional to the data rate; this is a very contradictory requirement, since more accurate and complex operations are usually required during start-up of a high-speed modem than with a low-speed version. Timing recovery, carrier recovery (if required), and training of the automatic equalizer are the most important start-up operations. The time that must be allowed for the equalizer training is usually the major delay in start-up; thus, it seems worthwhile to concentrate some of our efforts in this direction.

During the past few years, much attention has been given to partial response signaling.[1–6] Some of these waveforms are particularly attractive for data transmission over band-limited channels, since they have a spectral zero at the Nyquist frequency. These signals can easily be generated with virtually zero excess bandwidth, and yet their time response decays faster than that of a sin $(x)/x$ pulse. These advantages result from a controlled amount of intersymbol interference in these signals, which increases the number of levels in the baseband eye pattern. In some formats, the impulse response has odd symmetry and the spectrum is zero at dc (which is of particular interest if SSB modulation is used).[5] If a traditional transversal automatic equalizer with an MS (mean-square) gradient algorithm[6,7] is used for partial response signals, the initial convergence will be very slow. Chang has shown that the spread of the eigenvalues of the signal autocorrelation matrix can become very large for Class IV partial response signals, and he proposed a new equalizer structure that eliminates this problem and thus can give fast convergence.[8] Of course, with the appropriate modification of some parameters, his conclusions can be extended to other partial-response signals as well.

In this paper, we propose an alternate solution to the problem of fast partial-response signal equalization. Instead of changing the equalizer structure, we propose to modify the tap updating algorithm. The coefficients associated with the new algorithm are given by a simple matrix inversion.

The advantages of the new equalizer structure are not limited to partial-response signaling. Any knowledge of the channel or the signaling format can be used to speed up the convergence process. The classic transversal equalizer evolves as a special case (Nyquist signaling) of the more general configuration that is proposed in this paper.

II. A GENERAL MEAN-SQUARE ALGORITHM

Let us assume a transversal equalizer structure with N taps, with a tap signal input vector $\mathbf{x}_k$ and a tap coefficient vector $\mathbf{c}$ at time $t_0 + kT$. The equalizer output y_k is given by

$$y_k = \mathbf{x}_k^T \mathbf{c} \qquad (1)$$

and will not, in general, be exactly equal to the desired reference value d_k, the error being

$$e_k = y_k - d_k. \tag{2}$$

The MS error that appears at the output will thus be

$$\epsilon^2 = E\{e_k^2\}, \tag{3}$$

and we now choose a tap-updating algorithm that adjusts $\mathbf{c}$ in such a way that the MS error (3) is minimized. The gradient of ϵ^2 with respect to $\mathbf{c}$ can easily be obtained by combining (1) to (3),

$$\mathbf{g} = \frac{1}{2}\frac{\partial \epsilon^2}{\partial \mathbf{c}} = E\{e_k \mathbf{x}_k\} = A\mathbf{c} - \mathbf{v}, \tag{4}$$

where we have introduced the signal autocorrelation matrix,

$$A = E\{\mathbf{x}_k \mathbf{x}_k^T\}, \tag{5}$$

and the signal correlation vector,

$$\mathbf{v} = E\{\mathbf{x}_k d_k\}. \tag{6}$$

The optimum tap vector in the mean-square sense that minimizes (3) is obtained when we set the gradient to zero,

$$\mathbf{c}_{\text{opt}} = A^{-1}\mathbf{v}. \tag{7}$$

We propose a tap-updating algorithm of the form

$$\mathbf{c}_{m+1} = \mathbf{c}_m - Q_m \mathbf{g}_m, \tag{8}$$

where $\mathbf{g}_m$ is the gradient evaluated with tap-vector $\mathbf{c}_m$ after the mth iteration, and Q_m is a nonsingular matrix, which is discussed below. Obviously, an algorithm of the general form (8) will stop updating (turn itself off) when $\mathbf{c} = \mathbf{c}_{\text{opt}}$. If the algorithm converges at all, it will thus converge to the MS solution (7). The traditional steepest-descent gradient algorithm is obtained as a special case from (8) if we set $Q_m = \beta I$ (or a more general diagonal matrix instead of the unity matrix I if tapering should be included).

