# Adaptive Redundancy Removal in Data Transmission

By R. W. LUCKY

*This paper suggests an adaptive filter, similar to that used in automatic equalization, for use as a predictor in data compression systems. It discusses some of the applications of this adaptive predictor in digital data transmission. In the event of redundant data input to the system the predictor could be used to lower the transmitted power output required for a given error rate or to decrease the error rate while maintaining constant transmitted power. The action of these redundancy-removal and restoration systems is analyzed in simple cases involving Markov inputs.*

## I. INTRODUCTION

In the design, analysis, and testing of data transmission systems it is invariably assumed that the input digits are identically distributed, independent random variables. However, in many actual systems the input digits may arise from a physical source which imposes significant correlations in the data train. In these cases we know that the entropy of the source is less than when independent digits are presented. Accordingly, we should be able to use the redundancy in the input message to provide, in some sense, more efficient transmission. For example, we could imagine the redundancy being used to decrease bandwidth, to increase speed, to lower probability of error, or to lower average signal power.

Redundancy removal in analog transmission systems was investigated in the early 1950's by Oliver, Kretzmer, Harrison, and Elias[1-4]. Each of these papers relied on the theory of linear prediction as developed by Wiener in the early 1940's.[5] Figure 1 shows the basic idea. It is assumed that the input samples are taken from a stationary time series $\{x_n\}$. These samples are passed through a linear filter whose output $\hat{x}_n$ at time $t_n$ forms a linear prediction of the sample $x_n$ based on all preceding samples. The prediction $\hat{x}_n$ is subtracted from the actual sample $x_n$ and only the error $e_n$ is passed on for further processing and

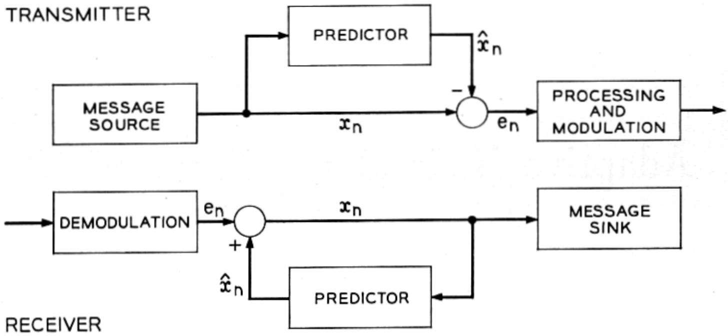



Fig. 1 — Predictive system.

transmission. Since the portion $\{\hat{x}_n\}$ "removed" from the input sequence is a deterministic function of the error sequence, no information has been lost and the original sequence can be reconstructed at the receiver by the feedback loop shown in the figure.

The philosophy of predictive systems has been widely studied for its application in bandwidth compression of telemetry data and of television; for example, see Kortman, Davisson, and O'Neal.[6-8] In these examples the error samples $e_k$ are quantized and transmitted by pcm. Because of redundancy, that is, predictability, in the source data, fewer digits per sample (and consequently less bandwidth) are required for transmitting the error samples than for transmitting the original samples for a given fidelity of reconstruction.

One of the difficulties with these data compression systems is in determining the predictor filter. Although the theory of linear prediction for stationary time series is well known, the practical determination of the statistical properties of the input data and the realization of the corresponding optimum filter are nearly impossible. Generally, an approximate average statistical description is used for the input data and a considerably simplified version of the optimum filter is constructed. Most existing compression schemes appear to use only linear or zero-order extrapolation of the previous sample to form the prediction of the succeeding sample. More complicated and adaptive prediction techniques have been confined to computer-processed data.

In this paper we describe a simply-instrumented adaptive filter for use as a predictor. This filter uses a finite tapped delay line whose coefficients are continually adjusted to provide a least squares prediction of incoming data. The coefficient settings are based on the statistics of a finite section of the past data (the learning period). As the statistics of the data during this learning period change, the coefficients are changed to provide an updated version of the predictor filter.

Although the most obvious applications of this adaptive predictor would be in the transmission of television or some other very redundant analog signal, we choose here to explore its application in digital data transmission. In the past, little attention seems to have been focused on the use of prediction in digital transmission. Presumably this is because the most effective use of prediction would be in the compression of the analog wave from which the digits are taken.

However, there do exist situations in which the input digital signal is not under the control of the transmission systems designer. This occurs notably in the design of data communications equipment. Although it has been common practice to use redundancy in speech signals to ease transmission system requirements (the TASI system is a dramatic example), nothing similar has been attempted with digital data signals. There would seem to be no compelling reason why any redundancy in digital signals should not be taken advantage of, as long as the error statistics of the output data were not adversely affected by the procedure. After describing a digital redundancy removal and restoration system we shall discuss its possible benefits to the customer and to the transmission plant.

## II. SYSTEM DESCRIPTION

Figure 2 shows a digital redundancy removal and restoration scheme. For simplicity we assume that the input digits $a_n$ are binary, although the technique obviously extends to multilevel transmission. The input sequence is passed through a shift-register transversal filter whose tap gains $c_k$ have been adjusted so that the filter output $\hat{a}_n$, where

$$\hat{a}_n = \sum_{k=1}^{N} c_k a_{n-k} , \tag{1}$$

is a linear least squares prediction of $a_n$. This prediction is subtracted from the actual sample $a_n$ and only the difference $e_n$ is passed to the modulator for transmission. Notice that, although $a_n$ is a binary variable taking on the values $\pm 1$, both $\hat{a}_n$ and $e_n$ are analog. Unless the digits $a_n$ are uncorrelated, the error samples $e_n$ will have smaller variance than the unit variance of the input data. Consequently, a linear modulator

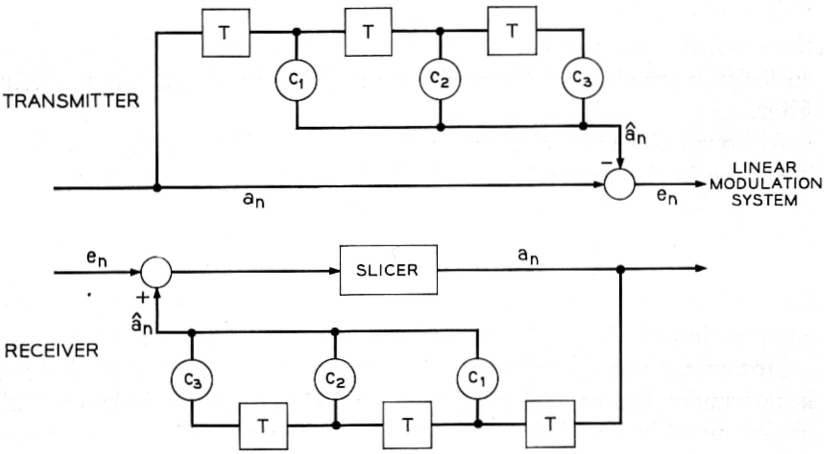



Fig. 2 — Digital redundancy removal and restoration.

will put out less line power in transmitting the error samples than in transmitting the original data.

