SU-SEL-65-073

Coding Schemes for Additive Noise Channels With Feedback

by

J. P. M. Schalkwijk

FACILITY FORM 602	N66-13559 (ACCESSION NUMBER)	(THRU)
	58 (PAGES)	1 (CODE)
	CR 68651 (NASA CR OR TMX OR AD NUMBER)	07 (CATEGORY)

GPO PRICE $ ____

CFSTI PRICE(S) $ ____

Hard copy (HC) 3.00

Microfiche (MF) .50

ff 653 July 65

August 1965

Scientific Report No. 10

Prepared under
National Aeronautics and Space Administration
Research Grant No. NsG-377

RADIOSCIENCE LABORATORY

STANFORD ELECTRONICS LABORATORIES

STANFORD UNIVERSITY · STANFORD, CALIFORNIA

SEL-65-073

CODING SCHEMES FOR ADDITIVE NOISE CHANNELS WITH FEEDBACK

by

J. P. M. Schalkwijk

August 1965

Scientific Report No. 10

Prepared under

National Aeronautics and Space Administration
Research Grant No. NsG-377

Radioscience Laboratory

Stanford Electronics Laboratories

Stanford University Stanford, California

ABSTRACT

13559

For one-way transmission that is not subject to a bandwidth constraint, orthogonal codes are known to achieve the capacity of the additive white gaussian noise channel. For a transmitted signal that is subject to a bandwidth constraint, no deterministic way of constructing a code achieving channel capacity has been known. The first such code--for the noiseless feedback case--is developed in this study. It is known that a noiseless feedback channel does not improve the capacity of the discrete memoryless forward channel; however the feedback channel may reduce the coding effort.

Two new coding schemes using feedback are developed and the influence of feedback noise on each is investigated. The WB coding scheme achieves capacity for the wideband ($W \rightarrow \infty$) case and the BL coding scheme achieves capacity for the bandlimited case. The latter scheme could be very important for satellite communications since it allows for a substantial decrease of the coding effort while permitting the satellite to transmit its information at a rate arbitrarily close to channel capacity.

Author

CONTENTS

ILLUSTRATIONS

SYMBOLS

a	a constant in the iterative expression of the Robbins-Munro procedure
a_n	sequence of coefficients
C	channel capacity
d	number of the time units of loop delay
$E(\cdot)$	expected value
$F(\cdot)$	regression curve
g	gain factor
$h(t)$	impulse response of the receiving filter
K	a constant
M	number of messages
N	number of iterations per message
$N(m,\sigma^2)$	normal distribution with mean m and variance σ^2
N_o	noise power spectral density
$n(t)$	additive white (gaussian) noise
$O(\cdot)$	of the order of
P_{av}	average power
P_e	probability of error
P_{peak}	peak power
R	rate of signaling
S/N	signal-to-noise ratio
s	time
$s(t)$	transmitted signal
T	coding delay
W	signal bandwidth

X	abscissa of the regression curve; a random variable
x	abscissa of the regression curve; deterministic
$Y(x)$	noisy observation of the value of the regression curve at x
$y(t)$	received waveform
Z	additive (gaussian) random variable
α	slope of the regression curve at the root
Δ	time unit
δ, ρ	a small positive number
$\delta(\cdot)$	impulse function
ϵ	a small positive number; an element of
θ	root of the regression curve, message point
σ	standard deviation
$\approx$	asymptotic equality
$\sim$	approximate equality; also, distributed like, as in the following example: $X \sim N(m,\sigma^2)$

ACKNOWLEDGMENT

The author wishes to thank Dr. A. M. Peterson for his encouragement and for the discussions on the presentation and practical applicability of the coding schemes developed. Thanks are also due Dr. T. Kailath for pointing out certain problems related to channel capacity and coding.

The interesting comments on stochastic approximation offered by Dr. J. H. Venter of Potchefstroom University, South Africa, during his stay at Stanford University, are gratefully acknowledged.

I. INTRODUCTION

A. BACKGROUND

A general discussion of feedback communication systems was given in 1961 by Green [Ref. 1] who distinguished 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 [Ref. 2] discusses a postdecision feedback system for the white gaussian noise channel. A decision is made when the a posteriori probability computed by the receiver exceeds a certain threshold determined by the probability of error. The transmitter is informed by means of postdecision feedback that the receiver has made its decision, and it then starts sending the next message.

As examples of predecision feedback systems, two interesting sequential schemes are now introduced.

Horstein [Ref. 3] discusses a coding scheme for the binary symmetric channel in which the transmitter sends a sequence of signals so as to maximize the a posteriori probability of the particular message being transmitted. When the a posteriori probability of some message, as computed by the receiver, exceeds a certain threshold, determined by the probability of error, the receiver makes its decision. The transmitter is continually informed about the probabilities computed by the receiver and changes its transmission accordingly.

Turin [Ref. 4] has a scheme applying to the white gaussian noise channel. It is similar to Horstein's scheme in that the receiver computes a likelihood ratio and makes its decision as this likelihood ratio exceeds a threshold set by the probability of error. The value of the likelihood ratio is fed back to the transmitter continually during the decision-making process. 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, such as to make this ratio increase as fast as possible. Average and peak power constraints are invoked. The average time $\overline{T}$ for deciding on a binary digit turns out to be $\overline{T} = \ln 2(P_{av}/N_o)^{-1}$, where P_{av} is the average power and N_o is the noise power spectral density. The probability 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

$$C = \frac{P_{av}}{N_o} \quad \text{nats/sec}^{*} \tag{1.1a}$$

The two coding schemes developed in Chapters II and III are predecision feedback systems: the first is applicable to the white gaussian noise channel and is designated the WB coding scheme; the second is applicable to the bandlimited white gaussian noise channel and is designated the BL coding scheme. Unlike those of Horstein and Turin, these schemes are not sequential in that the time T allocated to the transmission of a particular message is fixed beforehand.

The BL coding scheme discussed in Chapter III gives the first deterministic procedure to achieve the capacity C for the band-limited white gaussian noise channel:

$$C = W \ln\left(1 + \frac{P_{av}}{N_o W}\right) \quad \text{nats/sec} \tag{1.1b}$$

Elias [Ref. 5] has discussed a predecision feedback scheme applicable to this same channel. He divided the channel into K subchannels of bandwidth $w = W/K$. If noiseless feedback is available and if $K \to \infty$, information can be sent at a rate equal to that of Eq. (1.1b). However, since the signal bandwidth is w instead of W, the coding and decoding complexity for the feedback scheme becomes an arbitrarily small fraction of that required without feedback.

*"Nats" is defined as natural units of information in accordance with IEEE standards.

B. NEW RESULTS

Consider the situation represented by Fig. 1. The transmitter sends one of M possible signals of duration T. In the channel the signal is disturbed by additive noise $n(t)$ and therefore the receiver has to guess which of the M messages was actually sent. The word "channel" stands for physical perturbations in the transmission medium and in the receiver front end, as well as for transmitter constraints. Examples of transmitter constraints are an average power constraint, a peak power constraint, a constraint on the signal bandwidth, etc.

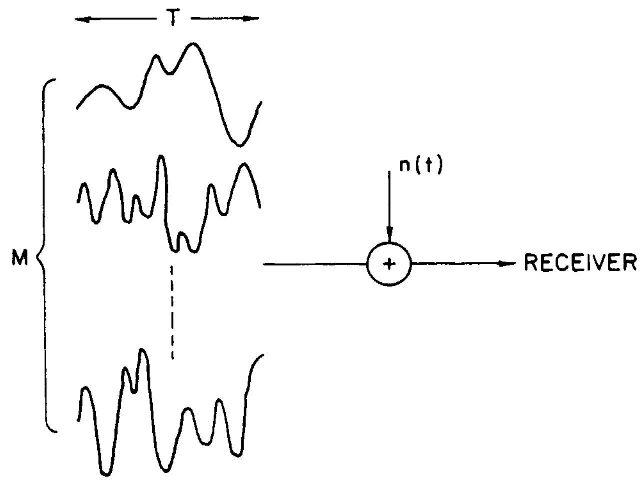


FIG. 1. TYPICAL COMMUNICATION PROBLEM.

The signaling rate is defined as $R = (\ln M)/T$, nats/sec. Consider a particular system. The probability of error will be a function of the number of possible transmitted signals M and of their duration T. It shall be assumed that the M signals are chosen so as to minimize the error probability, which is denoted by $P_{e,opt}(M,T)$. For fixed T the probability of error will increase with increasing M (greater probability of confusion), and for fixed M the probability of error decreases as a function of T (one has a longer look at the noise and can thus do a better job in predicting its effect).

Therefore, if $M = e^{RT}$ does not increase too fast with T (that is, if the rate R is low enough), $P_{e,opt}(M,T)$ can be made arbitrarily small by increasing T (and also M, since $M = e^{RT}$).

Shannon [Ref. 6] shows that there exists a critical rate C such that for any rate $R < C$ there exists a coding scheme for which the probability of error can be made arbitrarily small by making T large enough. This is not the case for any rate higher than C. That is,

$$\lim_{\substack{T \to \infty \\ M=e^{RT}}} P_{e,opt}(M,T) = \begin{cases} 0 & \text{for } R < C \\ 1 & \text{for } R > C \end{cases}$$

This critical rate C is called the channel capacity. For a more detailed consideration of channel capacity see the book by Wolfowitz [Ref. 7].

In the present work two channels are studied:

1. In one channel the disturbance is additive white gaussian noise with two-sided spectral density $N_o/2$. The only transmitter constraint is average power P_{av}. In this case the capacity is

$$C = \frac{P_{av}}{N_o} \qquad \text{nats/sec} \tag{1.1a}$$

It is known that orthogonal codes can be used to achieve this channel capacity. These codes use a bandwidth that grows exponentially with T.

2. In the second channel the disturbance is again additive white gaussian noise with spectral density $N_o/2$. The transmitter constraints are average power P_{av} and a constraint to $(-W,W)$ on the signal bandwidth. In this case the capacity is

$$C = W \ln \left(1 + \frac{P_{av}}{N_o W}\right) \qquad \text{nats/sec} \tag{1.1b}$$

No deterministic coding scheme is known that achieves this capacity.

When a noiseless feedback link is associated with these channels, one might expect to have a higher channel capacity because of additional flexibility. However, Shannon [Ref. 8] has shown that no increase in channel capacity can be obtained for memoryless channels by use of the noiseless feedback link. Still, some advantage should accrue from the presence of such a link, and in fact the coding procedures required to achieve a given probability of error (for any rate up to channel capacity) are much less complicated than those needed to achieve the same performance without a noiseless feedback link.

