

A coding scheme for additive noise channels with feedback, Part I: No bandwidth constraint

Citation for published version (APA):

Schalkwijk, J. P. M., & Kailath, T. (1968). A coding scheme for additive noise channels with feedback, Part I: No bandwidth constraint. *IEEE Transactions on Information Theory*, IT-12(2), 172-182.
<https://doi.org/10.1109/TIT.1966.1053879>

DOI:

[10.1109/TIT.1966.1053879](https://doi.org/10.1109/TIT.1966.1053879)

Document status and date:

Published: 01/01/1968

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

A Coding Scheme for Additive Noise Channels with Feedback—Part I: No Bandwidth Constraint

J. P. M. SCHALKWIJK, MEMBER, IEEE AND T. KAILATH, MEMBER, IEEE

Abstract—In some communication problems, it is a good assumption that the channel consists of an additive white Gaussian noise forward link and an essentially noiseless feedback link. In this paper, we study channels where no bandwidth constraint is placed on the transmitted signals. Such channels arise in space communications.

It is known that the availability of the feedback link cannot increase the channel capacity of the noisy forward link, but it can considerably reduce the coding effort required to achieve a given level of performance. We present a coding scheme that exploits the feedback to achieve considerable reductions in coding and decoding complexity and delay over what would be needed for comparable performance with the best known (simplex) codes for the one-way channel. Our scheme, which was motivated by the Robbins-Monro stochastic approximation technique, can also be used over channels where the additive noise is not Gaussian but is still independent from instant to instant. An extension of the scheme for channels with limited signal bandwidth is presented in a companion paper (Part II).

I. INTRODUCTION

IN CERTAIN COMMUNICATION problems we have the possibility of using a noiseless "feedback" link to improve communication over a noisy forward link. A good example is communication with a space satellite—the power in the ground-to-satellite direction can be so much larger than in the reverse direction that the first link can be taken to be an (essentially) noiseless link. Similar possibilities may also arise in special terrestrial situations.

It seems reasonable that the availability of a noiseless feedback link should substantially improve communication over the noisy forward link. Therefore, Shannon's result [4], Theorem 6, that the channel capacity of a *memoryless* noisy channel is not increased by noiseless feedback, is rather surprising. Still, some advantage should accrue from the presence of a noiseless feedback link and, in fact, the advantage is that noiseless feedback enables a substantial reduction in the complexity of coding and decoding required to achieve a given performance over the noisy link.

In this paper, we shall illustrate the simplifications obtained when the noisy link is an additive white Gaussian noise channel operated under an average power con-

straint. No restrictions are placed on the usable signal bandwidth. Such channels seem to be typical of those in space communications. In terrestrial communications we are often forced to impose bandwidth limitations on the transmitted signals. Channels with bandwidth constraints are discussed in a companion paper [17]. The communication scheme we shall present can also be used over non-Gaussian white noise channels. However, it is difficult to evaluate the effect of the feedback link in such cases because very few results are available for non-Gaussian (one-way) channels. We shall now briefly review the known results for the Gaussian case.

A. Additive White Gaussian Noise Channels

We shall assume that the noise in the channel is Gaussian and white with a (two-sided) spectral density $N_0/2$. The transmitted signals are required to have an average power P_{av} but no constraints are imposed on their bandwidth and peak power.

For this channel, the channel capacity is given by (e.g., Fano [9], Ch. VI)

$$\begin{aligned} C &= \frac{P_{av}}{N_0 \ln 2} \text{ bits/second} \\ &= \frac{P_{av}}{N_0} \text{ nats/second.} \end{aligned} \quad (1)$$

If one of M messages is to be transmitted over such a channel, the best code is universally believed to be a "regular-simplex" set of codewords (i.e., a set of M equal-energy signals with mutual cross-correlations of $-1/(M-1)$). For large M , an orthogonal signal set (for which the cross-correlations are zero rather than $-1/(M-1)$) performs almost as well. The ideal receiver for such signals is a bank of M correlation detectors, whose outputs are scanned to determine the correlator yielding the largest output. The error probability for an orthogonal (or simplex) signal set has been evaluated numerically for values of M from 2 to 10^6 . For larger values of M , the following asymptotic expression can be used. If T is the duration of each of the M signals, assumed equally likely a priori, then (cf. Fano [9], Chapter VI, and Zetterberg [12])

$$P_{e,orth}(M, T) = \frac{\text{constant}}{T^6} e^{-TE(R)} \quad (2)$$

This work was supported by NASA NSG 337, Tri Service Nonr 225 (83), and AFOSR Contract AF 49(638)-1517.

T. Kailath is with the Department of Electrical Engineering, Stanford University, Stanford, Calif.

J. P. M. Schalkwijk is with the Applied Research Laboratory, Sylvania Electronic Systems, a division of Sylvania Electric Products Inc., Waltham, Mass.

where

$$E(R) = \begin{cases} C/2 - R, & 0 \leq R \leq C/4 \\ (\sqrt{C} - \sqrt{R})^2, & \frac{C}{4} \leq R \leq C \end{cases} \quad (3)$$

$$1 \leq \beta \leq 2$$

R = the signalling rate = $\ln M/T$ nats/second

C = the channel capacity = P_{av}/N_0 nats/second.

Equation (2) shows that the error probability for orthogonal codes decreases essentially exponentially with T . As a result, for large T (low P_e), the choice of a suitable pair of values R and T to achieve a given P_e is essentially determined by the quantity $E(R)$. Equations (2) and (3) specify the "tradeoffs" that can be made between the signal duration T and the signaling rate R —for rates near channel capacity, $E(R)$ is small and we need a large T to achieve a given P_e ; if we are required to use a small value of T , the rate R must be suitably reduced.

We can now present the results we have obtained by assuming that a noiseless feedback link is available.

B. Summary of Our Results

We have developed a coding scheme that exploits the presence of the noiseless feedback link. The scheme was suggested by the Robbins-Monro stochastic approximation procedure for determining the zero of a function by noisy observations of its values at chosen points [2]. While we do not know if this coding scheme is "optimum," it has the virtue of great simplicity both in encoding and decoding. It also enables us to achieve the channel capacity P_{av}/N_0 of the white Gaussian channel (the capacity is the same, as we mentioned above, whether or not a noiseless feedback link is available). And most important it provides a dramatic reduction in the rate at which the error probability varies with signaling duration. Thus, for our scheme, we have

$$P_{e,fb} = \frac{1}{\sqrt{6\pi} \frac{C}{R} e^{2(C-R)T}} \exp \left[-\frac{3C}{2R} e^{2(C-R)T} \right] \quad (4)$$

showing that the error probability goes down much faster than exponentially with T , which is how the error probability behaves for one-way channels [cf. (2) and (3)]. As a sample comparison between the feedback and non-feedback cases, consider the values of T required to achieve, for example,

$$P_e = 10^{-7}, \quad R = 0.8C, \quad C = 1 \text{ bit/second.}$$

We have

$$T_{f.b.} = 15 \quad \text{and} \quad T_{orth} = 2030.$$

It is natural to wonder how our results are affected by delay and noise in the feedback link. The effect of the delay is merely a small increase in the error probability.

The effects of noise are more serious—we find that, provided the noise is smaller than a certain threshold value, the error probability is essentially unaffected; however, with noisier feedback, our scheme deteriorates rapidly and other coding techniques must be devised.