After demodulation at the receiver, the missing, predictable, component $\hat{a}_n$ must be added to the error sample $e_n$ before slicing, in order to recover $a_n$. This component is obtained by a bootstrap arrangement wherein the detected symbols are passed through a transversal filter identical to that at the transmitter in order to form the predictions $\hat{a}_n$. The receiver is similar in arrangement to the circuitry used in dc restoration.

There are two relatively simple ways in which this system could be used to improve transmission efficiency. As shown in Figure 2 the system lowers the average transmitted power without appreciably affecting the output data error rate. In this mode of operation any benefit from the data redundancy is used to lower the load requirements on the transmission plant. If many data sets were equipped with such circuitry, the average power handled by the plant would be lowered in a statistical fashion. Some sets, transmitting entirely random data, would require their normal power complement. Others, transmitting redundant data, would require considerably less. Notice that this is exactly the type of effect which now takes place for voice transmission.

As the input data becomes entirely redundant in the limit, the transmitted power goes to zero. In this case the input data consists of a periodic pattern. In spite of the zero-level line signal, the pattern is reconstructed exactly at the receiver (in the absence of noise). Such an eventuality would alleviate the problems now encountered with the transmission of periodic data. These data patterns normally lead to tones, that is, line spectra, in the transmission channel which cause certain overloading and other system malfunctions.

Currently the problem is being treated in wideband transmission by the introduction of digital scramblers.[9] In practice the zero-level transmitted signal would not be a satisfactory solution to the tone problem since some signal strength would be required for synchronizing and timing maintenance. However, proper design of the system could ensure that some minimum signal strength was maintained under all circumstances. For example, a nonlinear element in each predictor could be used to keep the predictions smaller than unity. As long as the same nonlinearity were used in both transmitter and receiver, the data signal would be reconstructed perfectly at the receiver.

The other simple way to use redundancy removal to aid transmission would be to keep the level of transmitted power constant while lowering the probability of error. In this case, compensating gain controls would be placed at the transmitter output and at the receiver input. These controls would be adjusted to keep the transmitted power constant regardless of signal redundancy. During periods of redundancy most of the voltage presented to the slicer at the receiver would come via the feedback predictor and therefore would be noiseless (in the absence of errors). Since the small error signal transmitted would be greatly amplified to keep line power constant, the total noise presented to the slicer after complementary deamplification would be much smaller than in normal transmission. Consequently, the error rate would be diminished during periods of redundant data transmission.

Complementary amplification and deamplification surrounding channel noise introduction are automatically accomplished in transmission over compandored facilities. Normally for these channels we would expect that the error rate would be independent of transmitted power level. In the redundancy removal system, however, this mechanism is defeated by using the noiseless feedback in the detection process.

There are further uses of redundancy removal in data transmission, but they appear to involve more complicated system arrangements. For example, the bit rate and bandwidth of the data signal could be lowered for redundant data. This could be accomplished by slicing the prediction $\hat{a}_n$ to obtain a closest *digital* prediction and then subtracting $\hat{a}_n$ from $a_n$ in digital form. The resulting error digits could then be processed by run-length encoding to achieve message compression. Of course we would then need a buffer to ensure a constant channel bit rate. We will not discuss this type of system further here.

Thus far we have alluded to the possible benefits of redundancy removal in data transmission. There is also one major drawback—that of error propagation. Since the estimate $\hat{a}_n$ at the receiver depends on the correct reception of all previous data, the compensation at the receiver is perfect only in the absence of errors. When an error occurs, the probability of error in succeeding bits tends to be larger and an error propagating effect occurs. Notice that this effect does not depend on the particular circuit configuration for its existence, but is a philosophical necessity in any redundancy removal operation. We analyze the effect of error propagation in a simple example in Section V. Normally we would not expect the error propagation to increase the entire error rate by more than a small algebraic factor.

## III. THE ADAPTIVE PREDICTION FILTER

In the theory of linear prediction developed by Wiener[5] and others it is assumed that the input samples $a_n$ are taken from a stationary time series with known covariance function $R(n)$, where

$$E[a_m a_n] = R(m - n). \tag{2}$$

The power output, which is the mean square prediction error, is

$$P = E[e_n^2] = E\left\{\left(a_n - \sum_{k=1}^{N} c_k a_{n-k}\right)^2\right\}. \tag{3}$$

The coefficients $c_k$; $k = 1, \ldots N$, which minimize this prediction error, can be obtained by the solution of the $N$ simultaneous equations

$$\sum_{k=1}^{N} c_k R(n - k) = R(n); \qquad n = 1, 2, \cdots, N. \tag{4}$$

In case of an infinite filter ($N = \infty$) the coefficients $c_k$ and the prediction error are given by a method involving factoring of the spectral density $G(f)$ of the input process. Under proper conditions the prediction error $P$ can be expressed in the form

$$P = \exp\left[\int_{-\frac{1}{2}}^{\frac{1}{2}} \log G(f)\, df\right] \tag{5}$$

(See Doob for the mathematical niceties of this result.[10]) Notice that if the input symbols are independent, $G(f) = 1, \mid f \mid \leq \frac{1}{2}$, and $P = 1$. Since the input power is also unity no gain is achieved by the prediction process. If, on the other hand, $G(f)$ is not flat the prediction error, $P$ is less than unity and power is saved.

While the mathematics of linear prediction for stationary time series serve as a guide to actual system performance, it is clear that the assumptions are philosophically inadmissible. Furthermore, since the data source is outside the designer's control, it would be extremely unlikely that the covariance function would be known in advance. For these reasons, Balakrishnan[11] in 1961 developed a mathematical formulation for a learning or adaptive predictor wherein the form of the prediction operator was dependent solely on the past data and not on any assumptions of stationarity or of prior knowledge of data statistics.

In Balakrishnan's formulation that prediction operator is chosen as optimum at time $t_n$ which works best when applied at times $t_{n-1}, \ldots, t_{n-L}$. Since all past information is available, we could "try out" all possible prediction operators on the previous data and select the operator for which

$$E_n = \sum_{j=1}^{L} [a_{n-j} - \hat{a}_{n-j}]^2 w_j \tag{6}$$

is minimum. The weights $w_j$ could be used to assign a relative importance to each past trial of the predictor.

For our finite linear predictor we have

$$E_n = \sum_{j=1}^{L} \left[ a_{n-j} - \sum_{k=1}^{N} c_k a_{n-j-k} \right]^2 w_j \, . \tag{7}$$

In order to develop a physical implementation for this adaptive filter we use a motivation based on a steepest descent approach. The derivatives of the error $E_n$ with respect to the coefficients $c_m$ are

$$\frac{\partial E_n}{\partial c_m} = - \sum_{j=1}^{L} 2w_j \left[ a_{n-j} - \sum_{k=1}^{N} c_k a_{n-j-k} \right] a_{n-j-m} \tag{8}$$

$$\frac{\partial E_n}{\partial c_m} = - \sum_{j=1}^{L} 2w_j e_{n-j} a_{n-j-m} \, . \tag{9}$$

Notice that these derivatives can be obtained by passing the product of sample $a_{n-m}$ and the error voltage $e_n$ through a filter with impulse response $\{w_j\}$. Thus we are led to the adaptive filter configuration shown in Figure 3. This configuration is entirely similar to that currently being used for equalization[12] and for echo suppression.[13, 14]

When the input samples $a_n$ are digital, the circuitry of Figure 3 is quite simple. The delay line becomes a shift register and the multipliers become simple polarity switches. However, the circuit is not limited to digital applications, but could be used in such analog functions as telemetry or television compression systems.