In Chapter II a coding scheme is presented for signaling over additive white gaussian noise channels that have no bandwidth limitation (channel 1) on the transmitted signals. For this coding scheme the error probability for large T is approximately

$$P_e \approx \frac{\exp\left[-\frac{3C}{2R}\, e^{2(C-R)T}\right]}{\left[6\pi \frac{C}{R}\, e^{2(C-R)T}\right]^{\frac{1}{2}}} \tag{1.2a}$$

This equation can be compared with the asymptotic error probability for the best codes (orthogonal codes) without a feedback link. In this case [see, e.g., Fano, Ref. 9, Chapter VI],

$$P_e \approx \frac{\text{Const}}{T^{\beta}}\, e^{-TE(R)} \tag{1.2b}$$

where

$$E(R) = \begin{cases} \frac{C}{2} - R & \text{for } 0 \leq R \leq \frac{C}{4} \\ (\sqrt{C} - \sqrt{R})^2 & \text{for } \frac{C}{4} \leq R \leq C \end{cases} \quad , \; 1 \leq \beta \leq 2$$

A comparison of these two expressions indicates that although the channel capacity is not increased by noiseless feedback, the coding delay T for the scheme developed here is only a fraction of the coding delay for orthogonal codes.

The WB coding scheme discussed in Chapter II has two difficulties associated with it. One is that for rates close to channel capacity the number N of iterations (transmissions) per message becomes very high (e.g., 10^9). Secondly, the peak power approaches infinity for rates close to channel capacity. With the BL coding scheme discussed in Chapter III, however, these difficulties can be avoided.

The latter coding scheme applies for channel 2, i.e., bandlimited signals. It is the first deterministic coding procedure to achieve capacity for this particular channel. The following exact expression for the error probability has been derived:

$$P_e = 2 \operatorname{erfc} \left\{ \sqrt{3} \left[\frac{\frac{N-1}{N} + \frac{P_{av}}{N_0 W}}{\exp\left(\frac{R}{W}\right)} \right]^{N/2} \right\} \tag{1.2c}$$

This expression can be compared with the bounds on the best achievable P_e for one-way communication as plotted by Slepian [Ref. 10]. A considerable improvement due to noiseless feedback is found.

Both Chapters II and III consider the deterioration in performance due to feedback noise. If one wants the probability of error to vanish, one finds that the rate of signaling approaches zero. On the other hand, when the rate of transmission is held constant, the probability of error has a minimum achievable value different from zero.

II. A FEEDBACK COMMUNICATION SYSTEM WITH NO CONSTRAINT ON THE BANDWIDTH

In this chapter a coding scheme for additive white gaussian noise channels with noiseless feedback is developed. There is an average power constraint on the transmitted signals. Let P_{av} be the average transmitted power. The additive (zero mean) white gaussian noise has double-sided spectral density $N_o/2$, and so the covariance function is $(N_o/2)\ \delta(t-s)$. There are no restrictions on the bandwidth or on the peak power of the signals. Such assumptions are usually unrealistic for terrestrial communication channels but seem to be quite appropriate in space communication problems. In the next chapter, however, channels that do have bandwidth and peak-power constraints shall be treated.

In the following sections it will be shown that the additive white gaussian noise channel can be converted to an equivalent time-discrete channel and that the channel capacity $C = P_{av}/N_o$ of the additive white gaussian noise channel can be achieved. The probability of error for finite coding delay T will be calculated and the results compared with those for orthogonal codes for one-way transmission. Some properties of the WB coding scheme, such as bandwidth, peak power, loop delay, and nongaussian statistics, will be discussed. The deterioration in performance of the WB coding scheme in the presence of feedback noise will be considered.

A. CHANNELS WITH ADDITIVE WHITE GAUSSIAN NOISE

In the WB coding scheme the information is transmitted by suitably modulating the amplitude of a known basic signal waveform $\phi(t)$.

Let the signal $s(t)$, see Fig. 2, be of the form

$$s(t) = \Sigma\, X_i\, \phi(t-i\Delta), \quad i = 1, 2, \ldots$$

Assume that $\phi(t)$ satisfies

$$\int \phi(t-i\Delta)\, \phi(t-j\Delta)\, dt = \delta_{ij} \tag{2.1}$$

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

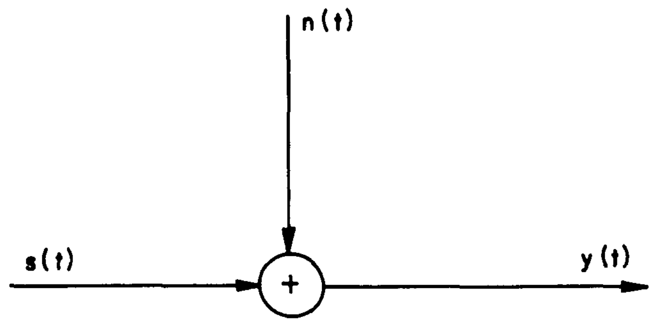


FIG. 2. MODEL FOR THE ADDITIVE NOISE CHANNEL.

Reception is achieved using a filter matched to $\phi(t)$, that is, $h(t) = \phi(-t)$. The output of this matched filter at $t = i\Delta$, $i = 1, 2, \ldots,$ is the sequence $< Y_i(X_i) >$, where $Y_i(X_i) = X_i + Z_i$. The additive random variables

$$Z_i = \int n(t)\, \phi(t-i\Delta)\, dt \tag{2.2}$$

will be shown to be independent for different values of i. Their variance is $\sigma^2 = N_o/2$, where $N_o/2$ is the (two-sided) spectral density of the white gaussian noise $n(t)$. Note that the covariance function of the noise $n(t)$, which is the Fourier transform of $N_o/2$, is

$$E\left[n(t)\, n(s)\right] = \frac{N_o}{2}\, \delta(t-s)$$

Therefore

$$E\left[Z_i Z_j\right] = E \iint n(t)\, n(s)\, \phi(t-i\Delta)\, \phi(s-j\Delta)\, dt\, ds$$

$$= \iint E\left[n(t)\, n(s)\right] \phi(t-i\Delta)\, \phi(s-j\Delta)\, dt\, ds$$

$$= \frac{N_o}{2} \iint \delta(t-s)\, \phi(t-i\Delta)\, \phi(s-j\Delta)\, dt\, ds$$

$$= \frac{N_o}{2} \iint \phi(t-i\Delta)\, \phi(t-j\Delta)\, dt = \frac{N_o}{2}\, \delta_{ij}$$

Next it will be shown that using the type of modulation discussed above, the white gaussian noise channel can be replaced by a time-discrete channel where a sequence of numbers $< X_i >$ is transmitted and a sequence of numbers $< Y_i(X_i) > = < X_i + Z_i >$ is received. The time unit is Δ. It has been shown that the random variables Z_i are independent for different values of i and that their variance is $\sigma^2 = N_o/2$. Since Eq. (2.2) is a linear functional of $n(t)$, Z_i is a gaussian random variable whenever the additive noise $n(t)$ is gaussian. Furthermore, the transmitted energy is equal to ΣX_i^2 in this time-discrete channel model, because of the orthonormality condition (2.1).

The question is, does one lose information in not considering the noise component $\tilde{n}(t) = n(t) - \bar{n}(t)$, where $\bar{n}(t) = \Sigma Z_i\, \phi(t-i\Delta)$.

It is easily shown that $\int \tilde{n}(t)\, \phi(t-i\Delta)\, dt = 0$, $i = 1, 2, \ldots,$ so that $\tilde{n}(t)$ is orthogonal to the signal space.

Next let us find the correlation between $\bar{n}(t)$ and $\tilde{n}(t)$.

$$E\left[\bar{n}(t)\, n(t)\right] = E\left\{\left[\Sigma Z_i\, \phi(t-i\Delta)\right]\left[n(t) - \Sigma Z_i\, \phi(t-i\Delta)\right]\right\}$$

$$= \Sigma \left\{\phi(t-i\Delta)\, E\left[n(t) \int n(s)\, \phi(s-i\Delta)\, ds\right]\right\}$$

$$- \Sigma \left[\phi^2(t-i\Delta)\, E(Z_i^{\,2})\right]$$

$$= \Sigma \left\{\phi(t-i\Delta) \int E\left[n(t)\, n(s)\right] \phi(s-i\Delta)\, ds\right\}$$

$$- \frac{N_o}{2}\, \Sigma\, \phi^2(t-i\Delta)$$

$$= \Sigma \left[\phi(t-i\Delta) \frac{N_o}{2} \int \delta(t-s)\, \phi(s-i\Delta)\, ds \right] - \frac{N_o}{2} \Sigma\, \phi^2(t-i\Delta) = 0$$

It is seen that $\bar{n}(t)$ and $\tilde{n}(t)$ are uncorrelated; and since they are gaussian, $\tilde{n}(t)$ is independent of $\bar{n}(t)$ and hence of Z.

It has been shown that one can replace the additive white gaussian noise channel by a time-discrete channel where a sequence $< X_i >$ is sent, at integral values of the time unit Δ, and where a sequence $< Y_i(X_i) > = < X_i + Z_i >$ is received. The gaussian random variables Z_i have zero mean, variance $\sigma^2 = N_o/2$, and are uncorrelated for different values of i. Furthermore, the transmitted energy is equal to ΣX_i^2.

B. THE WB CODING SCHEME

The coding scheme developed in this section was suggested by the Robbins-Munro [Ref. 11] stochastic approximation technique which is described in Sec. 1. Theorems concerning stochastic approximation shall also be used when dealing with peak-power limitations, loop delay, and nongaussian noise.