Our coding scheme does *not* depend upon the Gaussian nature of the additive white noise. It can be applied to any additive white¹ noise channel and will enable us to signal, with arbitrarily low error probability, at rates up to P_{av}/N_0 nats/second. (We recall that $N_0/2$ is defined as the (two-sided) spectral density of the white noise.) This result thus yields a lower bound of P_{av}/N_0 for the channel capacity of additive white noise channels of spectral density $N_0/2$ and transmitter power P_{av} . The actual capacity of such channels may be much larger than this, but the capacity is usually too complicated to evaluate analytically. The capacity is the same with or without feedback. The error probabilities will, however, be considerably different. No results on the error probabilities seem to be available for one-way non-Gaussian white noise channels; when a noiseless feedback channel is available, the expression given above for $P_{e,fb}$ will continue to be valid, for large T , for non-Gaussian channels. Some further remarks on such channels are made in Section IV.

We should also mention that our coding scheme does not depend upon knowledge of $N_0/2$ (the noise spectral density) for its operation. Of course, this knowledge will be necessary for evaluating the performance of the scheme.

Before proceeding to the derivation of our results, we shall give a brief discussion of related work.

C. Other Studies of Noiseless Feedback Channels

A general discussion of feedback communication systems, with reference to earlier work by Chang and others, is given by Green [10] who distinguishes between post- and predecision feedback systems. In postdecision feedback systems the transmitter is informed only about the receiver's decision; in predecision feedback systems, the state of uncertainty of the receiver as to which message was sent is fed back. Postdecision feedback systems require less capacity in the backward direction; however, the improvement over one-way transmission will also be less than that obtainable with predecision feedback.

Viterbi [17] discusses a postdecision feedback system for the white Gaussian noise channel. The receiver computes the likelihood ratio as a function of time and makes a decision when the value of the likelihood ratio crosses one of a pair of thresholds. The transmitter is informed by means of postdecision feedback that the receiver has made a decision, and it then starts sending the next message. For rates higher than half the channel capacity,

¹ By white noise, we shall mean noise whose values at any two instants of time are statistically independent.

the reliability is increased roughly by a factor of four as compared to one-way communication.

Turin [15] has a predecision feedback scheme applying to the white Gaussian noise channel, and giving an even greater improvement over one-way communication than Viterbi's scheme does. The receiver again computes the likelihood ratio as a function of time, but now the value of the likelihood ratio is fed back continuously to the transmitter. The transmitted signal is a function of the binary digit (that is, 0 or 1) being sent and of the value of the likelihood ratio, and is adjusted so as to make this ratio cross, as quickly as possible, one of a pair of decision thresholds. Average and peak power constraints on the transmitted signals are studied. The average time \bar{T} for deciding on a binary digit turns out to be

$$\bar{T} = (\ln 2)(P_{av}/N_0)^{-1} \text{ seconds,}$$

where P_{av} is the average power and N_0 is the (one-sided) noise power spectral density. The possibility of error P_e vanishes if infinite peak power and infinite bandwidth are allowed. Hence, a rate is achieved that is equal to the channel capacity

$$C = \frac{P_{av}}{N_0} \text{ nats/second.}$$

In this scheme, the actual time required to make a decision is variable (though the mean value is \bar{T}). The variance and other parameters of this variable time do not appear to be readily computable. As opposed to this, our feedback scheme is a "block" scheme, with a decision being made after a preassigned interval, the interval being determined by the desired rate and error probability.

II. THE CODING SCHEME AND ITS EVALUATION

Our coding scheme was motivated by the Robbins-Monro [2] stochastic approximation procedure which we shall describe briefly.

A. The Robbins-Monro Procedure

Consider the situation indicated in Fig. 1. One wants to determine θ , a zero of $F(x)$, without knowing the shape of the function $F(x)$. It is possible to measure the values of the function $F(x)$ at any desired point x . The observations are noisy, however, so that instead of $F(x)$ one obtains $Y(x) = F(x) + Z$, where Z is some additive disturbance. The "noise" Z is assumed to be independent and identically distributed from trial to trial. To estimate θ , Robbins and Monro proposed the following recursive scheme. Start with an arbitrary initial guess X_1 and make successive guesses according to

$$X_{n+1} = X_n - a_n Y_n(X_n), \quad n = 1, 2, \dots$$

For the procedure to work, that is, for X_{n+1} to tend to θ , the coefficients a_n must satisfy $a_n \geq 0$, $\sum a_n = \infty$, and $\sum a_n^2 < \infty$. A sequence $\{a_n\}$ fulfilling these requirements is $a_n = 1/an$, $a > 0$.

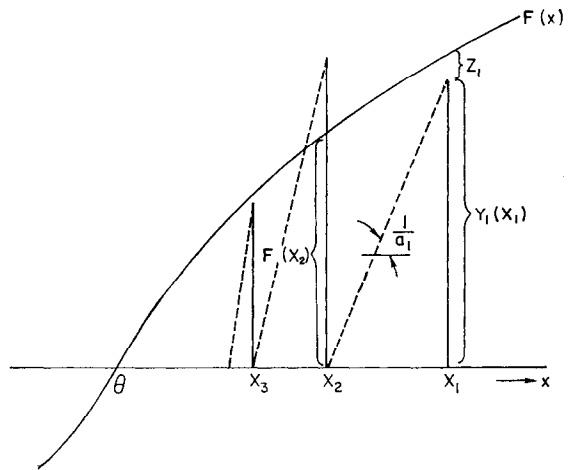


Fig. 1. The Robbins-Monro procedure.

The following additional requirements are needed on the function $F(x)$ and on Z .

- 1) $F(x) \geq 0$ according to $x \geq \theta$.
- 2) $\inf \{|F(x)|; \epsilon < |x - \theta| < 1/\epsilon\} > 0$ for all $\epsilon > 0$.
- 3) $|F(x)| \leq K_1 |x - \theta| + K_2$, where K_1 and K_2 are constants.
- 4) If $\sigma^2(x) = E[Y(x) - F(x)]^2$, then $\sup_x \sigma^2(x) = \sigma^2 < \infty$.

With these requirements, the following theorem can be established.

Theorem: When the above conditions on the a_n , the $F(x)$, and the Z_n are met, $Z_n \rightarrow \theta$ almost surely; and furthermore, if $E|X_1|^2 < \infty$, then $E|X_n - \theta|^2 \rightarrow 0$.

Robbins and Monro proved the convergence in mean square. The "convergence almost surely" was first proved by Wolfowitz [5]. A good proof of the preceding theorem is Dvoretzky's [3], where several types of stochastic approximation procedures are treated in a unified manner.²

The Robbins-Monro procedure is nonparametric, that is, no assumptions concerning the distribution of the additive disturbance, except for zero mean and finite variance, are necessary. However, it was shown by Sacks [6] that $\sqrt{n}(X_{n+1} - \theta)$ is normally distributed for large n . In fact, let the following assumptions, which complement the earlier requirements, be fulfilled.

- 5) $\sigma^2(x) \rightarrow \sigma^2(\theta)$ as $x \rightarrow \theta$.
- 6) $F(x) = \alpha(x - \theta) + \delta(x)$, where $\alpha > 0$ and $\delta(x) = O(|x - \theta|^{1+\rho})$, $\rho > 0$.
- 7) There exist $t > 0$ and $\delta > 0$ such that $\sup \{E|Z(x)|^{2+\delta}; |x - \theta| \leq t\} < \infty$.
- 8) $a_n = 1/an$, and $2a > a$.

Then we have the theorem [6]: Fulfillment of all the conditions (1-8) yields

$$\sqrt{n}(X_{n+1} - \theta) \sim N\left[0, \frac{\sigma^2}{a(2a - a)}\right].$$

This result will be used presently.

² A recent general survey of stochastic approximation methods is given by Venter [14].