In any event, the response of the system, involving accuracy and settling time as well as stability, is controlled by selection of the smoothing filters $W(\omega)$. Basically these filters must perform an averaging followed by an integration. If the data were stationary and the memory $L$ sufficiently long, the result of averaging the product of the error and sample voltages for the $m^{\text{th}}$ tap coefficient would give (see equation 8)

$$y_m(t) \cong E[a_{n-m}e_n] = R(m) - \sum_{k=1}^{N} c_k(t)R(m-k). \tag{10}$$

Then these voltages would be integrated for use as tap coefficients, so that the governing system equations would be

$$\dot{c}_m(t) = A\left[R(m) - \sum_{k=1}^{N} c_k(t)R(m-k)\right] \quad \text{for} \quad m = 1, \cdots, N. \tag{11}$$

This system would be stable for all $A$, since the covariance matrix, whose $nm^{\text{th}}$ entry is $R(n\text{-}m)$, must be positive definite (see Davenport and Root[15]). All voltages $y_m(t)$ would be asymptotically reduced to

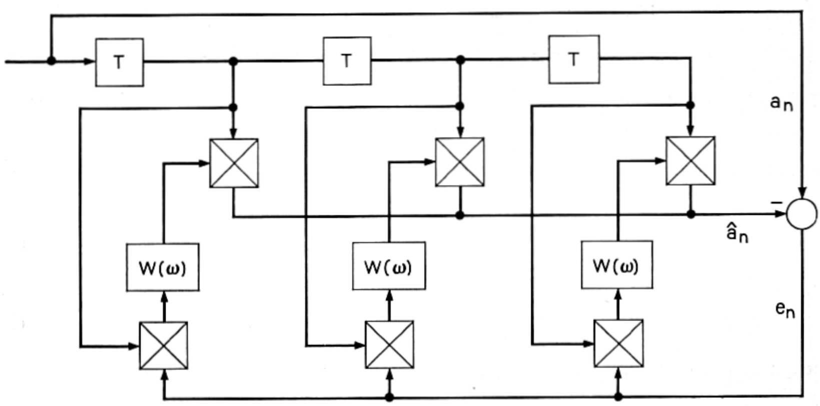



Fig. 3 — Adaptive prediction filter.

zero and the filter coefficients would asymptotically approach those of the optimum (least squares) linear predictor of equation (4).

For nonstationary data and realistic filters $W(\omega)$ the analysis of the nonlinear, multidimensional control system is extremely complicated. Let us study the dynamics of the one-dimensional system formed by using a one-tap predictor as a guide to the behavior of the system.

In order to put this analysis into proper perspective with regard to the system of Figure 2 we should observe that when the input data statistics change abruptly, both transmitter and receiver predictors undergo the same transients. If the predictors are identical, these transients cancel exactly at the receiver summer and no loss in noise margin is suffered. However, the statistics of the transmitted signal are affected by only the transmitter predictor. Therefore, the proper design of the adaptive predictor is crucial to obtaining desirable line power statistics, but not to the performance of the entire system.

## IV. THE ONE-TAP TRANSMITTER FOR BINARY DATA

Figure 4 shows a one-tap transmitter with a binary input signal of the form

$$\begin{aligned} s(t) &= \sum_{n=0}^{\infty} a_n r(t - nT) \\ a_n &= \pm 1 \\ r(t) &= \begin{cases} 1 & 0 \leqq t < T \\ 0 & \text{elsewhere} \end{cases}. \end{aligned} \tag{12}$$

The transmitted voltage is given by

$$e(t) = s(t) - c(t)s(t - T) \tag{13}$$

where

$$c(t) = Aw(t)*[s(t - T)e(t)]. \tag{14}$$

Because of the binary nature of the input $s^2(t) = 1$ and thus

$$c(t) = Aw(t)*[s(t)s(t - T) - c(t)]. \tag{15}$$

Let $m(t) = s(t)s(t-T)$; then the Laplace transform solution for $C(s)$ is*

* Some liberty has been taken with the shift-register starting state.

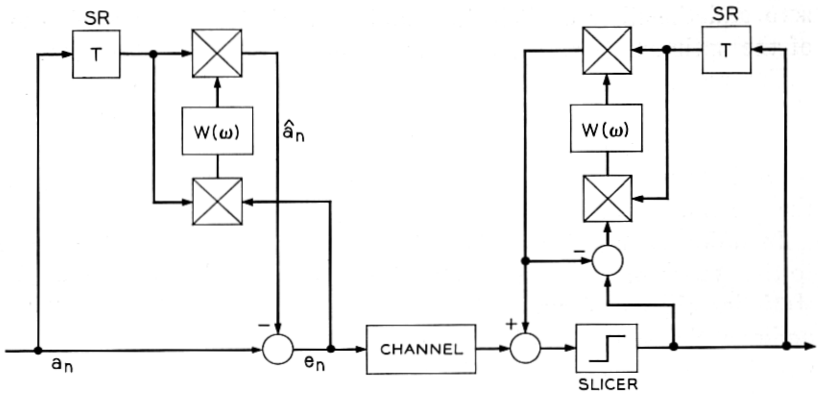



Fig. 4 — One-tap digital system.

$$C(s) = \frac{AW(s)M(s)}{1 + AW(s)}. \tag{16}$$

Now returning to equation (13) we multiply both sides by $s(t-T)$ to obtain

$$e(t)s(t - T) = m(t) - c(t). \tag{17}$$

Combining equations (16) and (17) gives

$$e(t)s(t - T) = m(t)*h(t) \tag{18}$$

where

$$H(s) = \frac{1}{1 + AW(s)}. \tag{19}$$

The output signal itself can be written by again multiplying equation (18) by $s(t-T)$

$$e(t) = s(t - T)[m(t)*h(t)]. \tag{20}$$

Notice that the special properties of binary sequences have been used in arriving at this solution, so that equation (20) does not hold for multilevel or analog input.

Figure 5(a) shows the mathematically equivalent transmitter given by equation (20) as well as its corresponding receiver. Since the second multiplier does not affect the transmitted power in any way, both transmitter and receiver can be simplified by its removal to result in the equivalent represented by Figure 5(b).* This final

* The systems differ in their noise performance, however.

equivalent system is amazingly simple and appears to bear little resemblance to the initial system of Figure 4. It is interesting to observe that, while the initial system was termed "adaptive," no one would seriously consider its equivalent in Figure 5(b) as being adaptive in any sense.