1. The Robbins-Munro Procedure

Consider the situation indicated in Fig. 3. 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 Munro proposed the following recursive scheme: Start with an arbitrary initial guess X_1 and make successive guesses according to

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

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

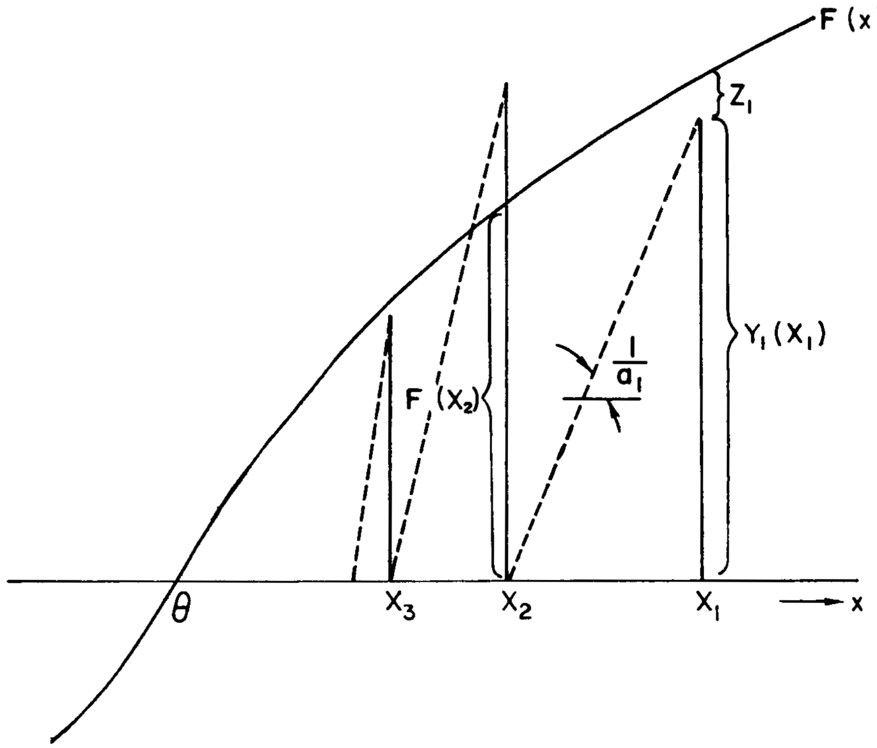


FIG. 3. THE ROBBINS-MUNRO PROCEDURE.

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

1. $F(x) \gtrless 0$ according to $x \gtrless \theta$.
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, $X_n \to \theta$ almost surely; and if $E | X_1 |^2 < \infty$, then $E | X_n - \theta |^2 \to 0$.

Robbins and Munro proved the convergence in mean square. The "convergence almost surely" was first proved by Wolfowitz [Ref. 12] in 1956. The best source for a proof of the preceding theorem is Dvoretsky's paper [Ref. 13], where several types of stochastic-approximation procedures are treated in a unified manner.*

The Robbins-Munro 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 [Ref. 15] that $\sqrt{n}\,(X_{n+1}-\theta)$ is normally distributed for large n. (This result shall be used later.) Let the following assumptions, which complement the earlier requirements, be fulfilled.

5. $\sigma^2(x) \to \sigma^2(\theta)$ as $x \to \theta$
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 $\sup \left\{ E|Z(x)|^{2+\delta};\ |x-\theta| \leq t \right\} < \infty$
8. $2\alpha > a$

Theorem (Sacks). Fulfillment of all the above conditions yields

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

2. Description of the WB Coding Scheme

The transmitter has to send one of M possible messages to a receiver. A noiseless feedback channel is available. The proposed scheme, as shown on Fig. 4, is described below.

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

*A recent general survey of stochastic approximation methods is given in Venter's thesis [Ref. 14].

$\sigma^2 = N_o/2$. 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 $F(X_{n+1}) = \alpha(X_{n+1}-\theta)$.

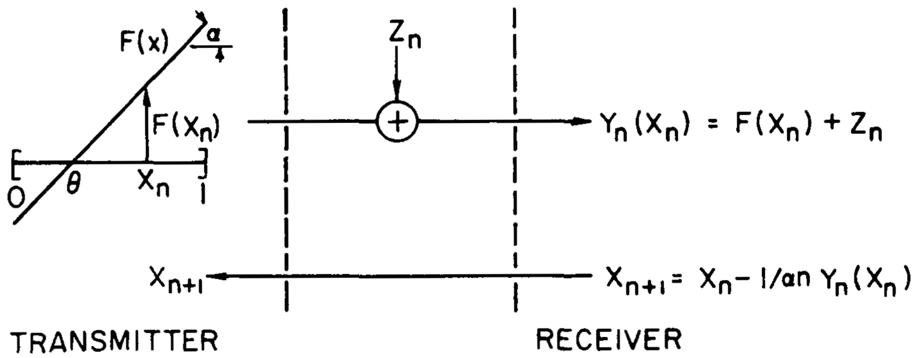


FIG. 4. PROPOSED CODING SCHEME FOR WIDEBAND SIGNALS.

From the theorem 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 $(\sigma/\alpha)^2$.

Straightforward computation shows that in the case where the additive disturbance is gaussian, X_{n+1} will be $N(\theta, \sigma^2/\alpha^2 n)$, that is, normal with mean θ and variance $\sigma^2/\alpha^2 n$.

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. 5. 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

$$P_e = 2 \operatorname{erfc}\left(\frac{\frac{1}{2}M^{-1}}{\sigma/\alpha\sqrt{N}}\right) \qquad (2.3a)$$

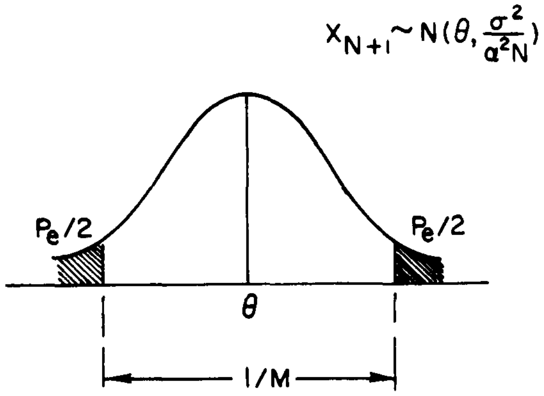


FIG. 5. DERIVATION OF THE ERROR PROBABILITY.

How large can one choose M in order for the probability of error to vanish for increasing N? The distribution in Fig. 5 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. Setting $M(N) = N^{\frac{1}{2}(1-\epsilon)}$ yields the probability of error

$$P_e = 2 \text{ erfc} \left[\frac{\alpha}{2\sigma} N^{\epsilon/2} \right] \tag{2.3b}$$

and as $N \to \infty$,

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

The critical rate determined by $\epsilon = 0$ will be

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

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

$$R_{crit} = \frac{\ln N}{2T} = A \qquad \text{nats/sec}$$

Now 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 Z. Therefore, using $E(\cdot)$ to denote averaging over the noise process gives

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

Now $T = (1/2A) \ln N$, and assuming a uniform prior distribution for the message point θ, $E(X_1 - \theta)^2$ will be $1/12$. Moreover, since $E(X_{i+1} - \theta)^2 = \sigma^2 / \alpha^2 i$, substitution in the formula for the average power leads to

$$P_{av} = \frac{2A}{\ln N} \left(\frac{\alpha^2}{12} + \sigma^2 \sum_{i=1}^{N-1} \frac{1}{i} \right) \tag{2.4}$$

Therefore

$$\lim_{N \to \infty} P_{av}(N) = 2\sigma^2 A = N_o A \qquad \text{or} \qquad A = \frac{P_{av}}{N_o}$$

and thus A is constrained to P_{av}/N_o and the critical rate becomes

$$R_{crit} = A = \frac{P_{av}}{N_o} \qquad \text{nats/sec} \tag{2.5}$$

From the above it is seen that the WB coding scheme presented here achieves channel capacity.

C. OPTIMUM P_e FOR FINITE N

In this section the value of the slope α, given R/C and given N, that minimizes the probability of error shall be determined. From Eq. (2.3), minimizing the probability of error is equivalent to maximizing

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

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

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

To compute $\frac{d\epsilon}{d(\alpha^2/N_o)}$, an expression for ϵ is needed. Substituting $R = (1-\epsilon)R_{crit} = (1-\epsilon)A$, and $\sigma^2 = N_o/2$ into formula (2.4) leads to

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

from which

$$\epsilon = 1 - \frac{R}{C} \left[\frac{\alpha^2}{6N_o} + \sum_{i=1}^{N-1} \frac{1}{i} \right] (\ln N)^{-1} \qquad (2.6)$$

Hence, $d\epsilon/[d(\alpha^2/N_o)] = -(R/6C)(\ln N)^{-1}$, which yields

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

Therefore, the optimum value α_o^2 of α^2 is

$$\alpha_o^2 = 6N_o \left(\frac{R}{C} \right)^{-1} \qquad (2.7)$$

Substituting for α_o^2 in the formula for the probability of error gives

$$P_e = 2 \text{ erfc} \left[\left(3 \frac{C}{R} N^{\epsilon} \right)^{\frac{1}{2}} \right]$$

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 formula (2.6) is positive. Note that for relative rates approaching unity, the number of transmissions per message becomes very high.

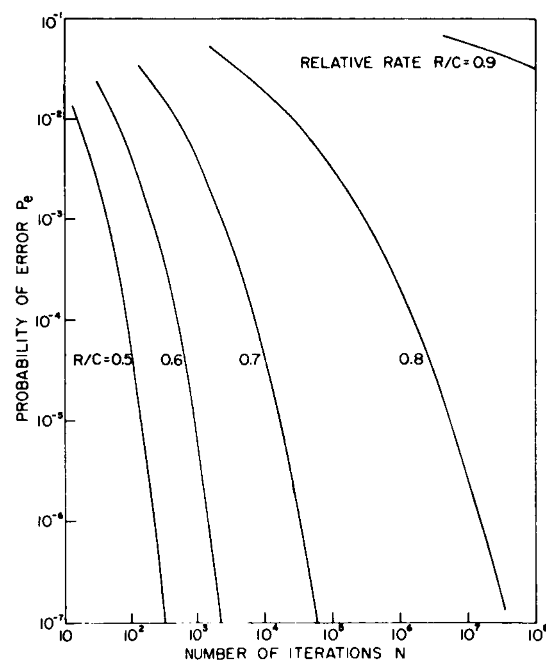


FIG. 6. THE PROBABILITY OF ERROR AS A FUNCTION OF THE NUMBER OF ITERATIONS.

Formula (2.7) gives the optimum value of the slope α as a function of the relative rate R/C and the noise power spectral density $N_o/2$. Figure 7 shows curves of the probability of error vs the slope squared relative to its optimum value.

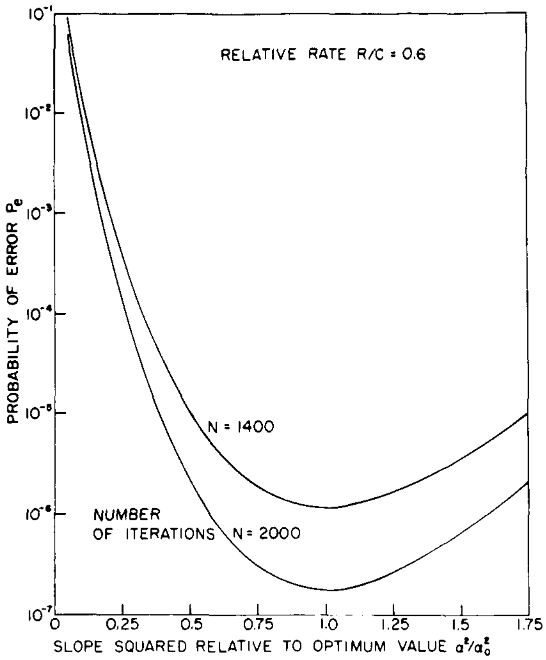


FIG. 7. THE PROBABILITY OF ERROR VERSUS THE SLOPE SQUARED RELATIVE TO ITS OPTIMUM VALUE.