B. An Equivalent Discrete-Time Channel

To apply the stochastic approximation procedure to the communication channel, we shall need to obtain a discrete-time equivalent of the additive white Gaussian noise channel. This can be done in many ways. We shall present a mathematically convenient method—later we shall comment on its physical viability.

To obtain a discrete-time equivalent, we shall assume that the message information is transmitted by suitably modulating the amplitude of a known basic waveform, $\phi(t)$. The signal in the channel (see Fig. 2) will thus be of the form

$$s(t) = \sum_i x_i \phi(t - i\Delta), \quad i = 0, 1, \dots$$

where Δ is a time interval that will be suitably chosen later. We shall require the basic waveform $\phi(t)$ to have unit energy and to be orthogonal for shifts Δ , i.e., $\phi(t)$ should satisfy

$$\int \phi(t - i\Delta) \phi(t - j\Delta) dt = \delta_{ij}. \quad (5)$$

The integral extends over all values of t for which the integrand is different from zero.

Reception will be achieved using a filter matched to $\phi(t)$, that is, $h(t) = \phi(-t)$. The output of this matched filter at $t = i\Delta$, $i = 1, 2, \dots$, will be the sequence $\{Y_i(X_i)\}$ where $Y_i(X_i) = X_i + Z_i$, and

$$Z_i = \int n(t) \phi(t - i\Delta) dt.$$

It can be easily checked that the $\{Z_i\}$ will be uncorrelated zero mean random variables with

$$E[Z_i Z_j] = \frac{N_0}{2} \delta_{ij}.$$

When the additive noise is Gaussian, the $\{Z_i\}$ will be Gaussian and, therefore, also independent. In the Gaussian case, it is easy to see that the discrete-time channel thus obtained (where a sequence of numbers $\{X_i\}$ is transmitted and sequence $\{Y_i(X_i) = X_i + Z_i\}$ is received) is completely equivalent to the original continuous-time channel. This follows from the fact that the matched filter for Gaussian white noise channels computes the likelihood ratio, which is a sufficient statistic and therefore preserves all the information in the received waveform that is relevant to the decision making process.

Finally, we note that by virtue of the orthonormality of $\phi(t - i\Delta)$ and $\phi(t - j\Delta)$, $i \neq j$, the transmitted energy in $s(t) = \sum x_i \phi(t - i\Delta)$ is $\sum x_i^2$.

We can now describe our coding scheme.

C. The Coding Scheme

The transmitter has to send one of M possible messages to a receiver. A noiseless feedback channel is available. We shall proceed as follows (see also Fig. 3).

Divide the unit interval into M disjoint, equal-length "message intervals." Pick as the "message point" θ ,

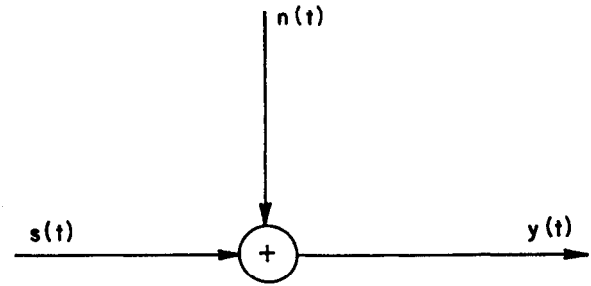


Fig. 2. Model for the additive noise channel.

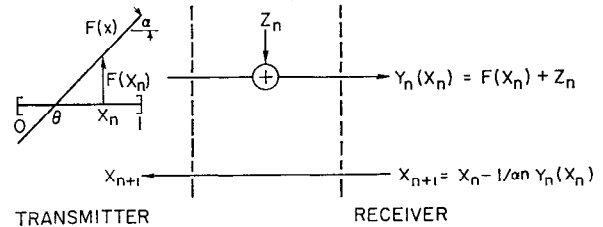


Fig. 3. Proposed coding scheme for wideband signals.

the midpoint of the message interval corresponding to the particular message being transmitted. Through this message point θ , put a straight line $F(x) = \alpha(x - \theta)$, with slope $\alpha > 0$. Start out with $X_1 = 0.5$ and send to the receiver the "number" $F(X_1) = \alpha(X_1 - \theta)$, as discussed in Section II-B. At the receiver one obtains the "number" $Y_1(X_1) = \alpha(X_1 - \theta) + Z_1$, where Z_1 is a Gaussian random variable with zero mean and variance $N_0/2 = \sigma^2$, say. The receiver now computes $X_2 = X_1 - (a/1)Y_1(X_1)$, where a is a constant which will be specified soon, and retransmits this value to the transmitter which then sends $F(X_2) = \alpha(X_2 - \theta)$. In general, one receives $Y_n(X_n) = F(X_n) + Z_n$ and computes $X_{n+1} = X_n - (a/n)Y_n(X_n)$. The number X_{n+1} is sent back to the transmitter, which then will send $F(X_{n+1}) = \alpha(X_{n+1} - \theta)$.

From Sacks' theorem [6], quoted earlier, on asymptotic distributions of stochastic approximation procedures, it follows that the best value for a is $a = 1/\alpha$ and that in this case $\sqrt{n}(X_{n+1} - \theta)$ converges in distribution to a normal random variable with zero mean and variance $(\sigma/\alpha)^2$.

In the Gaussian case, the distribution of $(X_{n+1} - \theta)$ can be computed directly for any n without reference to Sacks' theorem. With $a_n = 1/\alpha n$, the recursion relation

$$X_{n+1} = X_n - \frac{1}{\alpha n} Y_n(X_n), \quad Y_n(X_n) = \alpha(X_n - \theta) + Z_n$$

is easily solved to yield

$$X_{n+1} = \theta - \frac{1}{\alpha n} \sum_{i=1}^n Z_i.$$

Since the Z_i are independent, $N(0, \sigma^2)$, X_{n+1} will be Gaussian with mean θ and variance $\sigma^2/\alpha^2 n$. We may also point out here that X_{n+1} is (in the Gaussian case) the maximum likelihood estimate of θ , given $Y_1(X_1), \dots, Y_n(X_n)$. The estimate X_{n+1} is also unbiased and efficient (i.e., it achieves the Cramér-Rao lower bound on the

variance of any estimate of θ). This interpretation of X_{n+1} will be used in Section III-C; it also leads to the coding algorithm for the band-limited signal case (see Schalkwijk [16]).

Now suppose that N iterations are made before the receiver makes its decision as to which of the M messages was sent. What is the probability of error? The situation is presented in Fig. 4. After N iterations,

$$X_{N+1} \sim N(\theta, \sigma^2/\alpha^2 N).$$

The length of the message interval is $1/M$. Hence, the probability of X_{N+1} lying outside the correct message interval is

$$P_e = 2 \operatorname{erfc} \left(\frac{\frac{1}{2} M^{-1}}{\sigma/\alpha \sqrt{N}} \right), \quad (6)$$

where

$$\operatorname{erfc} x = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt.$$

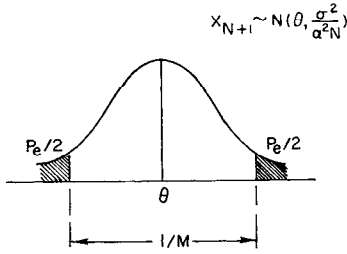


Fig. 4. The error probability is the shaded area.