Figure 5(b) has an intriguing interpretation. The input data is first subjected to the nonlinear operation of delay and multiplication. The output of the multiplier is

$$m(t) = \sum_n a_n a_{n-1} r(t - nT). \tag{21}$$

This voltage has a mean value given by $R(1)$ in the stationary case. If the filter $W(\omega)$ has been designed as a low pass filter, then the filter $1/[1 + AW(\omega)]$ in the equivalent circuit is a high pass filter. Thus the dc component of $m(t)$ is removed before transmission and reinserted via a dc restorer at the receiver. In other words, a nonlinear operation on the input signal has converted the correlation into a spectral line which can then be removed by a time invariant linear filter. It would seem that some generalization of this concept should be possible, but as yet none has been found.

The equivalent circuit can be used for design purposes in selecting

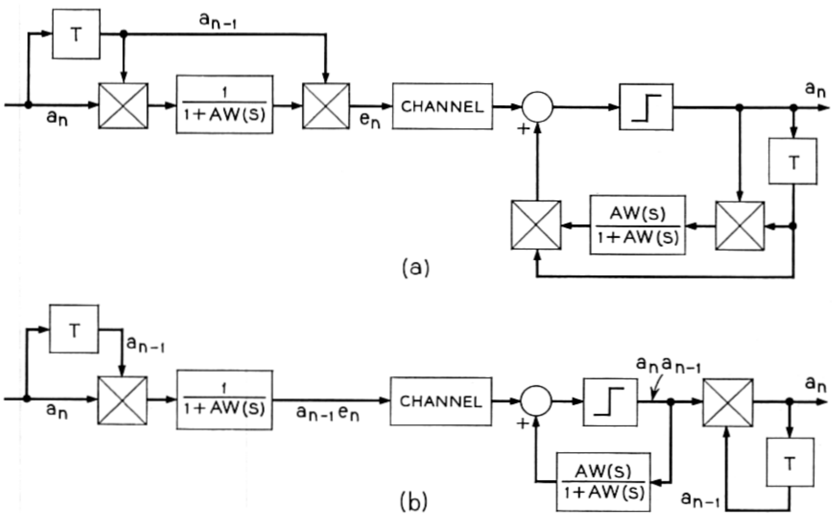



Fig. 5 — Equivalent binary one-tap systems. (a) Equivalent system. (b) Simplified equivalent.

$W(\omega)$, or for calculating line power or transient response. Here are the results of a few straightforward examples.

*Example 1*

Simple RC filter, dotting pattern input applied at time zero:

$$W(s) = \frac{\alpha}{s+\alpha}; \qquad \alpha = \frac{1}{RC}$$

$$a_n = \begin{cases} +1, & n \text{ even} \\ -1, & n \text{ odd} \end{cases}. \tag{22}$$

A deterministic sequence is to be transmitted. We find that the output of the equivalent circuit is

$$e(t)s(t-T) = -\left[\frac{1}{A+1}u(t) + \frac{A}{A+1}e^{-\alpha(A+1)t}\right]. \tag{23}$$

Thus the error voltage transmitted in the original circuit becomes

$$e(t) = \left[\sum_{n=0}^{\infty}(-1)^n r(t-nT)\right]\left[\frac{1}{A+1}u(t) + \frac{A}{A+1}e^{-\alpha(A+1)t}\right]. \tag{24}$$

The error voltage does not approach zero because of the lack of an integration in the smoothing filter.

*Example 2*

Simple RC filter, markov input:

If the input is a first order Markov process the one-tap predictor becomes the optimum linear predictor. (We study this case more thoroughly in the next section.) The covariance function of the input time series is taken to be

$$R(n) = R^{|n|}. \tag{25}$$

Since we now are dealing with a random input, our concern is with the transmitted power level rather than the exact waveform as in the previous example. The transmitted power is the same in Figures 4 and 5b, so we use the simpler structure of the latter diagram for analysis.

When the input Markov process is subjected to delay and multiplication, it can be shown that the resultant symbols $(a_n a_{n-1})$ have mean value $R$ and are uncorrelated. The spectral density of the multiplier output $m(t)$ is given by

$$S_M(\omega) = R^2\ \delta(\omega) + (1 - R^2)T\ \frac{\sin^2 \frac{\omega T}{2}}{\left(\frac{\omega T}{2}\right)^2}. \tag{26}$$

This spectral density can be multiplied by $|\ H(\omega)\ |^2$ and integrated to give the transmitted power. The power becomes

$$P = \frac{R^2}{(1 + A)^2} + (1 - R^2) \cdot \left\{\frac{1}{(1 + A)^2} + \left[1 - \frac{1}{(1 + A)^2}\right]\left[\frac{1 - e^{-\alpha(1+A)T}}{\alpha(1 + A)T}\right]\right\}. \tag{27}$$

Ideally, of course, this power should be $(1-R^2)$, but the crude RC filter is unable to approximate this result unless the gain is high and the time constant $(1/\alpha)$ is large.

Better results in both examples could be achieved by an improved selection of the filter characteristic $W(\omega)$. We can see from the equivalent circuit that the best choice of $W(\omega)$ makes $1/[1 + AW(\omega)]$ an efficient high pass filter with a transmission zero at $\omega = 0$. Of course this must be compromised with any requirement on the filter response time.

In this section we stress the use of the equivalent circuit as a method of analysis rather than as an implementable system. Clearly, if one were to build a one-tap binary predictor, the circuit of Figure 5(b) would be preferred to that of the original system. However we believe that such a restricted system would not be of great practical interest.

While the implementation of the simple equivalent circuit cannot be extended to wider application, it is hoped that the easy analysis of the simple system conveys some insight into the performance of multiloop systems. This would be particularly true if there were small interaction between taps on the multiloop system. Such a situation would occur if the covariance $R(n)$ decreased rapidly with $n$.

## V. ERROR PROPAGATION

When noise is added in the transmission channel there is some probability of the received digits being incorrectly detected by the slicer. Even though the transmitted power might have been substantially reduced by the redundancy removal, the probability of an initial error is identical to that of a full power system. Once an error has been made, however, the probability of making subsequent errors is increased because of the incorrect symbol being used in redundancy restoration. Thus, errors tend to bunch together in the received data. Besides increasing the average probability of error this error propagation considerably complicates the problems of error control in the entire system.

Error propagation in dc restoration circuits has been examined by Zador, Aaron, and Simon.[16, 17] It appears to be a very complicated problem, in general, which is even more confused by the presence of the adaptive, pattern sensitive filters in the redundancy removal system we are considering here. Therefore, we shall attempt the analysis of only the simplest meaningful theoretical model. Both transmitter and receiver will have one-tap transversal filters as shown in Figure 4. The input data is taken to be a binary first order Markov process, with zero mean and covariance

$$R(n) = R^{|n|}.$$

The transition matrix for this process is:

$$\begin{array}{cc|c|c|} & & \multicolumn{2}{c}{a_{n+1}} \\ & & +1 & -1 \\ \hline \multirow{2}{*}{$a_n$} & +1 & \frac{1+R}{2} & \frac{1-R}{2} \\ \cline{3-4} & -1 & \frac{1-R}{2} & \frac{1+R}{2} \\ \hline \end{array} .$$

The ideal linear predictor for this time series is simply $\hat{a}_n = Ra_{n-1}$ and the average transmitted power using this predictor is $1 - R^2$. Since the ideal predictor uses only a single tap filter, the assumption of single tap filters in the actual system is not particularly restrictive. If additional taps were used, their gains would be small and their effect on error propagation would not be significant.