For the WB coding scheme one can also write down an asymptotic expression for the probability of error, similar to expression (1.2b), for orthogonal codes. From Eq. (2.3) and substituting $\sigma^2 = N_o/2$, the probability of error becomes

$$P_e = 2 \text{ erfc} \left[\left(\frac{\alpha^2}{2N_o} N^{\epsilon} \right)^{\frac{1}{2}} \right]$$

By using the optimum α_o^2 of α^2 given by Eq. (2.7), the above equation is asymptotically equal to

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

Furthermore, $N = e^{2AT}$, $R = (1-\epsilon)A$, where A is asymptotically equal to C as was shown in Sec. B2. Thus, P_e is asymptotically equal to

$$P_e \approx \frac{\exp\left[-\frac{3C}{2R} e^{2(C-R)T}\right]}{\left[6\pi \ \frac{C}{R} \ e^{2(C-R)T}\right]^{\frac{1}{2}}} \tag{1.2a}$$

For one-way channels, the best P_e (achieved by orthogonal codes) is given by Eq. (1.2b).

An alternate comparison can be based on the blocklength L, in binary digits, which is defined as follows. Let $2^L = M$. After N iterations, M is $N^{\frac{1}{2}(1-\varepsilon)}$ and hence $L = \frac{1}{2}(1-\varepsilon) \log_2 N$. Figure 8 gives curves of the probability of error vs the blocklength L.

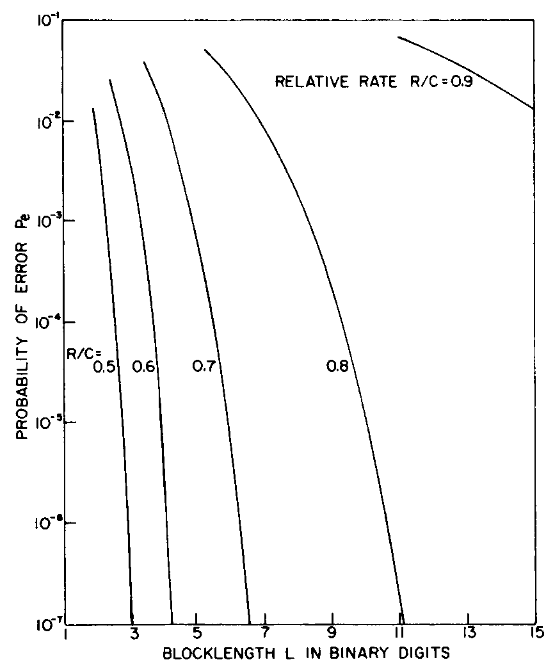


FIG. 8. THE PROBABILITY OF ERROR AS A FUNCTION OF THE BLOCKLENGTH IN BINARY DIGITS.

Plotting $\log_{10} P_e$ vs L for orthogonal codes gives

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

so that

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

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

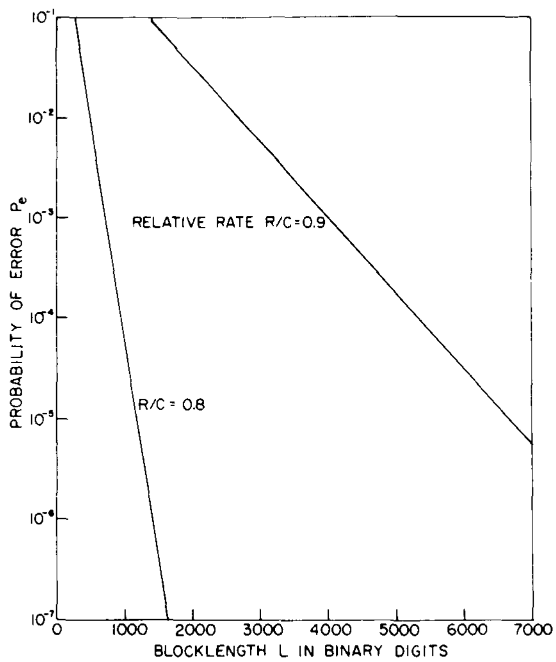


FIG. 9. THE ASYMPTOTIC EXPRESSION FOR THE PROBABILITY OF ERROR FOR ORTHOGONAL CODES AS A FUNCTION OF THE BLOCKLENGTH IN BINARY DIGITS.

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/sec, these blocklengths correspond to a coding delay T as given below:

$$T = \begin{cases} \dfrac{1625}{0.8} = 2031.25 \text{ sec} & \text{(orthogonal codes)} \\ \\ \dfrac{12}{0.8} = 15 \text{ sec} & \text{(WB coding scheme)} \end{cases}$$

D. SOME PROPERTIES OF THE WB CODING SCHEME

1. Bandwidth

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

From Sec. IIB, $N = e^{2AT}$ iterations are made in T sec. 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 $N = 2WT$,

$$W = \frac{1}{2T} e^{2AT} \tag{2.8}$$

where

$$A = C \left(\frac{\alpha^2}{6N_o} + \sum_{i=1}^{N-1} \frac{1}{i} \right)^{-1} \ln N \tag{2.9}$$

which follows from substituting $\sigma^2 = N_o/2$ into Eq. (2.4). From Eq. (2.9), A is asymptotically equal to C for large N and hence $W \approx 1/(2T)\, e^{2CT}$. It is thus seen that W grows exponentially with T and $\lim_{T\to\infty} W(T) = \infty$.

Substituting $T = 1/(2A) \ln N$ into Eq. (2.8) leads to an expression for W in terms of the number of iterations:

$$W = A \frac{N}{\ln N} \text{ cps} \tag{2.10}$$

2. Peak Power

It is known a priori that θ must lie in the interval $[0, 1]$. Restricting the Robbins-Munro procedure [Ref. 11] to this interval will limit the peak power for fixed bandwidth W. This is done with the aid of the following theorem of Venter [Ref. 14].

Theorem (Venter). 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:

$$X_{n+1} = \begin{cases} X_n + a_n Y_n(X_n) & \text{if } X_n + a_n Y_n(X_n) \in D \\ \begin{bmatrix} \text{the point on the} \\ \text{boundary of } D \\ \text{closest to} \\ X_n + a_n Y_n(X_n) \end{bmatrix} & \text{if } X_n + a_n Y_n(X_n) \notin D \end{cases}$$

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 the WB 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, consider the basic signal $\phi(t)$ which has bandwidth W and which satisfies the orthonormality condition (2.1) for $\Delta = 1/(2W)$:

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

With $N = e^{2AT}$,

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

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

Hence for large T (or N), the P_{peak} goes to infinity while the average power remains finite. As in Turin's scheme [Ref. 4], infinite peak power is required in order to achieve zero probability of error. Figure 10 shows the expected instantaneous transmitter power as a function of time for finite value of the coding delay T. In the case where the additive disturbance is gaussian, the variance of the instantaneous transmitted power is three times its expected value.

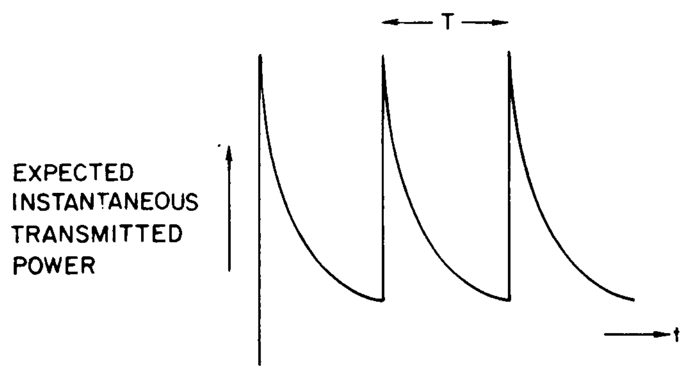


FIG. 10. THE EXPECTED INSTANTANEOUS TRANSMITTED POWER AS A FUNCTION OF TIME.

3. 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 derived that

$$X_{n+1} - \theta = - \frac{1}{\alpha n} \sum_{i=1}^{n} Z_i \qquad (2.11)$$

In other words, 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

$$X_{n+d+1} = \frac{(n-1)\, X_{n+d} + X_n}{n} - \frac{1}{\alpha n}\, Y_n(X_n) \qquad (2.12)$$

It follows easily that

$$X_{n+d+1} - \theta = - \frac{1}{\alpha n} \sum_{i=1}^{n} Z_i \qquad (2.13)$$

One must complete d more iterations 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.

4. <u>Nongaussian Noise</u>

If the additive white noise is gaussian the WB 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 $N_o/2$. However, as shown in Sec. IIC, knowledge of $N_o/2$ permits one to choose the slope α in an optimum fashion in the nonasymptotic case.

Stochastic approximation in general, and the Robbins-Munro procedure in particular, are nonparametric. Therefore the WB coding scheme will also work in the case of nongaussian white noise.

What about the probability of error? Sacks' theorem [Ref. 15] 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 chapter are still valid for large N.

Finally, does one achieve channel capacity when the additive noise is nongaussian? The critical rate of our system is still $R_{crit} = P_{av}/N_o$, and this gives a lower bound on the channel capacity for the nongaussian case.

E. INFLUENCE OF FEEDBACK NOISE ON WB 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 $N_o'/2$ is the (two-sided) power spectral density of the additive white gaussian noise in the forward channel, and $N_o''/2$ is the corresponding quantity for the feedback link.

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 (2.4) for the average power changes in that an additional term $\alpha^2(N_o''/2)(N/T)$ due to the feedback noise appears, and also σ^2 changes slightly, that is, $\sigma^2 = (N_o' + \alpha^2 N_o'')$ instead of $\sigma^2 = N_o'/2$. 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 Eq. (2.3) will also only change slightly provided that all other quantities in Eq. (2.3) remain the same.