D. Achieving Channel Capacity

Equation (6) shows, not unexpectedly, that P_e can be driven to zero by increasing the number N of iterations. However, if this is done without increasing M , the signaling rate (which we shall define as $R = \ln M/T$ nats/second) will go to zero. This tradeoff of rate for reliability had seemed quite natural and inevitable, until Shannon pointed out 1) that a constant rate R could be maintained if M was increased along with T (which is monotonically related to N), according to the formula $M = e^{TR}$, 2) that if R were not too high, i.e., M did not increase too rapidly with T (or N), then the degradation in performance introduced by increasing M could be more than compensated for by the good effects of increasing T , and therefore, 3) that for such rates, arbitrarily low error probability could be achieved by taking T (or N) large enough. The largest such rate, at which arbitrarily low error probabilities can be achieved, was called the channel capacity [11].

To apply Shannon's observations to our problem, we inquire how rapidly we can increase M with N while still enabling the probability of error to vanish for increasing N . The distribution in Fig. 4 squeezes in at a rate $1/\sqrt{N}$ (this being the standard deviation). Therefore, if $1/M$ is decreased at a rate slightly less than $1/\sqrt{N}$, one can "trap" the Gaussian distribution within the message interval and thus make the probability of correct detection go to unity. We therefore set

$$M(N) = N^{1/2(1-\epsilon)}. \quad (7)$$

The consequent error probability is

$$P_e = 2 \operatorname{erfc} \left[\frac{\alpha}{2\sigma} N^{\epsilon/2} \right] \quad (8)$$

and as $N \rightarrow \infty$,

$$\lim_{N \rightarrow \infty} P_e = \begin{cases} 0 & \text{for } \epsilon > 0 \\ 1 & \text{for } \epsilon < 0. \end{cases}$$

The critical rate (determined by $\epsilon = 0$) in nats per second will be

$$R_{\text{crit}} = \left[\frac{\ln M(N)}{T} \right]_{\epsilon=0} = \frac{\ln N}{2T} \text{ nats/second}. \quad (9)$$

However, in order to keep R_{crit} finite as $T \rightarrow \infty$, N must grow exponentially with T . Thus, setting $N = e^{2AT}$, with A being some constant, gives

$$R_{\text{crit}} = \frac{\ln N}{2T} = A \text{ nats/second}. \quad (10)$$

But what prevents us from choosing A arbitrarily large and thereby achieving an arbitrarily high rate of error-free transmission? The answer is that A is limited by the average power constraint P_{av} , which has not as yet been taken into account. The effect of P_{av} on A can be seen by calculating the average transmitted power with the proposed scheme. The transmitted power will depend upon the additive noise. Therefore, using $E[\cdot]$ to denote averaging over the noise process gives

$$P_{\text{av}}(N) = \frac{1}{T} E \left[\alpha^2 (X_1 - \theta)^2 + \sum_{i=1}^{N-1} \alpha^2 (X_{i+1} - \theta)^2 \right]. \quad (11)$$

If we assume a uniform prior distribution for the message point θ , $E(X_1 - \theta)^2$ will be $1/12$. Furthermore, we have seen that $E(X_{i+1} - \theta)^2 = \sigma^2/\alpha^2 i$. Substitution in the formula for the average power leads to [also using (10)]

$$P_{\text{av}}(N) = \frac{2A}{\ln N} \left(\frac{\alpha^2}{12} + \sigma^2 \sum_{i=1}^{N-1} \frac{1}{i} \right). \quad (12)$$

Therefore,

$$A = \frac{P_{\text{av}}}{N_0} \left[\frac{\alpha^2}{6N_0 \ln N} + \frac{\sum_{i=1}^N \frac{1}{i}}{\ln N} \right]^{-1} < \frac{P_{\text{av}}}{N_0}$$

$$\lim_{N \rightarrow \infty} P_{\text{av}}(N) = 2\sigma^2 A = N_0 A \quad \text{or} \quad A = \frac{P_{\text{av}}}{N_0}. \quad (13)$$

Therefore, A cannot be arbitrarily large but is constrained to be less than or equal to P_{av}/N_0 . The critical rate, therefore, is

$$R_{\text{crit}} = A = \frac{P_{\text{av}}}{N_0} \text{ nats/second}$$

which is just the channel capacity of the (one-way) additive white Gaussian noise channel.

It may be useful to view this result in the following way. The noise variance in our scheme goes down as $1/N$, which is no faster than the rate at which the noise variance goes down with a simple repetitive coding scheme

(i.e., sending each message N times). However, with simple repetition the signal power increases with N (and would therefore violate any average power constraint for large enough N), while with our scheme the transmitted power decreases suitably with N so as to meet the average power constraint. Figure 5 is a sketch of the behavior of the expected instantaneous transmitter power as a function of time.

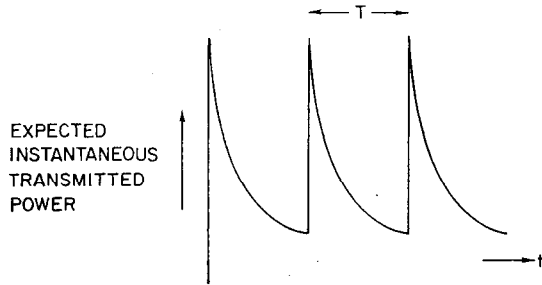


Fig. 5. The expected instantaneous transmitted power as a function of time.

Our feedback scheme cannot signal at a higher rate than is possible for the one-way channel, but it can achieve the same performance with considerably less coding and decoding complexity. As far as the coding goes, our recursive scheme for determining the transmitted waveform is somewhat simpler than the scheme for generating orthogonal waveforms. However, the real simplicity is in the decoding: with M orthogonal signals, ideal decoding requires searching for the largest of M matched filter outputs, a laborious operation for large M ; in our scheme, we just have to check in what amplitude range the output of a single matched filter lies.

We can make our complexity comparisons more quantitative by comparing the expressions for the error probability with and without feedback. Before making this comparison we note that we have not as yet specified the slope α of our straight-line coding function. As far as achieving the rate P_{av}/N_0 , any value of α will do. However, in evaluating the error probability for a fixed number of iterations N there is an optimum value of α .

E. Optimum P_e for Finite N

The value of the slope α , given R/C and given N , that minimizes the probability of error is easily determined. From (8), minimizing the probability of error is equivalent to maximizing

$$\frac{\alpha^2}{4\sigma^2} N^\epsilon = \frac{\alpha^2}{2N_0} N^\epsilon.$$

Now, differentiating with respect to α^2 , one has for the optimum α ,

$$\frac{d}{d(\alpha^2/N_0)} \left(\frac{\alpha^2}{2N_0} N^\epsilon \right) = \frac{1}{2} N^\epsilon + \frac{\alpha^2}{2N_0} \frac{dN^\epsilon}{d\epsilon} \frac{d\epsilon}{d(\alpha^2/N_0)} = 0.$$

To compute $d\epsilon/d(\alpha^2/N_0)$, an expression for ϵ is needed. Using (7) to get $R = \ln M(N)/T = (1 - \epsilon)A$ and setting $\sigma^2 = N_0/2$, we get from (12),

$$R = (1 - \epsilon)A = (1 - \epsilon)C \left[\frac{\alpha^2}{6N_0} + \sum_{i=1}^{N-1} \frac{1}{i} \right]^{-1} \ln N$$

from which

$$\epsilon = 1 - \frac{R}{C} \left[\frac{\alpha^2}{6N_0} + \sum_{i=1}^{N-1} \frac{1}{i} \right] (\ln N)^{-1}. \quad (14)$$

Hence, $d\epsilon/[d(\alpha^2/N_0)] = -(R/6C)(\ln N)^{-1}$, and using this, we will have

$$\frac{d}{d(\alpha^2/N_0)} \left(\frac{\alpha^2}{2N_0} N^\epsilon \right) = \frac{1}{2} N^\epsilon - \frac{\alpha^2}{2N_0} N^\epsilon (\ln N) \frac{R}{C} \frac{1}{6 \ln N} = 0.$$

Therefore, the optimum value of α^2 , say α_0^2 , is

$$\alpha_0^2 = 6N_0 \left(\frac{R}{C} \right)^{-1}. \quad (15)$$

Substituting for α_0^2 in the formula for the probability of error we finally have

$$P_e = 2 \operatorname{erfc} \left[\left(3 \frac{C}{R} N^\epsilon \right)^{1/2} \right]. \quad (16)$$

Figure 6 gives curves for the probability of error as a function of the number N of iterations. The parameter R/C is the rate relative to channel capacity. The curves start at that value of N beyond which ϵ as given by (14) is positive. Note that for relative rates approaching unity, the number of transmissions per message becomes very high.