We will assume that noise samples $\xi_k$, uncorrelated Gaussian random variables with zero mean and variance $\sigma^2$, are added to the transmitted symbols in the channel. We further assume that sufficient smoothing is done at the transmitter so that the tap gain may be fixed at its optimum value, $R$. Thus the transmitted samples are

$$e_k = a_k - Ra_{k-1} . \tag{28}$$

Now at the receiver we shall write the received symbols as $\beta_k a_k$. The parameter $\beta_k = \pm 1$ indicates the absence $(+1)$ or the presence $(-1)$ of an error at time $t_k$. If the tap gain at the receiver is denoted by the parameter $c$, the detected symbols can be written

$$\beta_k a_k = \text{sgn}\,[a_k - a_{k-1}(R - c\beta_{k-1}) + \xi_k]. \tag{29}$$

Thus the error parameter $\beta_k$ is

$$\beta_k = \text{sgn}\,[1 - a_k a_{k-1}(R - c\beta_{k-1}) + \eta_k] \tag{30}$$

where $\eta_k = \xi_k a_k$ has the same statistical properties as $\xi_k$. The probability of error at time $t_k$ is the probability that $\beta_k = -1$, which is the probability that $\eta_k$ is such that the term in brackets is negative.

Now we must turn our attention to the behavior of the receiver tap gain $c$. If no errors are made, then this gain is identical to the transmitter gain and as $k \to \infty$, $c \to R$. However, because of the presence of errors, the receiver tap gain tends to be different from the transmitter tap gain. At time $t_k$ the output voltage of the multiplier at the receiver is

$$v_k = \beta_k a_k \beta_{k-1} a_{k-1} - c. \tag{31}$$

The random variables $v_k$ are averaged to determine the movement of $c$. Notice that, since $| \beta_k a_k \beta_{k-1} a_{k-1} | = 1$, the magnitude of $c$ cannot exceed unity except as a transient starting state. This eliminates any possibility of a runaway in $c$ resulting from unusual error patterns.

We assume that the action of the loop at the receiver is to reduce to zero the expectation of the multiplier output voltage at time infinity. Thus

$$E[v_\infty] = 0 = \lim_{k\to\infty} E[\beta_k a_k \beta_{k-1} a_{k-1}] - c_\infty . \tag{32}$$

This type of final behavior would be exhibited by systems in which $W(\omega)$ consisted of a long term averaging followed by an integration. The expectation of the term in brackets in equation (32) depends on $c_\infty$ itself, so in general we end with a fairly complicated equation requiring a trial and error solution for $c_\infty$. By taking the limit as $k \to \infty$ of the expectation we eliminate the dependence on time and on the initial probability distributions for the random variables involved.

Define a vector random variable $\bar{\alpha}_k = (a_k , \beta_k)$ taking on the four possible states (+1, +1), (+1, −1), (−1, +1) and (−1, −1), denoted by states 1 through 4, respectively. Because $a_k$ is Markov and since the expression for $\beta_k$ in equation (30) involves only $a_k$, $a_{k-1}$, $\beta_{k-1}$, and $\eta_k$, we conclude that $\bar{\alpha}$ is also Markov. The four-by-four transition matrix $\pi$ for $\bar{\alpha}$ has entries $p_{ij}$ which may be calculated from the original transition matrix for the input symbols $a_k$ and from equation (30) for the probabilities of error in various states. Table I lists these transition probabilities. If the 4-entry row vector $\bar{w}^{(k)}$ gives the probabilities of $\bar{\alpha}_k$ assuming each of the four possible states, then

$$\bar{w}^{(k)} = \bar{w}^{(k-1)}\pi. \tag{33}$$

In terms of the initial state distribution $\bar{w}^{(0)}$

$$\bar{w}^{(n)} = \bar{w}^{(0)}\pi^n. \tag{34}$$

For $|R| < 1$ it is clear from standard Markov chain theory (see, for example, Reference 18) that steady-state probabilities exist for

TABLE I—TRANSITION PROBABILITIES FOR $\bar{\alpha}_k = (a_k, \beta_k)$

$$Q(x) = \int_x^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$$

$$p_{11} = p_{33} = \left(\frac{1+R}{2}\right)\left[1 - Q\left(\frac{1-R+c}{\sigma}\right)\right]$$

$$p_{12} = p_{34} = \left(\frac{1+R}{2}\right)Q\left(\frac{1-R+c}{\sigma}\right)$$

$$p_{13} = p_{31} = \left(\frac{1-R}{2}\right)\left[1 - Q\left(\frac{1+R-c}{\sigma}\right)\right]$$

$$p_{14} = p_{32} = \left(\frac{1-R}{2}\right)Q\left(\frac{1+R-c}{\sigma}\right)$$

$$p_{21} = p_{43} = \left(\frac{1+R}{2}\right)\left[1 - Q\left(\frac{1-R-c}{\sigma}\right)\right]$$

$$p_{22} = p_{44} = \left(\frac{1+R}{2}\right)Q\left(\frac{1-R-c}{\sigma}\right)$$

$$p_{23} = p_{41} = \left(\frac{1-R}{2}\right)\left[1 - Q\left(\frac{1+R+c}{\sigma}\right)\right]$$

$$p_{24} = p_{42} = \left(\frac{1-R}{2}\right)Q\left(\frac{1+R+c}{\sigma}\right)$$

the transition matrix $\pi$, that is, $\bar{w}^{(n)}$ approaches a constant vector $\bar{w}$ as $n \to \infty$ independent of $\bar{w}^{(0)}$. The steady-state probabilities of the four possible states can be obtained by the solution of the equations given by

$$\bar{w}\pi = \bar{w}. \tag{35}$$

Some algebraic manipulation yields the probabilities

$$w_1 = P(a_\infty = +1, \beta_\infty = +1) = \frac{\frac{1}{2}(1 - p_{22} - p_{24})}{1 - p_{22} + p_{12} - p_{24} + p_{14}} \tag{36}$$

$$w_2 = P(a_\infty = +1, \beta_\infty = -1) = \tfrac{1}{2} - w_1 \tag{37}$$

$$w_3 = P(a_\infty = -1, \beta_\infty = +1) = w_1 \tag{38}$$

$$w_4 = P(a_\infty = -1, \beta_\infty = -1) = \tfrac{1}{2} - w_1 \tag{39}$$

where the transition probabilities $p_{12}$, $p_{14}$, $p_{22}$, and $p_{24}$ are given in Table I as functions of $c$, $R$, and $\sigma$.