Figure 11 is a plot of the relative rate

$$\frac{R}{C} = (1 - \epsilon)\left\{\frac{1}{\ln N}\left[\frac{\alpha_o^2}{6N_o'} + \sum_{k=1}^{N-1}\frac{1}{k} + \alpha_o^2\,\frac{N_o''}{N_o'}\left(N + \sum_{k=1}^{N-1}\frac{1}{k}\right)\right]\right\}^{-1} \qquad (2.14)$$

vs the number N of iterations for different values of N_o''. 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 (2.14) follows from formula (2.4) adding the additional term $\alpha^2(N_o''/2)(N/T)$. For α_o^2 the optimum value for noiseless feedback is used, that is, the value given by Eq. (2.7).

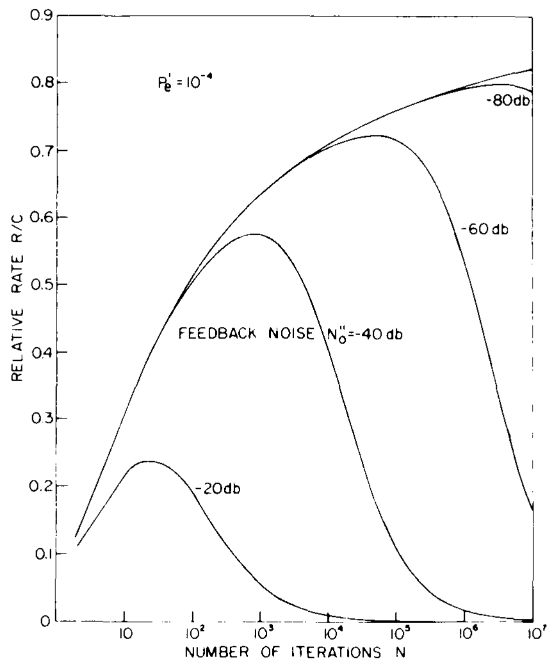


FIG. 11. THE RELATIVE RATE VS THE NUMBER OF ITERATIONS FOR THE CASE WHERE X_n IS SENT BACK.

It is seen from Fig. 11 that for noiseless feedback the relative rate approaches unity with increasing N; while 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 is optimum for $N_o'' = 0$.

The feedback power P_{fb} is

$$P_{fb} = \frac{1}{\alpha_o^2 T} \quad \frac{1}{2} (N_o' + \alpha_o^2 N_o'') \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{12} \frac{N}{T} \tag{2.15}$$

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 Eq. (2.4) is only slightly affected in that now $\sigma^2 = \frac{1}{2}(N_o' + N_o'')$ instead of $\sigma^2 = N_o'/2$ and the same is true for the relative rate, assuming N_o'' small compared to N_o'.

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

$$X_{n+1}' = X_{n+1}'' + \frac{1}{\alpha} \sum_{i=1}^{n} \frac{1}{i} Z_i'' \tag{2.16}$$

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

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

and the variance, say σ_t^2, of X_{N+1}' is equal to

$$\sigma_t^2 = \frac{\sigma^2}{\alpha^2 N} + \frac{\sigma''^2}{\alpha^2} \sum_{i=1}^{N} \left(\frac{1}{i}\right)^2 \tag{2.17}$$

The formula for the probability of error is

$$P_e = 2 \text{ erfc } \left[\frac{\frac{1}{2} M(N)^{-1}}{\sigma_t} \right] = 2 \text{ erfc } \left[\frac{N^{-\frac{1}{2}(1-\epsilon)}}{2\sigma_t} \right] \tag{2.18}$$

Again, as in the noiseless feedback case, let us find the optimum value α_o^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,

$$\epsilon = 1 - \frac{R}{C} (\ln N)^{-1} \left(\frac{\alpha^2}{6N_o} + \sum_{k=1}^{N-1} \frac{1}{k} \right) \tag{2.6}$$

where now $N_o = N_o' + N_o''$. It is desired to minimize the probability of error with respect to σ^2. From Eq. (2.18) this is equivalent to minimizing $\sigma_t^2 N^{-\epsilon}$. Setting the derivative equal to zero,

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

yields

$$\sigma_o^2 = 6N_o \left(\frac{R}{C} \right)^{-1} \tag{2.19}$$

which has the same form as Eq. (2.7) for noiseless feedback.

Figure 12 shows curves of the probability of error P_e vs the number of iterations N, with the parameter being the power spectral density N_o''/N_o' of the feedback noise relative to the corresponding quantity for the forward 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

$$P_{fb} = P_{av} - \frac{\alpha_o^2}{12T} + \frac{N_o'}{2} \frac{N}{T} \tag{2.20}$$

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.

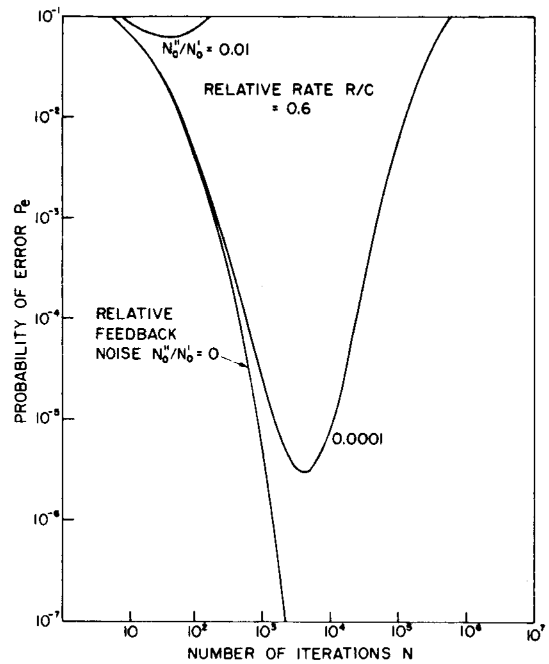


FIG. 12. PROBABILITY OF ERROR VS NUMBER OF ITERATIONS.

Note that the performance of the WB coding scheme, which achieves channel capacity with noiseless feedback, has been analyzed in the presence of feedback noise. Our results for noisy feedback, however, do not preclude the existence of systems that perform better with noisy feedback information.

III. A FEEDBACK COMMUNICATION SYSTEM WITH A CONSTRAINT ON THE BANDWIDTH

Let T be the time in seconds necessary for the transmission of a particular message. For the WB coding scheme discussed in the last chapter, as for orthogonal codes in one-way transmission, the bandwidth W(T) of the transmission is an exponential function of the coding delay T. In order to make the probability of error vanish for a fixed relative rate smaller than one, a large bandwidth is required.

Suppose now that one is given a fixed bandwidth W, which the transmission is not supposed to exceed. With this additional transmitter constraint imposed, the channel capacity C is no longer P_{av}/N_o as in Eq. (1.1a), but is now given by Eq. (1.1b), or W ln $[1 + (P_{av}/N_oW)]$, nats/sec. For small values of P_{av}/N_oW the latter capacity approaches that of Eq. (1.1a) as it should, for when $W \to \infty$ both channels are identical.

Shannon [Ref. 16] derives the capacity formula, Eq. (1.1b), by a random coding argument, and up till now no deterministic way was known for constructing a code achieving the critical rate for a band-limited white gaussian noise channel with or without feedback. In the present chapter the first such code will be developed for the case where noiseless feedback is available.

As in the preceding chapter, an optimization for finite block-length is carried through, the results are compared with bounds on one-way transmission plotted by Slepian [Ref. 10], and the deterioration of the present scheme due to feedback noise is considered.

A. THE BL CODING SCHEME

In the WB coding scheme discussed in the last chapter the variance of the estimate X_{N+1} for the message point θ was inversely proportional to the number N of iterations. The critical rate was R_{crit} = (ln N)/2T nats/sec, and in order to achieve a constant rate one had to choose $N = e^{2AT}$, that is, the number of transmissions had to increase exponentially with time.

Now suppose one has to meet a bandwidth constraint W in cycles per second. In this case the number of independent transmissions can only increase linearly with time. The highest number of independent transmissions per second is approximately equal to $2W$. Substituting $N = 2WT$ in the formula for the critical rate above gives $R_{crit} = (\ln 2WT)/2T$ nats/sec. Hence, $R_{crit} \to 0$ with increasing T, and so the system discussed in the last chapter has to be modified in order to achieve a constant rate different from zero in the bandlimited case.

Two useful observations can be made at this point. First, while the critical rate approaches zero when one takes $2W$ iterations per second the asymptotic relation $R_{crit}(T) \approx P_{av}(T)/N_o$ is still valid. In other words, both the rate and the average power approach zero for increasing T. The limit of their ratio, however, is equal to the constant N_o. The second observation is that X_{N+1} can be looked at as the maximum likelihood estimate of θ having observed $Y_1(X_1)$ through $Y_n(X_n)$, and assuming gaussian noise, as explained in Sec. IID.

With the above two observations in mind, a coding scheme can now be constructed for the bandlimited white gaussian noise channel.

Suppose that transmissions take place at integer values of time, the time unit being $1/(2W)$ sec. Numbers are sent again by amplitude modulating some basic waveform of bandwidth W and unit energy. The disturbance is white gaussian noise (with spectral density $N_o/2$), and reception takes place using a matched filter.

The coding scheme starts out the same as in Sec. IIB2.

At the transmitter:

1. Divide the unit interval $[0, 1]$ into M disjoint message intervals of equal length. Let θ be the midpoint of the message interval corresponding to the particular message to be transmitted.
2. At instant one, transmit $\alpha(X_{11} - \theta)$, where $X_{11} = 0.5$ and α is some constant to be determined later.

At the receiver:

1. Receive $Y_{11}(X_{11}) = \alpha(X_{11} - \theta) + Z_{11}$, where Z_{11} is as before a gaussian random variable with mean zero and variance $\sigma^2 = N_o/2$.
2. Compute $X_{12} = X_{11} - \alpha^{-1}Y_{11}(X_{11})$, then set $X_{21} = X_{12}$ and send X_{21} back to the transmitter.

Up to this point everything is the same as for the coding scheme of Sec. IIB2. In other words, $X_{21} - \theta = -(1/\alpha)Z_{11}$, where X_{21} is the maximum likelihood estimate of θ having observed $Y_{11}(X_{11})$.