Equation (15) gives the optimum value of the slope α as a function of the relative rate R/C and the noise power spectral density $N_0/2$. Figure 7 shows curves of the probability of error vs. the slope squared relative to its optimum value.

We can also write down an asymptotic expression for the probability of error, similar to the expression quoted in Section I for orthogonal codes. From (8), and substituting $\sigma^2 = N_0/2$, the probability of error is

$$P_e = 2 \operatorname{erfc} \left[\left(\frac{\alpha^2}{2N_0} N^\epsilon \right)^{1/2} \right].$$

By using the optimum α_0^2 of α^2 given by (15) and using the well-known asymptotic formula for the *erfc* function, we obtain (asymptotically for large N)

$$P_e \approx \frac{\exp \left[-\frac{3}{2} \left(\frac{R}{C} \right)^{-1} N^\epsilon \right]}{\left[6\pi \left(\frac{R}{C} \right)^{-1} N^\epsilon \right]^{1/2}}.$$

Furthermore, $N = e^{2AT}$ and $R = (1 - \epsilon)A$, where A is asymptotically equal to C . We therefore have

$$P_e \approx \frac{\exp \left[-\frac{3C}{2R} e^{2(C-R)T} \right]}{\left[6\pi \frac{C}{R} e^{2(C-R)T} \right]^{1/2}}. \quad (17)$$

It will be convenient to make the comparison with orthogonal codes on the basis of a "blocklength L " in binary digits, which will be defined as follows. Let $2^L = M$.

After N iterations, M will be $N^{1/(1-\epsilon)}$, and hence, $L = \frac{1}{2}(1 - \epsilon) \log_2 N$. Figure 8 gives curves of the probability of error vs. the blocklength L .

Similar curves can be obtained for orthogonal codes by using the relations

$$\log_{10} P_e \approx -\frac{TE(R)}{\ln 10} \quad \text{and} \quad 2^L = M = e^{RT},$$

which yield

$$\log_{10} P_e \approx L \frac{E(R)}{R} \log_{10} 2. \quad (18)$$

This expression for $\log_{10} P_e$ is plotted in Fig. 9 for several values of R/C .

For example, let the relative rate be $R/C = 0.8$. Suppose a probability of error $P_e = 10^{-7}$ is required. The asymptotic expression for the probability of error for orthogonal codes indicates a blocklength of approximately $L = 1625$ binary digits (see Fig. 9). Figure 8 shows that the WB coding scheme requires a blocklength of only $L = 12$ binary digits. For relative rates closer to unity an even more marked difference is obtained.

If $C = 1$ bit/second, these blocklengths correspond to a coding delay T :

$$T = \begin{cases} \frac{1625}{0.8} = 2031 \text{ seconds (orthogonal codes)} \\ \frac{12}{0.8} = 15 \text{ seconds (with our feedback scheme).} \end{cases}$$

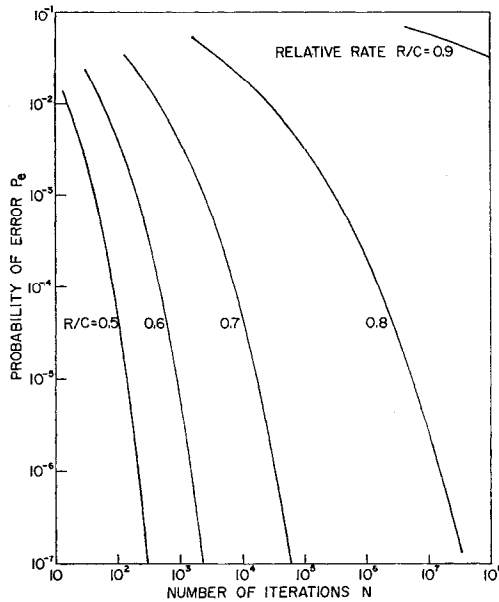


Fig. 6. The probability of error as a function of the number of iterations.

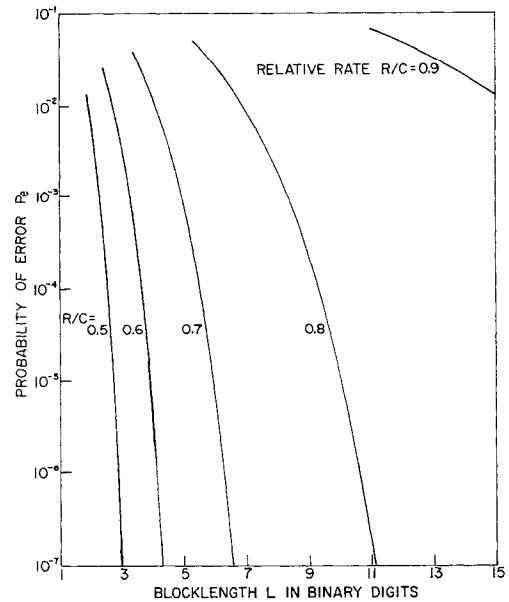


Fig. 8. The probability of error as a function of the blocklength in binary digits.

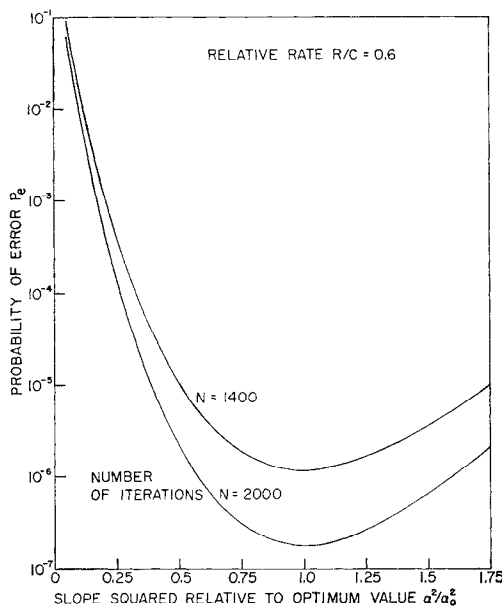


Fig. 7. The probability of error vs. the slope squared relative to its optimum value.

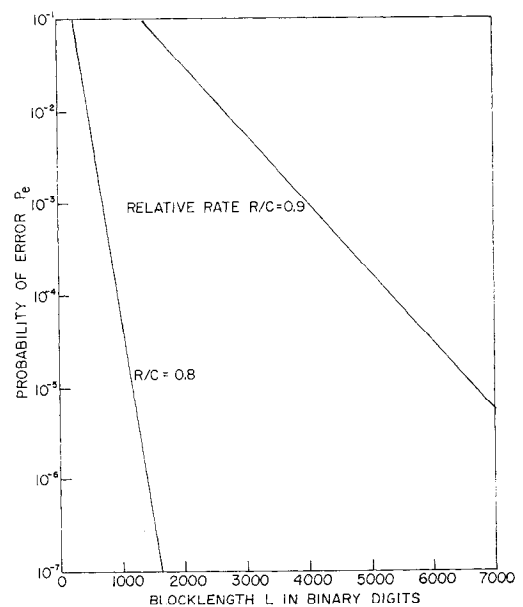


Fig. 9. The asymptotic expression for the probability of error for orthogonal codes as a function of the blocklength in binary digits.