The expected value of the multiplier output at time infinity can now be written in terms of the steady-state probabilities $w_i$ and the transition probabilities $p_{ij}$.

$$E[v_\infty] = w_1[p_{11} - p_{12} - p_{13} + p_{14}] + w_2[p_{22} + p_{23} - p_{21} - p_{24}]$$
$$+ w_3[p_{32} + p_{33} - p_{31} - p_{34}] + w_4[p_{41} + p_{44} - p_{42} - p_{43}] - c. \tag{40}$$

Again some algebraic manipulation yields the result

$$E[v_\infty] = \frac{R[1 - p_{14} - p_{24} - p_{22} - p_{12}] + 2[p_{14} - p_{12}] + 4[p_{22}p_{12} - p_{24}p_{14}]}{1 - p_{22} + p_{12} - p_{24} + p_{14}} - c. \tag{41}$$

The value of the tap gain at time infinity can be found by trial and error. A value of $c$ is assumed, the transition probabilities are computed and $E[v_\infty]$ is found. The value of $c$ for which $E[v_\infty] = 0$ is $c_\infty$. Notice that under suitable assumptions $E[v_\infty]$ gives the rate of change of the coefficient $c$ in the dynamic action of the system.

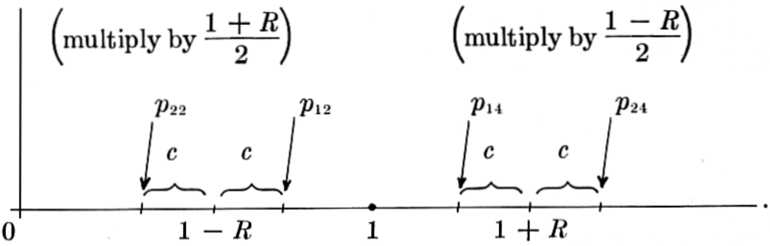
The probability of error after the system has settled is simply the probability that $\bar{\alpha}_\infty$ is in a state where $\beta_\infty = -1$, which is simply $(w_2 + w_4)$.

$$P_e = \frac{p_{12} + p_{14}}{1 - p_{22} + p_{12} - p_{24} + p_{14}}. \tag{42}$$

The transition probabilities here must be computed using $c_\infty$.

Expressions (41) and (42) have been written in terms of only those transition probabilities which involve errors. Thus, as $\sigma \to 0$, each of the transition probabilities in (41) and (42) approaches zero,

$c_\infty \to R$, and $P_e \to 0$. Each of these probabilities can be visualized as the probability that the noise (zero mean, variance $\sigma^2$) is greater than the one of these four thresholds:




Thus $p_{24}$ is the smallest transition probability, while $p_{22}$ is the largest.

If the transition probabilities are small, it can be seen from equation (42) that $P_e$ is principally determined by $(p_{12} + p_{14})$, which is minimized by $c = R$. Also we notice from equation (42) that the tap gain $c$ approaches $R$ very closely for small transition probabilities. In general, however, $c = R$ will not be the best setting to minimize the error probability in equation (42), nor is it the setting to which the loop settles. Unfortunately it appears that these are not compensating offsets. For example, in Figure 6 we have plotted $P_e$ and $E[v_\infty]$ against $c$, for a case in which $R = 0.4$ and $\sigma = 0.4$. Although neither

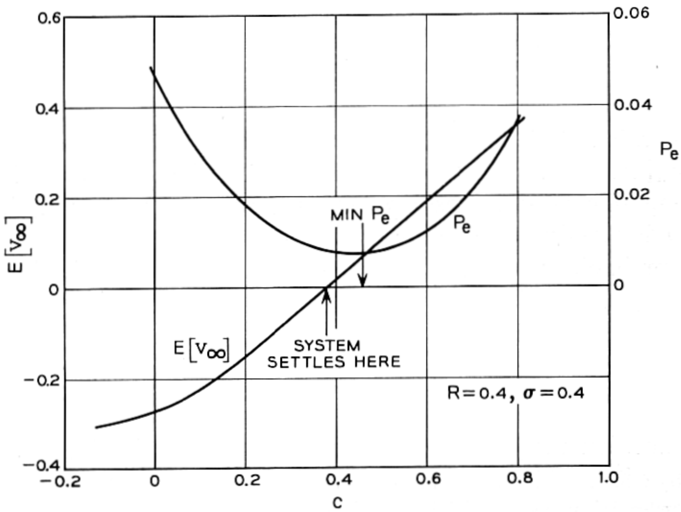



Fig. 6 — Probability of error and $E[v\infty]$ vs receiver tap gain $c$.

effect is very significant, it can be seen that the system settles ($E[v_\infty] = 0$) for a value of $c$ somewhat smaller than $R$, while the minimum error probability is obtained at a value of $c$ somewhat larger than $R$.

In all but the most severe noise conditions the approximation of $c_\infty = R$ would be satisfactory and we would have

$$P_e|_{c=R} = \frac{Q\left(\frac{1}{\sigma}\right)}{1 - \left(\frac{1+R}{2}\right)Q\left(\frac{1-2R}{\sigma}\right) - \left(\frac{1-R}{2}\right)Q\left(\frac{1+2R}{\sigma}\right) + Q\left(\frac{1}{\sigma}\right)}. \tag{43}$$

But $Q(1/\sigma)$ is the probability of error in the original system (no redundancy removal). If this probability, called $P_{e0}$, is small, then $Q(1 + 2R/\sigma)$ is much smaller and we have the very good approximation

$$P_e|_{P_{e0}\text{ small}} \cong \frac{P_{e0}}{\left[1 - \left(\frac{1+R}{2}\right)Q\left(\frac{1-2R}{\sigma}\right)\right]}. \tag{44}$$

The factor in the denominator gives the amplification of the original error rate due to error propagation. Finally if $R > 1/2$, then $Q(1 - 2R/\sigma)$ approaches unity and we get the severe dependence upon $R$

$$P_e|_{\substack{P_{e0}\text{ small}\\ R>\frac{1}{2}}} \cong \frac{2P_{e0}}{1-R}. \tag{45}$$

The most significant aspect of the error propagation behavior of the circuit is that the redundancy removal and restoration system has impressed the statistics of the input data (Markov here) upon the error statistics of the output. It is clear that this philosophy would hold in general. In the case of highly correlated input we would end with highly correlated errors. The problems of error control could be made quite severe in this manner.

## VI. EXPERIMENTAL RESULTS

A three-tap, adaptive transmitter and a similar receiver were designed and constructed by V. G. Koll. The system was designed for binary data transmission so that the multipliers in Figure 3 became polarity switches, while the delay line took the form of a shift register. The filters $W(s)$ consisted of simple RC low pass sections followed by integrators, that is,

$$W(s) = \frac{\alpha}{s(s+\alpha)}. \tag{46}$$

With this choice of smoothing, the steady-state error for a periodic input (period 3 or less here) was zero. It was in fact observed that during the transmission of periodic data the transmitter could be disconnected with no effect on the received data pattern.

The input data for the system was obtained by passing white Gaussian noise through a variable cutoff, low pass filter. If we assume an ideal low pass filter, with cutoff frequency $W$ Hz, then the autocorrelation function of the filter output is

$$R_1(\tau) = 2N_0W\left[\frac{\sin 2\pi W\tau}{2\pi W\tau}\right]. \tag{47}$$

This voltage is then sampled at rate $(1/T)$ and subjected to infinite clipping so as to produce the correlated input bits. Van Vleck and Middleton[19] show that the resulting autocorrelation is

$$R(n) = \frac{2}{\pi}\sin^{-1}\left[\frac{\sin 2\pi nWT}{2\pi nWT}\right]. \tag{48}$$

For a filter cutoff of $1/2T$ Hz the data is uncorrelated. By decreasing the filter cutoff frequency the redundancy in the data can be increased.