The algorithm (8) can be written in the form

$$\mathbf{c}_{m+1} = (I - Q_m A)\mathbf{c}_m + Q_m \mathbf{v}. \tag{9}$$

If we introduce the tap error vector

$$\boldsymbol{\phi}_m = \mathbf{c}_m - \mathbf{c}_{\text{opt}}, \tag{10}$$

it is easy to show that

$$\boldsymbol{\phi}_{m+1} = (I - Q_m A)\boldsymbol{\phi}_m = \prod_{n=1}^{m} (I - Q_m A)\boldsymbol{\phi}_1. \tag{11}$$

Obviously, the convergence can be controlled to some degree by the selection of Q_m, since this choice affects the eigenvalues of $I - Q_mA$. This is discussed in Section III.

III. SELECTION OF THE Q MATRIX

First we will make some general remarks without specifying that the signal format be partial response. If the channel characteristics, demodulating carrier phase, and timing instant were perfectly known, we would select $Q = Q_m = A^{-1}$ and then obtain $\mathbf{c}_{\text{opt}}$ after the first iteration. Actually, if that knowledge were available, it would, of course, be easier to preset the taps to $\mathbf{c}_{\text{opt}}$ or to introduce a suitable fixed compensation network. Unfortunately, the transmission channel is not completely known. If, however, the spread within these channels is not extreme, we can reduce the average equalizer training time if we select a good estimate of the optimum tap vector for initial presetting. For example,

$$\mathbf{c}_1 = E\{A_i^{-1}\mathbf{v}_i\}, \tag{12}$$

where the expected value is taken over the ensemble of channels, and a similar estimate for Q,

$$Q_m = \beta_m[E\{A_i\}]^{-1}. \tag{13}$$

For a particular channel with

$$A_i = E\{A_i\} - \Delta_i, \tag{14}$$

the critical matrix $B_i = I - Q_mA_i$ then becomes

$$B_i = (1 - \beta_m)I + \beta_m[E\{A_i\}]^{-1}\Delta_i. \tag{15}$$

If we select $\beta_m = 1$, the average B_i matrix is equal to the zero matrix. For a randomly selected channel, the B_i matrix will, on the average, be much closer to the zero matrix than with a gradient algorithm, and faster convergence may be obtained. Modern digital signal-processing techniques allow quick and easy change (or selection) of (12) and (13) with different read-only memories at a fraction of the costs that are involved when traditional compromise equalizers have to be changed. Furthermore, even different signal formats may be processed efficiently with the same modem.

IV. APPLICATIONS TO PARTIAL-RESPONSE SIGNALING

Our results have so far been quite general, as we have not yet defined a special signaling technique. As was already mentioned in Section I, we see a particular promising application to processing zero excess-bandwidth partial-response signals. For such signals, the A

matrix depends only on the amplitude characteristics of the channel and is independent of the phase parameters[8] (timing phase, carrier phase, and phase characteristic of the channel). Since the amplitude response of voice-grade channels is relatively constant over the transmission band (or at least under much better control than the phase response), we may select an estimate

$$Q_m = \beta_m A_0^{-1}, \tag{16}$$

where A_0 is the ideal partial-response signal matrix. More specifically, for Class IV signaling, the matrix elements of A_0 are

$$[A_0]_{ik} = \delta_{ik} - \tfrac{1}{2}\delta_{i,k+2} - \tfrac{1}{2}\delta_{i,k-2} \tag{17}$$

and for duobinary are

$$[A_0]_{ik} = \delta_{ik} + \tfrac{1}{2}\delta_{i,k+1} + \tfrac{1}{2}\delta_{i,k-1}. \tag{18}$$

The inverse of both (nonsingular) matrices may be expressed as

$$A_0^{-1} = (I-G)^{-1} = I + G + G^2 + \cdots, \tag{19}$$

which shows that the elements of A_0^{-1} are rational numbers, and are symmetric with respect to both matrix diagonals.

Having selected an ideal channel as an estimate for Q, our algorithm takes the form

$$\boldsymbol{\phi}_{m+1} = (I - \beta_m A_0^{-1} A)\boldsymbol{\phi}_m = B_m \boldsymbol{\phi}_m. \tag{20}$$

Obviously, if we have only phase distortion, then $A = A_0$ and we may select $\beta_m = 1$. The equalizer will then reach its optimum tap vector within a single iteration because the matrix B_m is zero.