Therefore, the use of feedback provides a considerable reduction in coding delay (and hence, coding and decoding complexity).³ The savings due to feedback become even more pronounced as we go to lower values of P_e and to rates closer to the channel capacity. However, we should point out two ways in which the comparison above is somewhat unfair. The first, and less important, is that we have obtained the value $T = 15$ seconds for the feedback channel by using the exact formula for $P_{e, fb}$, whereas the value $T = 2031$ seconds was obtained from the asymptotic formula for $P_{e, orth}$. However, for the error probability (10^{-7}) we are considering, the exact calculation for T (which is difficult to perform) would not give results much different from $T = 2031$. The second, and more serious objection, is that the value for orthogonal codes is based on a strict power limitation of P_{av} for all code words, whereas in the feedback scheme, it is only the *expected* average power that is limited to P_{av} . (The expectation is over the Gaussian noise variables.) The instantaneous average (over the time interval T) power is the sum of a large number of squared Gaussian variates (11); the expected (or mean) value of the average power is P_{av} and using a well-known relation ($Ex^4 = 3Ex^2$ for a Gaussian random variable x), we see that the variance is $3P_{av}$. If we make allowance for this variation by using, say, $P_{av} + 3\sqrt{3P_{av}}$ for the power with the feedback scheme, we will need a larger coding delay T and the reduction will not be in the ratio 2000 to 15. However, without making any recalculations, we feel it is fair to say that the use of feedback definitely produces an "order of magnitude" reduction in the necessary coding delay T .

III. FURTHER PROPERTIES AND EXTENSIONS

In this section we shall examine the bandwidth and peak power requirements of our coding scheme, study the effects of loop delay and feedback noise, and consider extensions to channels where the spectral density $N_0/2$ may not be known and/or where the additive white noise is not Gaussian.

A. Bandwidth of the Transmitted Signals

The feedback communication system described in this section has no constraint on the bandwidth of the transmitted signals. It will be shown presently why it is not possible to cope with a bandwidth constraint.

From Section II-D, $N = e^{2AT}$ iterations are made in T seconds. Suppose the transmitted signals have bandwidth W , then the number of iterations is at most equal to the number of degrees of freedom. The number of degrees of freedom of a waveform of bandwidth W and duration T is approximately equal to $2WT$. Putting $N = 2WT$,

³ As we mentioned in the Introduction, with $C = 1$, Turin's scheme [15] would require only an *average* coding delay of 1 second and the *average* signaling rate will be 1 bit/second. However, the actual coding delays may fluctuate considerably from the average value.

$$W = \frac{1}{2T} e^{2AT}$$

where

$$A = C \left(\frac{\alpha^2}{6N_0} + \sum_{i=1}^{N-1} \frac{1}{i} \right)^{-1} \ln N \quad (19)$$

which follows from substituting $\sigma^2 = N_0/2$ into (12). From (19) we see that A is asymptotically equal to C for large N , and hence, $W \approx 1/(2T)e^{2CT}$. That is, W grows exponentially with T and $\lim_{T \rightarrow \infty} W(T) = \infty$.

Substituting $T = 1/(2A) \ln N$ into (19) leads to an expression for W in terms of the number of iterations:

$$W = A \frac{N}{\ln N} \text{ c/s.} \quad (20)$$

B. Peak Power

It is known a priori that θ must lie in the interval $[0, 1]$. Restricting the Robbins-Monro procedure to this interval will limit the peak power for fixed bandwidth W . This can be done with the aid of the following theorem from Venter [14].

Theorem: Suppose D is a closed convex subset of R^p , p -dimensional Euclidean space, and it is known a priori that $\theta \in D$. Then modify the stochastic approximation procedure in the following way:

$$X_{n+1} = \begin{cases} X_n + a_n Y_n(X_n) & \text{if } X_n + a_n Y_n(X_n) \in D \\ \left[\begin{array}{l} \text{the point on the} \\ \text{boundary of } D \text{ clos-} \\ \text{est to } X_n + a_n Y_n(X_n) \end{array} \right] & \text{if } X_n + a_n Y_n(X_n) \notin D. \end{cases}$$

Whenever the original procedure converges, so does its restriction to D . The asymptotic rate of convergence for both procedures is the same.

A special case of this theorem, in which the closed convex subset is equal to the unit interval $[0, 1]$ and $p = 1$, is applicable to our coding scheme. Hence, the modified procedure is as follows.

$$X_{n+1} = \begin{cases} 0 & \text{if } X_n + a_n Y_n(X_n) \leq 0 \\ X_n + a_n Y_n(X_n) & \text{if } 0 < X_n + a_n Y_n(X_n) < 1. \\ 1 & \text{if } 1 \leq X_n + a_n Y_n(X_n) \end{cases}$$

In investigating how the peak power P_{peak} depends on the bandwidth W , let us consider a basic signal $\phi(t)$,

$$\phi(t) = \sqrt{2W} \frac{\sin 2\pi W t}{2\pi W t}.$$

This signal has bandwidth W and satisfies the orthogonality condition (5) for $\Delta = 1/2W$. With $N = e^{2AT}$ [A given by (19)],

$$\Delta = \frac{T}{N} = T e^{-2AT}$$

$$P_{\text{peak}} = 2W = \frac{1}{\Delta} = \frac{1}{T} e^{2AT}.$$

Hence, for large T (or N), the P_{peak} goes to infinity while the average power remains finite. A similar phenomenon will be discovered if the basic signal $\phi(t)$ is chosen to have a duration Δ . Of course, this exponentially growing (with T) peak power also occurs with one-way channels if orthogonal signals of finite duration or of the form $\sin 2\pi Wt/\sqrt{2W\pi t}$ are used. Since we are using matched filter reception, we can use pulse compression techniques to alleviate the peak power problem. However, this topic is somewhat apart from the main theme of this paper, and we shall therefore not pursue it any further.

C. Loop Delay

Up to this point, only instantaneous feedback has been considered. In a practical situation there will be feedback delay.

Let $F(x) = \alpha(x - \theta)$, and let the additive random variables Z_n be identically distributed. From the iterative relation,

$$X_{n+1} = X_n - \frac{1}{\alpha n} Y_n(X_n)$$

where $Y_n(X_n) = F(X_n) + Z_n$, it can easily be shown that

$$X_{n+1} - \theta = -\frac{1}{\alpha n} \sum_{i=1}^n Z_i. \quad (21)$$

This means that (when the Z_n are Gaussian) X_{n+1} is the maximum likelihood estimate of θ , based on the observations $Y_1(X_1)$ through $Y_n(X_n)$.

Now suppose there are d units of loop delay, so that $Y_n(X_n)$ can first be used to determine X_{n+d+1} . The first time one can use received information is when computing X_{d+2} .

Let us choose as X_{n+d+1} the maximum likelihood estimate of θ , based on observations $Y_1(X_1)$ through $Y_n(X_n)$. The iterative relation now becomes

$$X_{n+d+1} = \frac{(n-1)X_{n+d} + X_n}{n} - \frac{1}{\alpha n} Y_n(X_n). \quad (22)$$

It follows easily that

$$X_{n+d+1} - \theta = -\frac{1}{\alpha n} \sum_{i=1}^n Z_i. \quad (23)$$

One must complete d more transmissions in order to obtain the same variance as in the case of instantaneous feedback, and thus, the influence of the delay will become negligible for large values of n .