The action of the adaptive redundancy remover is shown in Figure 7 for two different values of filter cutoff. Notice that as the redundancy is increased the transmitted waveform has longer periods of near zero voltage where predictability is good and occasional peaks where the predictor is "surprised." Except for a few minor discontinuities the reconstructed signal before slicing at the receiver is the same as the original input waveform at the transmitter. The relative power saving as a function of filter cutoff is shown in Figure 8.

In order to predict system performance in Gaussian noise we make the crude approximation that the input process is Markov with $R(1)$ as given in equation (48). According to this approximation the transmitted power should be $1 - R(1)^2$. This value is also shown in Figure 8 in comparison with the actual measured power output. Since the exact correlation function is known, the theoretical signal power output could be computed precisely through equation (4). However, we have no corresponding means of computing the degree of error propagation

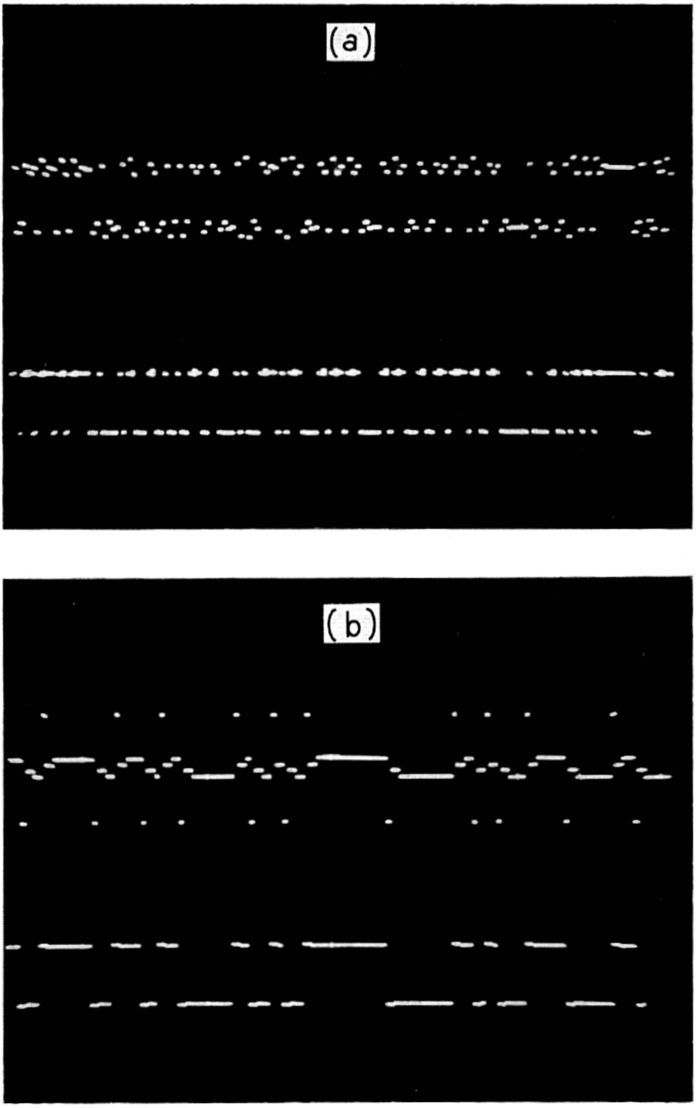



Fig. 7 — Transmitted and reconstructed signals. (a) Filter cutoff $\omega T = 0.4$ [little redundancy, $R(1) = 0.15$]. (b) Filter cutoff $\omega T = 0.1$ [moderate redundancy, $R(1) = 0.77$].

for the non-Markov source. The approximate curve of signal power in Figure 8 is shown only as a way of evaluating the Markov approximation for later use in predicting error propagation values.

Bandlimited white Gaussian noise was added to the transmitted signal, and error rates were experimentally determined by V. G. Koll at a number of filter cutoff (redundancy) positions. The results of these tests are shown in Figure 9 in curves of probability of error versus signal-to-noise ratio. Beside these measured curves have been plotted theoretically computed curves which are based on the Markov approximation and on the use of equation (43) for $P_e$.

Although all necessary information for performance determination is contained in Figure 9, it is instructive to plot two additional curves of probability of error versus filter cutoff. These curves are shown in Figure 10. In one curve the transmitter and receiver gains are held constant so that the line power decreases according to the curve of Figure 8 while the probability of error increases with increasing redundancy because of the effects of error propagation. In the other curve of Figure 10 the transmitter and receiver gains have been adjusted with increasing redundancy so as to hold line power constant. In this case the probability of error decreases with increasing redundancy.

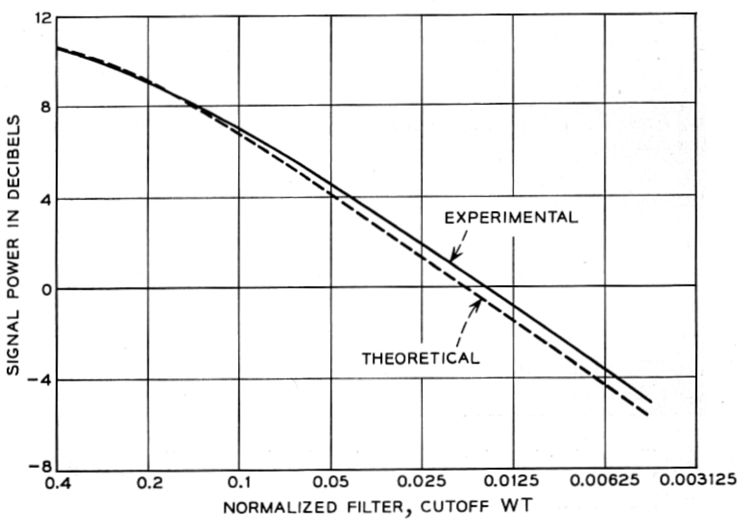



Fig. 8 — Signal power saving by redundancy removal.