Now in order to prevent the expected power per transmission to decrease, as it did in the WB coding scheme, the next transmission is $g\alpha(X_{21} - \theta)$ instead of $\alpha(X_{21} - \theta)$, where the constant g will be determined presently. The receiver obtains the noisy observation

$$Y_{21}(X_{21}) = g\alpha(X_{21} - \theta) + Z_{21}$$

and then computes

$$X_{22} = X_{21} - (g\alpha)^{-1}Y_{21}(X_{21})$$

One now has two independent estimates of θ:

$$X_{21} = \theta - \frac{1}{\alpha} Z_{11} \qquad \text{and} \qquad X_{22} = \theta - \frac{1}{g\alpha} Z_{21}$$

For the value X_{31} to be sent back to the transmitter, one takes the maximum likelihood estimate of θ having observed $Y_{11}(X_{11})$ and $Y_{21}(X_{21})$, that is,

$$X_{31} = \frac{\left(\frac{1}{g\alpha}\right)^2 X_{21} + \left(\frac{1}{\alpha}\right)^2 X_{22}}{\left(\frac{1}{\alpha}\right)^2 + \left(\frac{1}{g\alpha}\right)^2} = \frac{X_{21} + g^2 X_{22}}{1 + g^2}$$

What is the variance of our successive maximum likelihood estimates X_{11}, X_{21}, X_{31}? It is known that

$$X_{21} \sim N\left(\theta, \frac{\sigma^2}{\alpha^2}\right) \qquad \text{and} \qquad X_{31} \sim N\left(\theta, \frac{\sigma^2}{\alpha^2} \frac{1}{1 + g^2}\right)$$

If, however, $g = (\alpha^2 - 1)^{\frac{1}{2}}$ is chosen, then

$$X_{31} \sim N\left(\theta, \frac{\sigma^2}{(\alpha^2)^2}\right)$$

In general, X_{i1}, $i = 2, 3, \ldots,$ is sent back. The next transmission is

$$\alpha^{i-1} (\alpha^2 - 1)^{\frac{1}{2}} (X_{i1} - \theta) \tag{3.1}$$

but the receiver obtains

$$Y_{i1}(X_{i1}) = \alpha^{i-1} (\alpha^2 - 1)^{\frac{1}{2}} (X_{i1} - \theta) + Z_{i1}$$

and then computes

$$X_{i2} = X_{i1} - \left[\alpha^{i-1} (\alpha^2 - 1)^{\frac{1}{2}}\right]^{-1} Y_{i1}(X_{i1})$$

and

$$X_{i+1,1} = \frac{X_{i1} + (\alpha^2 - 1) X_{i2}}{\alpha^2}$$

The maximum likelihood estimate $X_{i+1,1}$ is normally distributed with mean θ and variance $\sigma^2/[(\alpha^2)^i]$, that is,

$$X_{i+1,1} \sim N \left[\theta, \frac{\sigma^2}{(\alpha^2)^i}\right] \tag{3.2}$$

From this point on, the analysis is very similar again to that of Chapter II. Suppose the transmitter sends one of M possible messages, that is, the interval $[0, 1]$ is divided into M disjoint equal-length message intervals. The message point θ is the midpoint of the message interval corresponding to the particular message being transmitted. The probability of the receiver deciding on the wrong message interval (i.e., the probability of X_{N+1} lying outside the correct interval) is [see Eq. (2.3a)]:

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

Now pick $M = \alpha^{N(1-\epsilon)}$, that is, $R = (\ln M)/N = (1 - \epsilon) \ln \alpha$, nats/dimension (the time unit was $1/(2W)$ sec). This gives for the probability of error

$$P_e = 2 \operatorname{erfc} \left(\frac{\alpha^{N\epsilon}}{2\sigma} \right) \tag{3.3}$$

and thus

$$\lim_{N\to\infty} P_e(N, \epsilon) = \begin{cases} 0 & \text{for } \epsilon > 0 \\ 1 & \text{for } \epsilon < 0 \end{cases}$$

In other words the critical rate is equal to $R_{crit} = \ln \alpha$, nats/dimension. Putting $\alpha = e^A$ gives $R_{crit} = A$.

Next let us derive an expression for the average power, P_{av}.

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

$$= \frac{1}{T} \left\{ \alpha^2 E(X_{11} - \theta)^2 + \sum_{i=2}^{N} \left[\alpha^{i-1} (\alpha^2 - 1)^{\frac{1}{2}} \right]^2 \frac{\sigma^2}{(\alpha^2)^{i-1}} \right\}$$

Substituting $T = N/(2W)$ sec, $\sigma^2 = N_o/2$, $\alpha = e^A$, and $E(X_{11} - \theta)^2 = 1/12$ (assuming a uniform prior distribution for θ), one gets

$$P_{av} = \frac{We^{2A}}{6N} + \frac{N-1}{N} N_o W (e^{2A} - 1) \tag{3.4}$$

Hence, asymptotically,

$$R_{crit} = \begin{cases} A = \frac{1}{2} \ln \left(1 + \frac{P_{av}}{N_o W} \right) & \text{nats/dimension, or} \\ 2WA = W \ln \left(1 + \frac{P_{av}}{N_o W} \right) & \text{nats/sec} \end{cases} \tag{3.5}$$

which is the channel capacity as computed by Shannon [Ref. 16]! This result proves that the BL coding system presented here achieves capacity for the bandlimited white gaussian noise channel. It is the first deterministic coding procedure to do so.

B. OPTIMIZATION FOR FINITE BLOCKLENGTH

As in Chapter II, let us now investigate how far one falls short of the ideal when only permitting a finite coding delay N (in time units of $1/(2W)$ sec).

In Sec. A the slope α_i at the i^{th} transmission was taken as

$$\alpha_i = \alpha^{i-1+\delta_{i1}} (\alpha^2 - 1)^{\frac{1}{2}(1-\delta_{i1})}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

In order to make an optimization possible, an additional factor a is introduced, hence

$$\alpha_i = a\alpha^{i-1+\delta_{i1}} (\alpha^2 - 1)^{\frac{1}{2}(1-\delta_{i1})} \qquad i = 1,2,3,\ldots \tag{3.6}$$

The receiver now has

$$Y_{i1}(X_{i1}) = \alpha_i(X_{i1} - \theta) + Z_{i1}$$

and computes

$$X_{i2} = X_{i1} - \alpha_i^{-1} Y_{i1}(X_{i1}),$$

$$X_{21} = X_{12},$$

and

$$X_{i+1,1} = \frac{X_{i1} + (\alpha^2 - 1)X_{i2}}{\alpha^2} \qquad \text{for } i = 2, 3, \ldots$$

Effectively, the introduction of the factor a reduces N_o by a factor a^2.

By Eq. (3.3) minimizing the probability of error is equivalent to maximizing the expression

$$a^2 \frac{\alpha^{2N\epsilon}}{2N_o} \qquad (3.7)$$

where ϵ can be obtained from

$$R = (1 - \epsilon) \ln \alpha \qquad \text{nats/dimension} \qquad (3.8)$$

and $\sigma^2 = N_o/2$ was substituted for the variance.

Substituting $\alpha = e^A$ in Eq. (3.4) and allowing for the additional factor a leads to the following expression for the average power.

$$P_{av} = a^2 \left[\frac{W\alpha^2}{6N} + \frac{N-1}{N} \frac{N_o}{a^2} W(\alpha^2 - 1) \right]$$

which can be modified as

$$\frac{P_{av}}{N_o W} = a^2 \frac{\alpha^2}{6NN_o} + \frac{N-1}{N} (\alpha^2 - 1) \qquad (3.9)$$

Now assuming C, R, W, N_o, and N constant, let us maximize expression (3.7) with respect to a^2. Note that C and W constant implies P_{av}/N_oW constant, for

$$C = W \ln \left(1 + \frac{P_{av}}{N_o W} \right) \qquad (1.1b)$$

Having gone through these preliminaries, one is now ready to perform the optimization. Set the derivative of $a^2(\alpha^{2N\epsilon}/2N_o)$ equal to zero,

$$\frac{d}{da^2}\left(a^2 \frac{\alpha^{2N\epsilon}}{2N_o}\right) = \frac{\alpha^{2N\epsilon}}{2N_o} + a^2 \frac{d}{d\alpha^2}\left(\frac{\alpha^{2N\epsilon}}{2N_o}\right)\frac{d\alpha^2}{da^2} + a^2 \frac{d}{d\epsilon}\left(\frac{\alpha^{2N\epsilon}}{2N_o}\right)\frac{d\epsilon}{d\alpha^2}\frac{d\alpha^2}{da^2} = 0 \qquad (3.10)$$

From Eq. (3.8) it follows that

$$\epsilon = 1 - \frac{2R}{\ln \alpha^2} \qquad \text{and} \qquad \frac{d\epsilon}{d\alpha^2} = \frac{2R}{\alpha^2} \, \frac{1}{(\ln \alpha^2)^2}$$

and from Eq. (3.9) it follows that

$$\frac{d\alpha^2}{da^2} = - \frac{\alpha^2}{a^2 + 6N_o(N-1)}$$

Making these substitutions in Eq. (3.10) and putting the result equal to zero finally gives, after some algebra, the following simple expression for the optimum value a_o^2 of a^2:

$$a_o^2 = 6N_o \tag{3.11}$$

For the probability of error, substituting $\sigma^2 = N_o/2$ and $a_o^2 = 6N_o$ in Eq. (3.3), one has

$$P_e = 2 \text{ erfc} \left[\left(3\alpha^{2N\epsilon} \right)^{\frac{1}{2}} \right]$$

Solving for α^2 from Eqs. (3.9) and (3.11) gives

$$\alpha^2 = \frac{N-1}{N} + \frac{P_{av}}{N_o W}$$

By Eq. (3.8) one has

$$\frac{R}{2W} = \ln \alpha^{1-\epsilon} \qquad \text{or} \qquad \alpha^{1-\epsilon} = \exp\left(\frac{R}{2W}\right)$$

where R is now in nats/second. Hence,

$$\alpha^{\epsilon} = \frac{\alpha}{\exp\left(\frac{R}{2W}\right)} = \frac{\left(\frac{N-1}{N} + \frac{P_{av}}{N_o W} \right)^{\frac{1}{2}}}{\exp\left(\frac{R}{2W}\right)}$$

and finally,

$$P_e = 2 \operatorname{erfc} \left\{ \sqrt{3} \left[\frac{\frac{N-1}{N} + \frac{P_{av}}{N_o W}}{\exp\left(\frac{R}{W}\right)} \right]^{N/2} \right\} \qquad (1.2c)$$

This final result will be compared in the next section with the bounds on one-way communication as obtained by Slepian [Ref. 10].

C. COMPARISON WITH SLEPIAN'S RESULTS

In 1963 Slepian [Ref. 10] plotted lower bounds on communication in the one-way case based on a geometrical approach to the coding problem for bandlimited white gaussian noise channels used by Shannon [Ref. 6]. That is, there is no one-way communication system whose performance is any better than that plotted by Slepian. Figures 13 through 18 compare Slepian's curves (dashed lines) with the results described by Eq. (1.2c) (solid lines). Note that the solid curves are exact; that is, they are not a bound as Slepian's curves are. The graphs presented in this section are described below.