D. Non-Gaussian Noise

If the additive white noise is Gaussian, our coding scheme will permit error-free transmission at any rate less than channel capacity. For the scheme to work it is not necessary to know the noise power spectral density $N_0/2$. However, as shown in Section II, knowledge of $N_0/2$ permits one to choose the slope α in an optimum fashion in the nonasymptotic case.

Stochastic approximation, in general, and the Robbins-Monro procedure, in particular, are nonparametric. There-

fore, the coding scheme will also work in the case of non-Gaussian white noise.

What about the probability of error? Sacks' theorem [6] on the asymptotic distribution of X_{n+1} implies that X_{n+1} is asymptotically Gaussian with the required variance. Hence, all the calculations given earlier in this Section are still valid for large N .

Finally, does one achieve channel capacity when the additive noise is non-Gaussian? The critical rate of our system is still $R_{\text{crit}} = P_{\text{av}}/N_0$, and this gives a lower bound on the channel capacity for all non-Gaussian white noise channels with noise of spectral density $N_0/2$.

E. Influence on Feedback Noise on Wideband Coding Scheme

In the case of noiseless feedback it is immaterial whether X_{n+1} or $Y_n(X_n)$ is sent back to the transmitter. This is not true in the case of noisy feedback. The following notation is adopted for this case: a single prime refers to the forward direction and a double prime to the feedback link. Thus, $N'_0/2$ is the (two-sided) power spectral density of the additive white Gaussian noise in the forward channel, and we shall write $N''_0/2$ for the spectral density of the white Gaussian noise $n''(t)$ in the feedback link. The noises in the forward and feedback links are assumed to be independent.

The estimates of θ obtained by the receiver and transmitter are denoted by X'_n and X''_n , respectively. $Y'_n(X''_n)$ is the noisy observation made by the receiver. This value is sent back to the transmitter which obtains $Y''_n(X''_n) = Y'_n(X''_n) + Z''_n$, where Z''_n is the additive noise in the feedback link.

The influence of feedback noise is mainly a reduction in relative rate R/C in the case where the receiver's estimate X_{n+1} is sent back to the transmitter. The probability of error increases only slightly. When the receiver's observation $Y'_n(X''_n)$ is sent back to the transmitter, the feedback noise reduces the rate only slightly and its main effect is an increase in the error probability.

Consider first the case where X_{n+1} is sent back. Equation (12) for the average power changes in that an additional term $\alpha^2(N''_0/2)(N/T)$ due to the feedback noise appears, and also σ^2 changes to $\sigma'^2 = \frac{1}{2}(N'_0 + \alpha^2 N''_0)$ instead of $\sigma^2 = N'_0/2$. If it is assumed that the feedback noise is small compared to the additive disturbance in the forward channel, then σ'^2 will only change slightly. The error probability in (8) will also only change slightly provided that all other quantities in (8) remain the same.

Figure 10 is a plot of the relative rate

$$\frac{R}{C} = (1 - \epsilon) \left\{ \frac{1}{\ln N} \left[\frac{\alpha_0^2}{6N'_0} + \sum_{k=1}^{N-1} \frac{1}{k} \right] + \alpha_0^2 \frac{N''_0}{N'_0} \left(N + \sum_{k=1}^{N-1} \frac{1}{k} \right) \right\}^{-1} \quad (24)$$

vs. the number N of iterations for different values of N''_0 . The upper curve is for noiseless feedback. The probability of error for noiseless feedback is $P'_e = 10^{-4}$. In the case of noisy feedback it is only slightly higher.

Equation (24) follows from (12), adding the additional term $\alpha^2(N''_0/2)(N/T)$. For α_0^2 the optimum value for noiseless feedback is used, that is, the value given by (15).

It is seen from Fig. 10 that for noiseless feedback the relative rate approaches unity with increasing N ; however, in the case of noisy feedback, the curve for noiseless feedback is followed for some time after which the relative rate drops to zero quite suddenly. (Note that no optimization in the presence of feedback noise is attempted. The particular system we use is optimum for $N''_0 = 0$.)

The feedback power P_{fb} is

$$P_{fb} = \frac{1}{\alpha_0^2 T} \frac{1}{2} (N'_0 + \alpha_0^2 N''_0) \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{12} \frac{N}{T} \quad (25)$$

and is again hardly affected by the feedback noise.

Now consider the case where $Y'_n(X''_n)$ is sent back. The average transmitted power as given by (12) is only slightly affected in that now $\sigma^2 = \frac{1}{2}(N'_0 + N''_0)$ instead of $\sigma^2 = N'_0/2$, and the same is true for the relative rate, assuming N''_0 small compared to N'_0 .

What is the influence of the feedback noise on the error probability? X''_{n+1} as used by the transmitter is equal to

$$X''_{n+1} = X''_n - \frac{1}{\alpha n} Y'_n(X''_n)$$

where $Y'_n(X''_n) = Y'_n(X''_n) + Z'_n$ in which $Y'_n(X''_n) = F(X''_n) + Z'_n$ is the noisy observation made by the receiver. A simple derivation shows that

$$X'_{n+1} = X''_{n+1} + \frac{1}{\alpha} \sum_{i=1}^n \frac{1}{i} Z'_i \quad (26)$$

where X'_{n+1} is the estimate of the message point θ computed by the receiver. Hence,

$$X'_{n+1} \sim N\left[\theta, \frac{\sigma^2}{\alpha^2 n} + \frac{\sigma'^2}{\alpha^2} \sum_{i=1}^n \left(\frac{1}{i}\right)^2\right]$$

and the variance, say σ_i^2 , of X'_{n+1} is equal to

$$\sigma_i^2 = \frac{\sigma^2}{\alpha^2 N} + \frac{\sigma'^2}{\alpha^2} \sum_{i=1}^N \left(\frac{1}{i}\right)^2. \quad (27)$$

The formula for the probability of error is

$$P_e = 2 \operatorname{erfc} \left[\frac{\frac{1}{2} M(N)^{-1}}{\sigma_i} \right] = 2 \operatorname{erfc} \left[\frac{N^{-1/2(1-\epsilon)}}{2\sigma_i} \right]. \quad (28)$$

Again, as in the noiseless feedback case, let us find the optimum value of α_0^2 of α^2 . (Note that in the earlier case, where the receiver's estimate X_{n+1} is sent back, such an optimization was not attempted for nonzero feedback noise.) As before,

$$\epsilon = 1 - \frac{R}{C} (\ln N)^{-1} \left(\frac{\alpha^2}{6N_0} \sum_{k=1}^{N-1} \frac{1}{k} \right) \quad (14)$$

where now $N_0 = N'_0 + N''_0$. It is desired to minimize the probability of error with respect to α^2 . From (28), this is equivalent to minimizing $\sigma_i^2 N^{-\epsilon}$. Setting the derivative equal to zero,

$$\frac{d}{d\alpha^2} (\sigma_i^2 N^{-\epsilon}) = -\frac{1}{\alpha^2} \sigma_i^2 N^{-\epsilon}$$

$$+ \sigma_i^2 N^{-\epsilon} (\ln N) \left[\frac{R}{C} (\ln N)^{-1} \frac{1}{6N_0} \right] = 0$$

yields

$$\alpha_0^2 = 6N_0 \left(\frac{R}{C} \right)^{-1} \quad (29)$$

which has the same form as (15) for noiseless feedback.