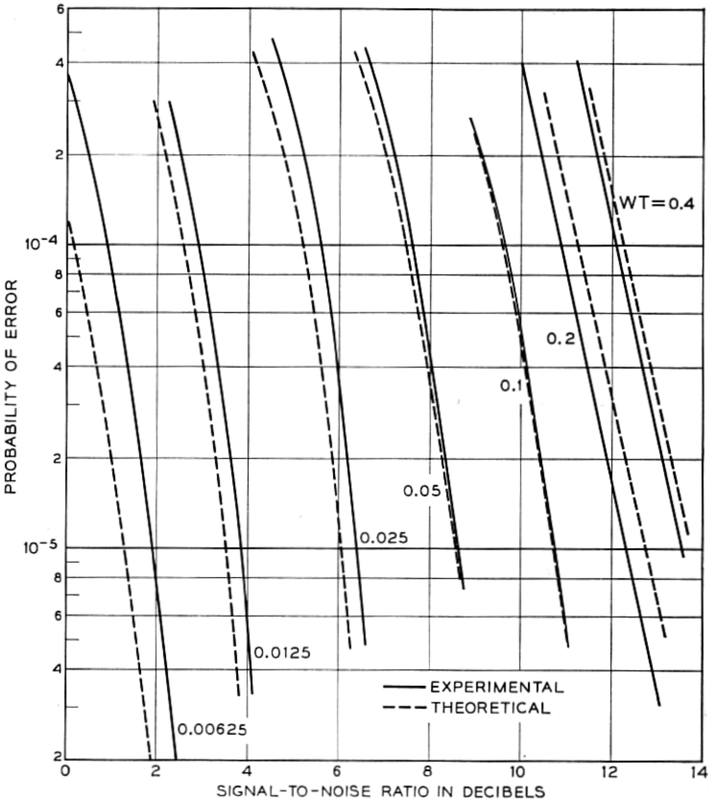



Fig. 9 — Performance of redundancy removal system at various values of normalized filter cutoff $\omega T$.

## VII. CONCLUSION

We have advanced two main points. First we suggest the possibility of using an easily-implemented adaptive predictor for data compression systems. Second, we investigated the use of this adaptive predictor in digital transmission.

We have seen that the predictor can be used to increase transmission efficiency for redundant data either by decreasing signal power for a given error rate or by decreasing probability of error for a given signal power. Although the required circuitry for the digital application is quite simple, it is nearly impossible to make an economic evaluation of the system because of the complete lack of knowledge of the prevalence and degree of redundancy in customer input data.

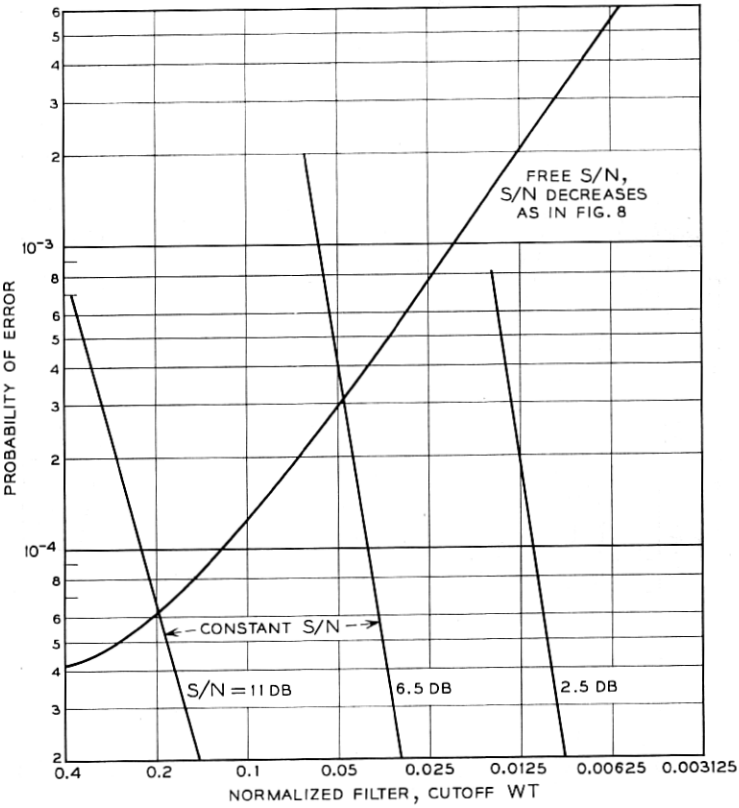



Fig. 10 — Probability of error vs filter cutoff for constant and for free S/N.

## VIII. ACKNOWLEDGEMENT

The author is indebted to V. G. Koll who designed, constructed, and tested the three-tap experimental ssytem. Mr. Koll also made the photos and the experimental performance curves in Figures 7 through 10.

## REFERENCES

1. Oliver, B. N., "Efficient Coding," B.S.T.J., *31,* No. 4 (July 1952), pp. 724–750.
2. Kretzmer, E. R., "Statistics of Television Signals," B.S.T.J., *31,* No. 4 (July 1952), pp. 751–763.
3. Harrison, C. W., "Experiments with Linear Prediction in Television," B.S.T.J., *31,* No. 4 (July 1952), pp. 764–783.
4. Elias, P., "Predictive Coding," IRE Trans. Inform. Theory, *IT*-1, No. 1 (March 1955), pp. 16–33.

5. Wiener, N., *Extrapolation, Interpolation and Smoothing of Stationary Time Series,* Cambridge, Mass.: MIT Press, 1949.
6. Kortman, C. M., "Redundancy Reduction—A Practical Method of Data Compression," Proc. IEEE, *55,* No. 3 (March 1967), pp. 253–263.
7. Davison, L. D., "Theory of Adaptive Data Compression," in *Recent Advances in Communication Systems,* vol. 2, ed. A. V. Balakrishnan, New York: Academic Press, 1966.
8. O'Neal, J. B., "Predictive Quantizing Systems," B.S.T.J., *45,* No. 5 (May-June 1966), pp. 689–721.
9. Savage, J. E., "Some Simple Self-Synchronizing Digital Data Scramblers," B.S.T.J., *46,* No. 2 (February 1967), pp. 449–487.
10. Doob, J. L., *Stochastic Processes,* Chapter 12, New York: John Wiley and Sons, Inc., 1953.
11. Balakrishnan, A. V., "An Adaptive Non-Linear Data Predictor," 1962 Nat. Telemetering Conf. Proc.
12. Lucky, R. W. and Rudin, H. R., "An Automatic Equalizer for General-Purpose Communication Channels," B.S.T.J., *46,* No. 9 (November 1967), pp. 2179–2208.
13. Becker, F. K. and Rudin, H. R., "Application of Automatic Transversal Filters to the Problem of Echo Suppression," B.S.T.J., *45,* No. 10 (December 1966), pp. 1847–1850.
14. Sondhi, M. M. and Presti, A. J., "A Self-Adaptive Echo Canceller," B.S.T.J., *45,* No. 10 (December 1966), pp. 1851–1853.
15. Davenport, W. B. and Root, W. L., *An Introduction to the Theory of Random Signals and Noise,* New York: McGraw-Hill Book Co., Inc., 1958, p. 105.
16. Zador, P. L., "Error Probabilities in Data System Pulse Regenerator with DC Restoration," B.S.T.J., *45,* No. 6 (July-August 1966), pp. 979–984.
17. Aaron, M. R. and Simon, M. K., "Approximation of the Error Probability in a Regenerative Repeater with Quantized Feedback," B.S.T.J., *45,* No. 10 (December 1966), pp. 1845–1847.
18. Kemeny, J. G. and Snell, J. L., *Finite Markov Chains,* New York: D. Van Nostrand Co., Inc., 1959.
19. Van Vleck, J. H. and Middleton, D., "The Spectrum of Clipped Noise," Proc. IEEE, *54,* No. 1 (January 1966), pp. 2–19 (reprint).