1. Figure 13 shows the signal-to-noise ratio $S/N = 10 \log_{10} (P_{av}/N_o W)$ in decibels vs the rate R/W in dits/cycle, as given by Shannon's capacity formula, Eq. (1.1b).
2. Figures 14a-c indicate the additional signal-to-noise ratio, in decibels above the value indicated in Fig. 13, required for a finite coding delay N, as a function of the rate in dits per cycle. The probability of error for the three figures is respectively, $P_e = 10^{-2}$, 10^{-4}, and 10^{-6}. It is seen that a large improvement is obtained by going from $N = 5$ to $N = 15$, especially in the feedback scheme. Increasing the coding delay further does not result in much improvement.
3. Figures 15a,b are plots of the additional signal-to-noise ratio in decibels above the ideal value indicated in Fig. 13 vs the coding delay N, for different values of the probability of error P_e and for a rate of $R/W = 0.2$ dit per cycle. Figure

15b represents a plot for the bounds computed by Slepian. Note that the curves for the feedback scheme (Fig. 15a) indicate a much lower relative (to the ideal, given in Fig. 13) signal-to-noise ratio, except for extremely small values of N.

4. Figures 16a,b are plots of the probability of error vs the coding delay N, with the signal-to-noise ratio in decibels above the ideal as the parameter. The rate is $R/W = 0.2$ dit per cycle. Note the difference in shape between the two sets of curves.
5. Figure 17 is a plot of the relative rate R/C vs the rate R/W in dits per cycle for different values of the coding delay. The probability of error is $P_e = 10^{-4}$.
6. Figures 18a,b are plots of the relative rate R/C vs the coding delay N for different values of the signal-to-noise ratio.

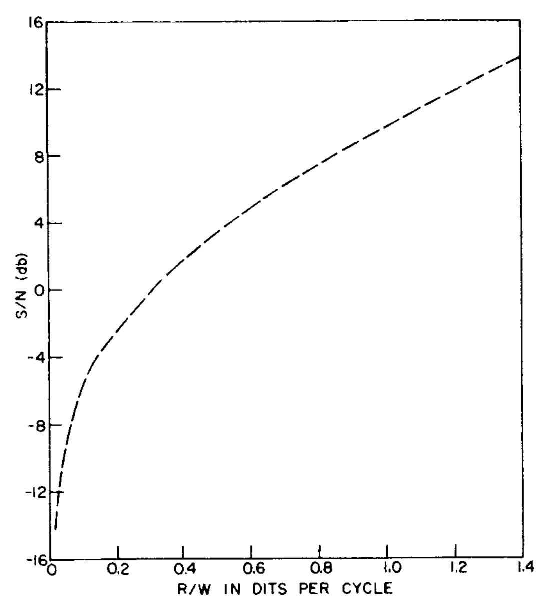
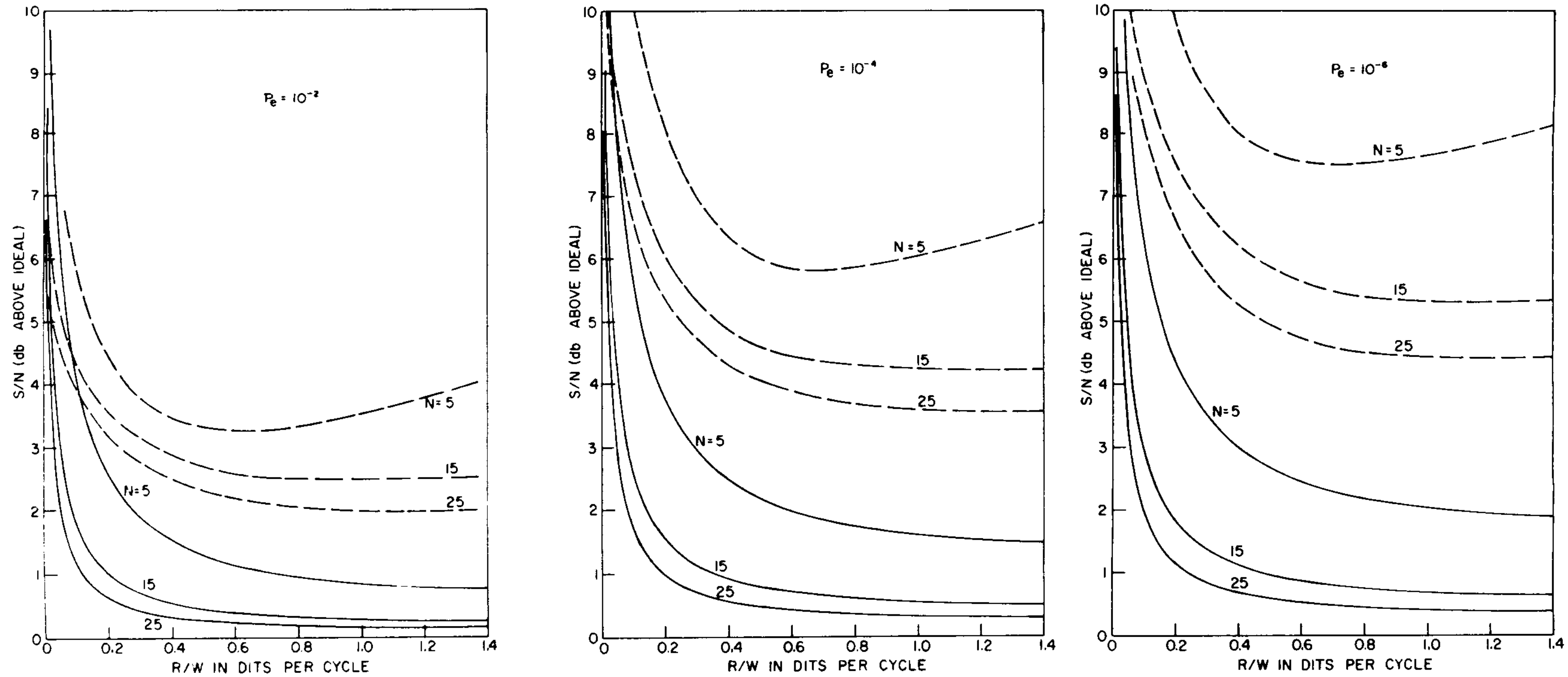
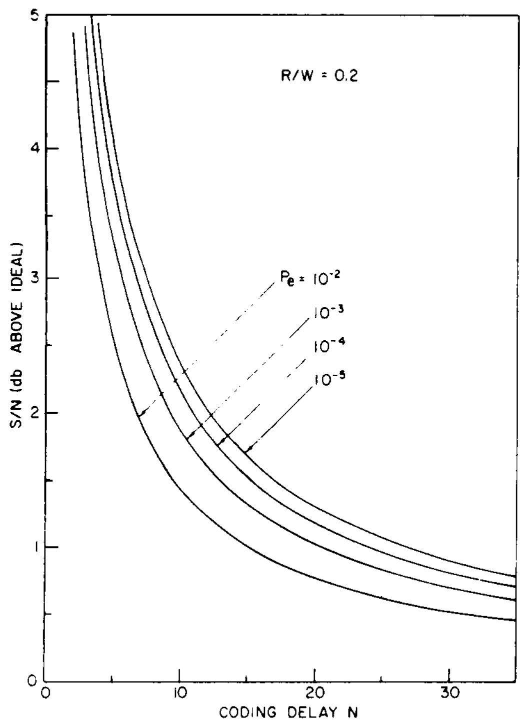


FIG. 13. THE SIGNAL-TO-NOISE RATIO REQUIRED BY SHANNON'S CAPACITY FORMULA.

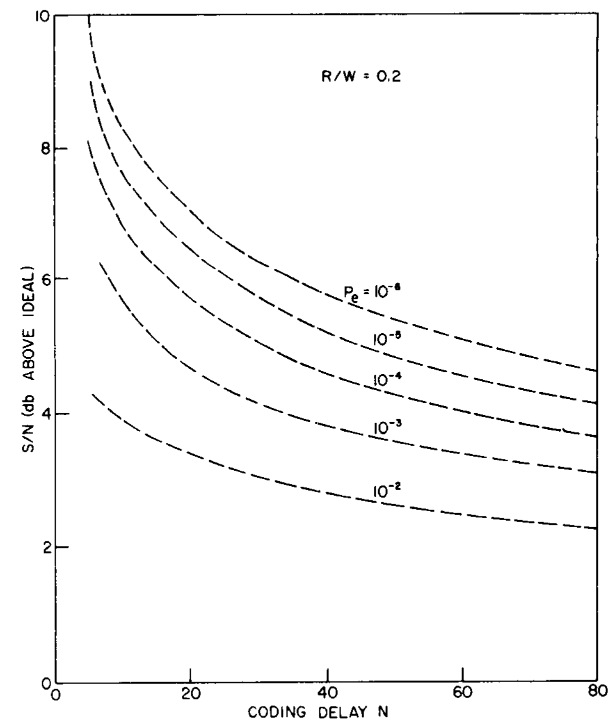


a. $P_e = 10^{-2}$ b. $P_e = 10^{-4}$ c. $P_e = 10^{-6}$

FIG. 14. THE ADDITIONAL SIGNAL-TO-NOISE RATIO REQUIRED WHEN USING A FINITE CODING DELAY.