If the channel has amplitude distortion, then $A \neq A_0$ and thus $B_m \neq 0$. The eigenvalues λ_i and eigenvectors $\mathbf{w}_i$ of $A_0^{-1}A$ are defined by

$$A_0^{-1} A \mathbf{w}_i = \lambda_i \mathbf{w}_i. \tag{21}$$

Following a procedure shown by Chang,[8] we premultiply both sides by $\mathbf{w}_i^{\dagger} A_0$ and obtain

$$\lambda_i = \frac{\mathbf{w}_i^{\dagger} A \mathbf{w}_i}{\mathbf{w}_i^{\dagger} A_0 \mathbf{w}_i}. \tag{22}$$

Since both A_0 and A are symmetric, positive, definite matrices, we conclude that all λ_i are positive. Chang[8] has further shown that the eigenvalues can be bounded by

$$|H(\omega)|^2_{\min} \leqq \lambda_i \leqq |H(\omega)|^2_{\max}, \tag{23}$$

where $H(\omega)$ is the quotient between the actual and the estimated channel transfer function. For convergence, the spectral radius ρ of

B_m must be less than unity; this is guaranteed if we choose

$$0 < \beta_m < \frac{2}{\lambda_{\max}}, \tag{24}$$

since the eigenvalues μ_i of B_m are given by

$$\mu_i = 1 - \beta_m \lambda_i. \tag{25}$$

If the amplitude distortion is small, $A_0^{-1}A$ is close to the identity matrix, and the eigenvalues λ_i will be closely scattered around unity. We would then select $\beta_m \approx \text{const} \approx 1$, and the matrix B would be close to the zero matrix with all eigenvalues μ_i close to zero. The tap error vector after m iterations is given, from (20),

$$\boldsymbol{\phi}_m = (I - \beta A_0^{-1}A)^m \boldsymbol{\phi}_0 = B^m \boldsymbol{\phi}_0 \tag{26}$$

if β is constant (this is, of course, replaced by the usual matrix product if $\beta = \beta_m$). The squared magnitude of $\boldsymbol{\phi}$ and the excess MS error are given by the expressions

$$|\boldsymbol{\phi}_m|^2 = \boldsymbol{\phi}_0^T B^{2m} \boldsymbol{\phi}_0 \tag{27}$$

$$\epsilon^2 = \boldsymbol{\phi}_0^T B^m A B^m \boldsymbol{\phi}_0, \tag{28}$$

and can be used to estimate the convergence speed. The procedure and the results are practically identical to those of Chang[8] and are, therefore, not repeated here.

V. IMPLEMENTATION

Figure 1 is a basic block diagram of the new equalizer. A version with only three taps is shown for clarity, although a much larger and preferably even number of taps would actually be used. The upper part represents a traditional MS equalizer with the well-known circuitry for correlating the error signal with each tap signal to obtain the gradient $\mathbf{g} = A\mathbf{c} - \mathbf{v}$. Because of the particular structure of A for partial response signals, the components of $\mathbf{g}$ are highly correlated. The tap corrections obtained with a direct gradient algorithm would thus not be independent of each other. This interaction (which would cause a very slow convergence) can be eliminated if the gradient is "decorrelated" in the fixed weighting matrix Q shown in the lower part of Fig. 1. The resulting outputs are now decoupled and can be used to adjust the taps free of interaction to obtain rapid convergence.

A somewhat different arrangement is shown in Fig. 2. Here the (same) Q matrix is placed between the tap signals and the correlators. It is obvious that this arrangement is equivalent to Fig. 1 because the error signal is common to all correlator inputs.