Figure 11 shows curves of the probability of error P_e vs. the number of iterations N , with the parameter being the power spectral density N''_0/N'_0 of the feedback noise relative to the corresponding quantity for the forward

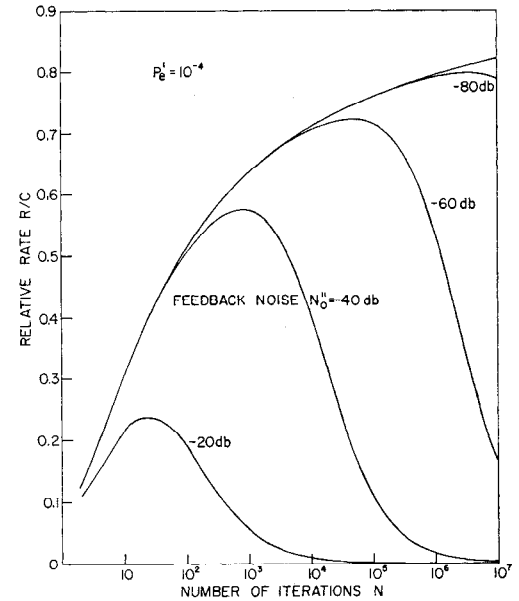


Fig. 10. The relative rate vs. the number of iterations for the case where X_n is sent back.

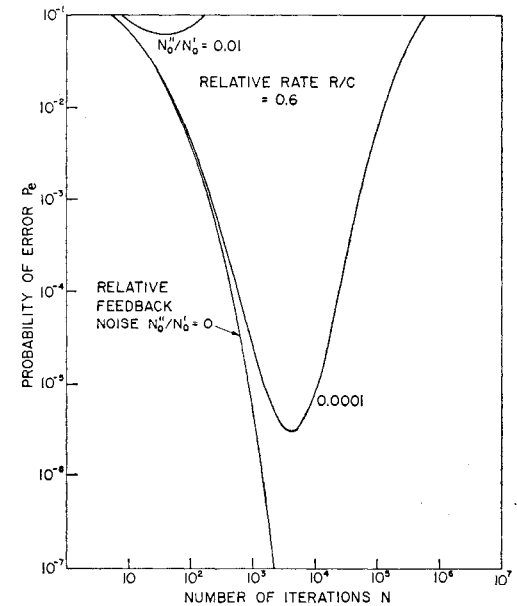


Fig. 11. Probability of error vs. number of iterations.

link. The P_e curves have a minimum for nonzero variance of the feedback noise, and it does not make sense to do more iterations per message than the value indicated by the minimum of the P_e curve.

The average feedback power P_{fb} is

$$P_{fb} = P_{av} - \frac{\alpha_0^2}{12T} + \frac{N_0' N}{2T}. \quad (30)$$

In conclusion, it should be observed that one can either 1) insist on a vanishing probability of error in which case the rate of signaling will approach zero, or 2) require a nonvanishing rate in which case there is a minimum achievable probability of error different from zero.

We should also point out that by using a different scheme, it may be possible to obtain much better results for noisy feedback channels than are yielded by our scheme.

IV. CONCLUDING REMARKS

There are several areas for further work that may be investigated. We shall briefly mention some of them. One is the question of whether our method of exploiting the feedback with a linear encoding function is the most efficient. Even if it should turn out to be the most efficient for Gaussian noise, we can ask whether for non-Gaussian additive white noise, we cannot achieve a rate greater than P_{av}/N_0 by using some other form of encoding function. For example, T. Cover of Stanford University has suggested subjecting the straight line to the nonlinear transformation that would convert a non-Gaussian probability density function into a Gaussian density function; if the stochastic approximation technique is applicable to the resulting curve, this might yield better results for the non-Gaussian case. (It should be pointed out here that asymptotically the shape of $F(x)$ does not matter, since Sacks' theorem on the asymptotic distribution of the estimate is true under very mild assumptions, given in Section II on $F(x)$. The shape of the regression function is important only in nonasymptotic calculations.)

Another possibility is to use sequential detection in combination with our scheme. Instead of making a decision after a prespecified number of iterations, we could wait until the matched filter output crossed a suitable threshold. Such operation would certainly result in some improvement in the error probability, but we suspect the gain may not be worth the extra complexity.

Our results should also be extendable to other situations where stochastic approximation techniques apply, e.g., channels with unknown gains, slowly changing random delays, etc. It may be mentioned that the coding scheme suggested by the Kiefer-Wolfowitz stochastic approximation technique for determining the minimum of an unknown function does not yield results as good as those obtained for the coding scheme in this paper (which was suggested by the Robbins-Monro procedure). J. Venter (Stanford University, 1965) showed in unpublished work

that a coding scheme based on the Kiefer-Wolfowitz procedure cannot achieve channel capacity.

Of course, a major question is the study of communication over noisy feedback links. When more general results, e.g., on the capacity, of such channels are known, it may be easier to look for efficient communication schemes with noisy feedback.

Finally, we mention an extension [16] of the results in this paper to the case of signals with a bandwidth constraint; the scheme to be used under this constraint is more complicated than the one given in this paper, but, of course, it also applies to the non-band-limited-signal case and, in fact, has some advantages—fixed peak power, fewer iterations, etc.

Note added in proof: J. Omura has pointed out a mistake in the coefficient in (17) that arises from an error in arguments based on (10)–(12). The correct formula is

$$P_e \sim \exp [-be^{2(C-R)T}]/[4\pi be^{2(C-R)T}]^{1/2}$$

where $2b = 3e^{-(1+\gamma)}$ and $\gamma = 0.577 \dots$ is Euler's constant. This causes small changes in the curves, but does not affect the main lines of the argument. The details will appear in Omura's thesis at Stanford University.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–424 and 623–657, July–October 1948.
- [2] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, vol. 22, pp. 400–407, September 1951.
- [3] A. Dvoretzky, "On stochastic approximation," *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, pp. 39–55, 1956.
- [4] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. on Information Theory*, vol. IT-2, pp. 8–19, September 1956.
- [5] J. Wolfowitz, "On stochastic approximation methods," *Ann. Math. Stat.*, vol. 27, pp. 1151–1156, December 1956.
- [6] J. Sacks, "Asymptotic distributions of stochastic approximation procedures," *Ann. Math. Stat.*, vol. 29, pp. 373–405, June 1958.
- [7] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Sys. Tech. J.*, vol. 38, pp. 611–656, May 1959.
- [8] P. Elias, "Channel capacity without coding," in *Lectures on Communication System Theory*, Baghdady, Ed. New York: McGraw-Hill, 1961.
- [9] R. M. Fano, *Transmission of Information*. New York: Wiley, 1961.
- [10] P. E. Green, "Feedback Communication Systems," in *Lectures on Communication System Theory*, Baghdady, Ed. New York: McGraw-Hill, 1961.
- [11] J. Wolfowitz, *Coding Theorems of Information Theory*. Berlin: Springer-Verlag, 1961.
- [12] L. H. Zetterberg, "Data transmission over a noisy Gaussian channel," *Trans. Roy. Inst. of Tech.*, no. 184, Stockholm, Sweden, 1961.
- [13] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. on Information Theory*, vol. IT-9, pp. 136–143, July 1963.
- [14] J. Venter, "On stochastic approximation methods," Ph.D. dissertation, University of Chicago, Ill., 1963.
- [15] G. L. Turin, "Signal design for sequential detection systems with feedback," *IEEE Trans. on Information Theory*, vol. IT-11, pp. 401–408, July 1965.
- [16] P. Schalkwijk, "Coding for additive noise channels with feedback—part II: band-limited channels," this issue, page 183.
- [17] A. J. Viterbi, "The effect of sequential decision feedback on communication over the Gaussian Channel," *Information and Control*, vol. 8, pp. 80–92, February 1965.