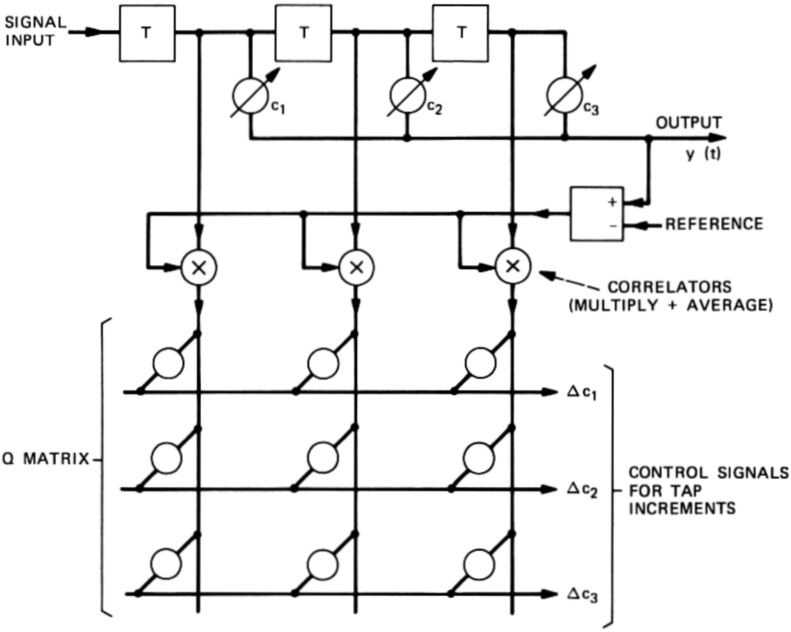

Fig. 1—Use of Q matrix in a first implementation of the new algorithm.

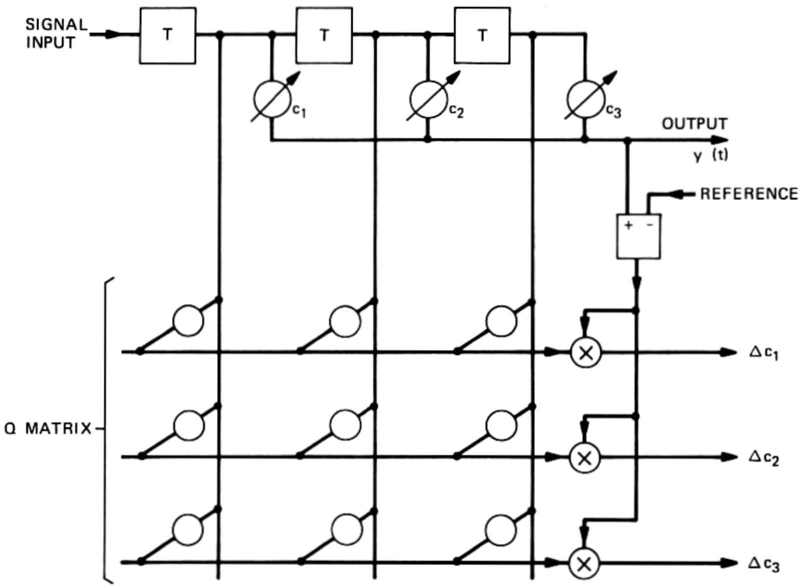

Fig. 2—Use of Q matrix in a second implementation of the new algorithm.

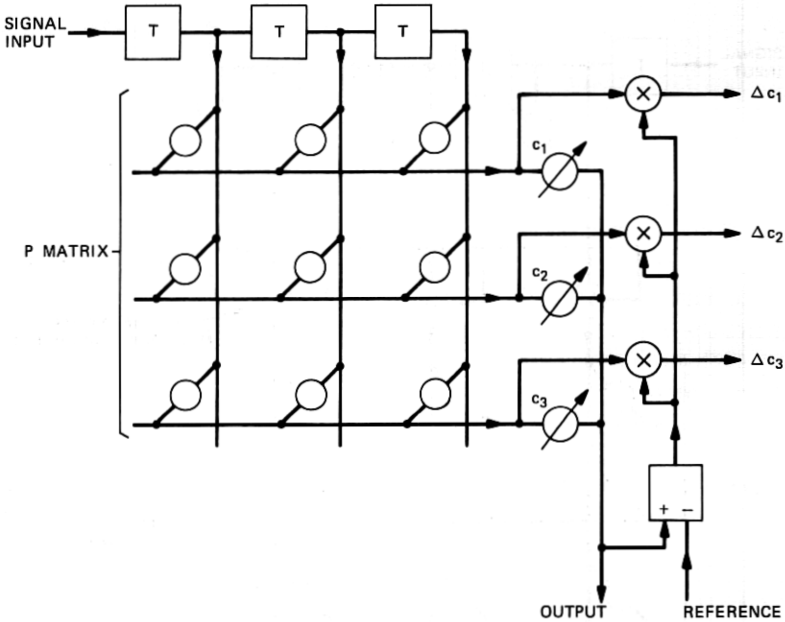

Fig. 3—Use of P matrix to obtain a modified equalizer structure.

As a comparison, we show in Fig. 3 a three-tap equalizer proposed by Chang.[8] Here the input signal is first transformed by a weighting matrix P into a set of orthonormal components that are then processed as in a conventional equalizer. It is, in fact, identical to a conventional equalizer except that the input signal vector $\mathbf{x}$ is replaced by $P\mathbf{x}$, where P is a suitably chosen matrix such that $P^TP = A_0^{-1}$. Initial presetting of the equalizer and constraining the range of the variable gain coefficients can, however, be quite complicated with this structure. These difficulties are avoided in the new proposal, since the equalizer structure remains unchanged and only the updating algorithm is modified.

Although the number of multiplications is the same in all three structures (Figs. 1 to 3), we would like to point out one very important difference: in Chang's structure, the accuracy of the transformation is very critical. The input signals must be represented by their full precision (say, 10 bits). To avoid round-off errors when all the products are summed up, more than 12 accurate bits of these products must be processed and can then (after summation) probably be rounded off to 10 bits for the transformed tap signal. Any errors introduced in this process will result in undesirable output noise that cannot be compensated for by the equalizer. In the new equalizer, however, only the tap increments are affected by errors in the matrix calculation. Obviously, a much lower accuracy is needed; in the extreme case, the correct polarity will be sufficient. Even with comparatively large errors, the tap-vector noise can be kept low by suitable scaling of the increments. We thus conclude that the new equalizer has inherent advantages as far as complexity, cost, size, speed requirements, and time-sharing possibilities are concerned. Furthermore, simpler algorithms can be derived from this one. If, for example, the structure of Fig. 2 is used for an algorithm with tap signals quantized to one bit, the matrix operation is reduced to controlling the addition and subtraction of the coefficients and may, for a medium-sized equalizer, be cheap and efficiently replaced by a suitably programmed read-only memory. For such nonmean-square algorithms, the optimum Q matrix will, in general, of course, be different from A_0^{-1}.

A few words should be said about the structure of the matrix $Q = A_0^{-1}$ in the special case of partial-response Class IV signaling. It can be shown that every second diagonal array is zero; furthermore, if N is even, every second row (or column) is a shifted version of the previous one. These properties are identical to those of Chang's P matrix (the remaining nonzero coefficients are, however, different from those of the P matrix) and can be used to simplify the signal-processing operations.

Finally, we would like to point out that an all-digital implementation for higher data speeds (group band) would still be extremely complex and expensive. For such applications, an array of operational amplifiers with suitable weighting resistors (probably mounted on a ceramic) would be a far more economic solution with the present state of the art.

VI. SIMULATIONS

To test the proposals made in this paper, a 15-tap MS equalizer was simulated. A 4800-b/s, SSB-modulated, partial-response, Class IV signal was selected. A parabolic delay characteristic was assumed with 1-ms delay difference between the passband channel center and edges. Table I shows the coefficients of the matrix $Q = A_0^{-1}$. In Fig. 4, the equalizer convergence is shown for four different timing phases that are separated by 90 degrees each. After each iteration, the MS distortion of the overall impulse response is plotted. It is seen that the minimum is basically achieved with the first iteration (the small residual error after the first iteration results from the time-domain truncation and round-off errors in the components of the initial impulse response during the simulation). Some additional simulations were made with an MS stochastic approximation algorithm (adjustments at the symbol rate) and with amplitude distortion; they also showed a clearly improved convergence when the decorrelation matrix was used.

Table I — Decorrelation matrix for $N = 15$ (elements multiplied with 36)

64	0	56	0	48	0	40	0	32	0	24	0	16	0	8
0	63	0	54	0	45	0	36	0	27	0	18	0	9	0
56	0	112	0	96	0	80	0	64	0	48	0	32	0	16
0	54	0	108	0	90	0	72	0	54	0	36	0	18	0
48	0	96	0	144	0	120	0	96	0	72	0	48	0	24
0	45	0	90	0	135	0	108	0	81	0	54	0	27	0
40	0	80	0	120	0	160	0	128	0	96	0	64	0	32
0	36	0	72	0	108	0	144	0	108	0	72	0	36	0
32	0	64	0	96	0	128	0	160	0	120	0	80	0	40
0	27	0	54	0	81	0	108	0	135	0	90	0	45	0
24	0	48	0	72	0	96	0	120	0	144	0	96	0	48
0	18	0	36	0	54	0	72	0	90	0	108	0	54	0
16	0	32	0	48	0	64	0	80	0	96	0	112	0	56
0	9	0	18	0	27	0	36	0	45	0	54	0	63	0
8	0	16	0	24	0	32	0	40	0	48	0	56	0	64

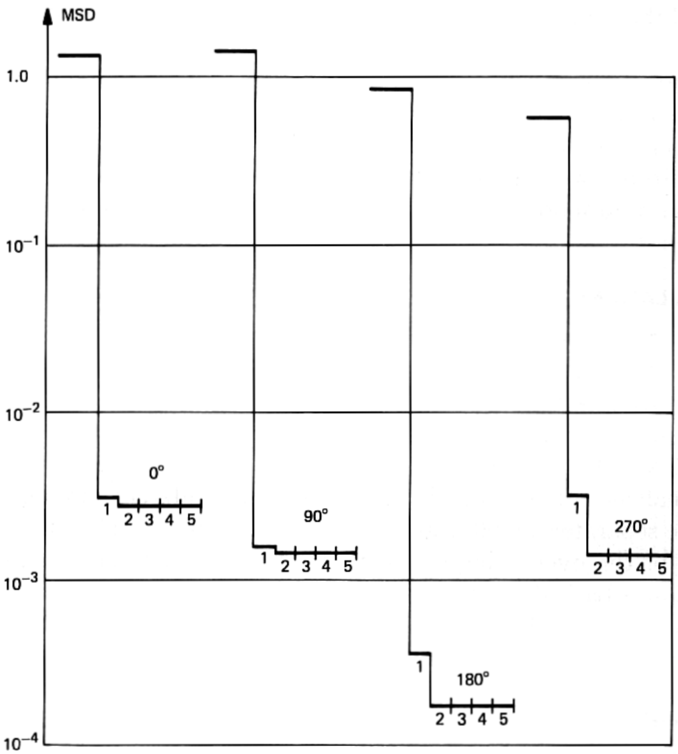

Fig. 4—Convergence of new algorithm for different timing phases.

VII. CONCLUSIONS

The spread of the eigenvalues of the signal autocorrelation matrix can become very large for signals with nonflat amplitude spectra. If the gradient with respect to the tap gains of an MS equalizer is calculated, its components are highly correlated. The resulting interaction of the individual gain adjustments may cause a very slow equalizer convergence. We have proposed a new MS algorithm that eliminates these problems. With this new algorithm, a set of uncorrelated tap increments is calculated so that the individual adjustments are decoupled from each other. This can be achieved in either of two ways: The gradient can first be calculated in the conventional way (correlation of error signal and tap signals) and then be multiplied by a decorrelation matrix Q to produce the final correction vector. On the other hand, tap signals can first be passed through such a matrix and the resulting outputs can then be correlated with the error signal. The matrix Q is chosen as the inverse of the expected (average) signal autocorrelation matrix A. The new algorithm was simulated for partial response signals with zero excess bandwidth. In the case of a flat channel, the optimum tap coefficients can then be determined within a single iteration, independent of delay distortion, carrier phase, and timing instant. This was confirmed by simulation. Compared with other recent proposals for fast start-up, the new algorithm features simpler presetting and requires less accuracy in the matrix transformation. Furthermore, this method can easily be extended to systems with nonmean-square cost functions.

REFERENCES

1. E. R. Kretzmer, "Generalization of a Technique for Binary Data Transmission," IEEE Trans. on Commun. Technology, *COM-14*, No. 1 (February 1966), pp. 67–68.
2. A. Lender, "Correlative Digital Communication Techniques," IEEE Trans. on Commun. Technology, *COM-12*, No. 4 (December 1964), pp. 128–135.
3. A. Lender, "The Duobinary Technique for High Speed Data Transmission," IEEE Trans. on Commun. and Elec., *82*, No. 66 (May 1963), pp. 214–218.
4. P. J. vanGerwen, "On the Generation and Application of Pseudo-Ternary Codes in Pulse Transmission," Philips Research Reports, *20* (August 1965), pp. 469–484.
5. P. Leuthold and F. Tisi, "Ein Einseitenbandsystem fuer Datenuebertragung," Archiv der Elektrischen Uebertragung, *21*, No. 7 (July 1967), pp. 354–362.
6. R. W. Lucky, J. Salz, and E. J. Weldon, *Principles of Data Communication*, New York: McGraw-Hill, 1968.
7. D. Hirsch and W. J. Wolf, "A Simple Adaptive Equalizer for Efficient Data Transmission," IEEE Trans. on Commun. Technology, *COM-18*, No. 1 (February 1970), pp. 5–12.
8. R. W. Chang, "A New Equalizer Structure for Fast Start-Up Digital Communication," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1969–2014.