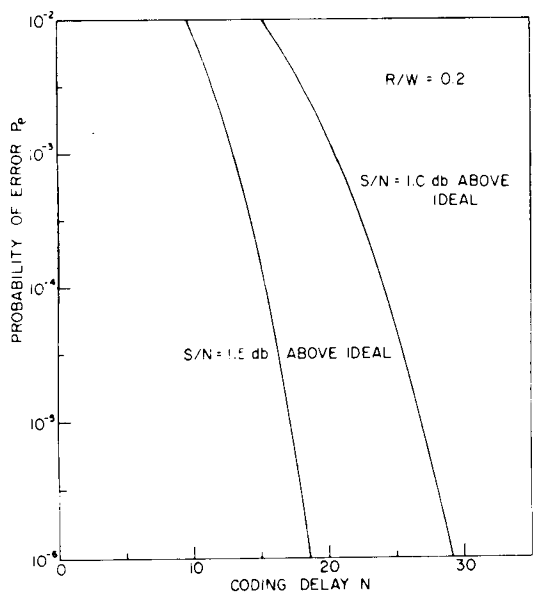


a. BL coding scheme

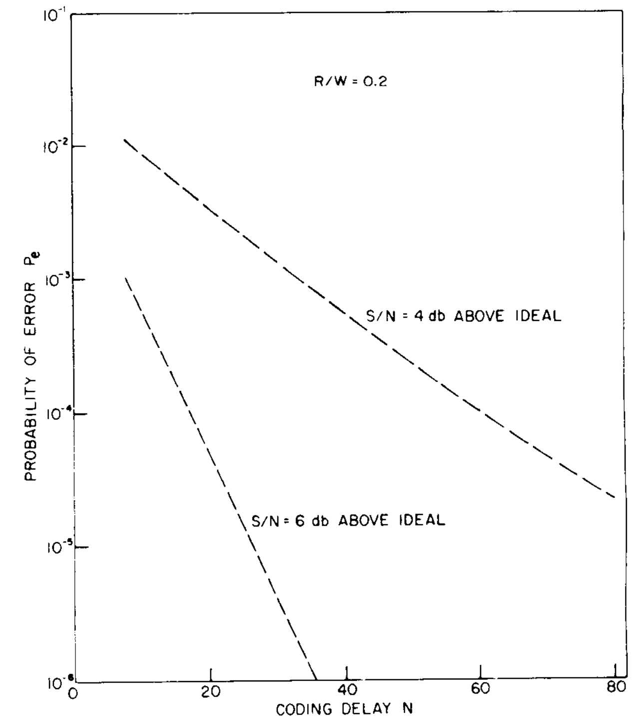


b. Bounds on one-way communication

FIG. 15. THE ADDITIONAL SIGNAL-TO-NOISE RATIO AS A FUNCTION OF THE CODING DELAY FOR DIFFERENT VALUES OF THE PROBABILITY OF ERROR.

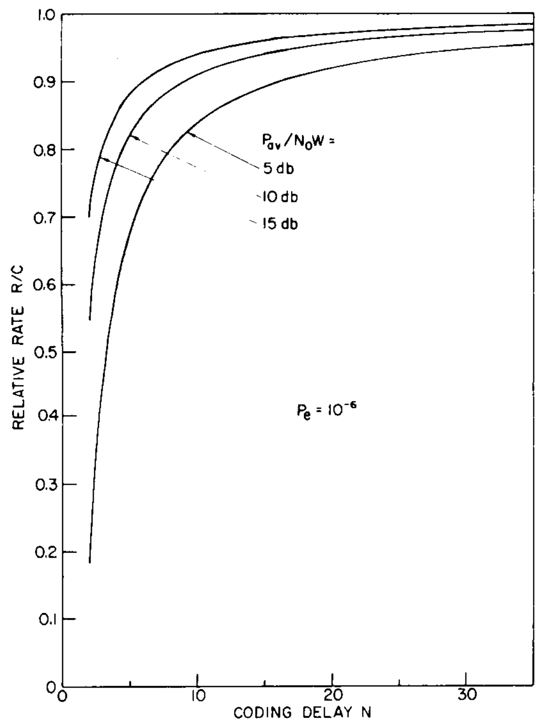


a. BL coding scheme

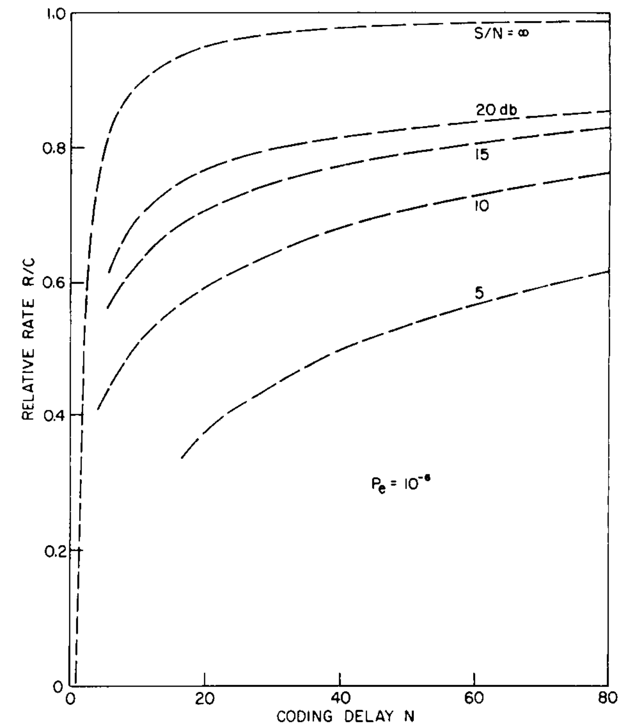


b. Bounds on one-way communication

FIG. 16. THE PROBABILITY OF ERROR AS A FUNCTION OF THE CODING DELAY FOR DIFFERENT VALUES OF THE RELATIVE SIGNAL-TO-NOISE RATIO.



a. BL coding scheme



b. Bounds on one-way communication

FIG. 18. THE RELATIVE RATE VERSUS THE CODING DELAY FOR DIFFERENT VALUES OF THE SIGNAL-TO-NOISE RATIO.

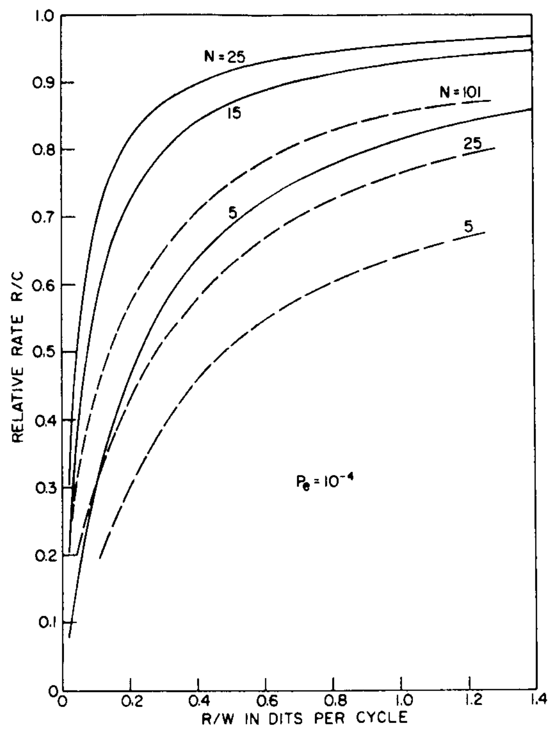


FIG. 17. THE RELATIVE RATE VERSUS THE RATE PER UNIT BANDWIDTH FOR DIFFERENT VALUES OF THE CODING DELAY.

D. INFLUENCE OF FEEDBACK NOISE ON THE BL CODING SCHEME

In this section only the configuration in which $Y'_n(X''_n)$ (the received "number") is sent back will be investigated. The results for the case where X_n (the receiver's estimate) is sent back are similar to those in the last chapter in that the rate drops off to zero quickly.

Using the same notation as in Chapter II, it follows easily that

$$X'_{N+1,1} = X''_{N+1,1} + \alpha_1^{-1} Z''_{11} + \frac{\alpha^2 - 1}{\alpha^2} \sum_{i=2}^{N} \alpha_i^{-1} Z_{i1} \qquad (3.12)$$

where $\sum_{1=2}^{1} = 0$, and α_i is given by Eq. (3.6). Hence:

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

The variance σ_t^2 of the estimate X'_{N+1} of θ, as computed by the receiver, is

$$\sigma_t^2 = \frac{1}{a^2} \left[\frac{\sigma'^2 + \sigma''^2}{\alpha^{2N}} + \frac{\sigma''^2}{\alpha^2} + \sigma''^2 \left(\alpha^{2(N-1)} - 1\right)\right] \qquad (3.13)$$

For the probability of error one has, from Eq. (3.3),

$$P_e = 2 \operatorname{erfc} \left[\frac{\alpha^{N(1-\varepsilon)}}{2\sigma_t}\right] \qquad (3.14)$$

where again $R = (1 - \varepsilon) \ln \alpha$, nats/dimension.

The expression for the signal-to-noise ratio in the forward direction is, from Eq. (3.9),

$$\frac{P_{av}}{N'_o W} = a^2 \frac{\alpha^2}{6NN'_o} + \frac{N-1}{N} \frac{N'_o + N''_o}{N'_o} (\alpha^2 - 1) \qquad (3.15)$$

Figure 19 presents curves for the probability of error P_e vs the coding delay N for $R/W = 0.2$ dit per cycle, and different values of the feedback noise relative to the forward noise, N_o''/N_o'. For a^2 the value $a^2 = 6N_o'$ as given by Eq. (3.11) is used. Hence, the curves present the degradation due to feedback noise of a system that is optimum for the noiseless feedback case.

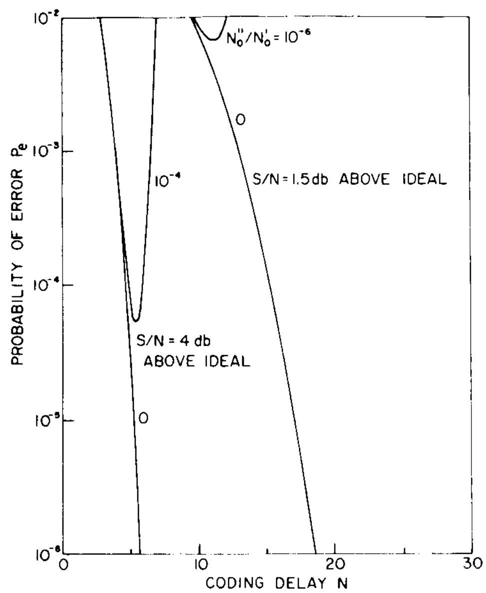


FIG. 19. THE PROBABILITY OF ERROR AS A FUNCTION OF THE CODING DELAY FOR DIFFERENT VALUES OF THE RELATIVE SIGNAL-TO-NOISE RATIO IN THE PRESENCE OF FEEDBACK NOISE.

IV. CONCLUDING REMARKS

The WB (wideband) coding scheme was suggested by the Robbins-Munro stochastic approximation procedure. In the gaussian case it turns out that this coding procedure determines the maximum likelihood estimate of the message point θ recursively. Since the maximum likelihood estimate approaches θ and the transmitted power is proportional to the square of the difference, the expected transmitted power per iteration decreases in this scheme. Retaining the maximum likelihood property but making up for the transmitted power in order to make the expected power per transmission a constant, leads to the BL (bandlimited) coding scheme. This simple scheme is the first deterministic procedure to achieve the channel capacity, Eq. (1.1b), of the bandlimited white gaussian noise channel.

It is believed that this approach of recursive maximum likelihood estimation to the coding problem with feedback has a much wider area of application; for example, channels with unknown parameters, fading channels, dependences between the noises in forward and feedback links, and so on. The method is ideally suited for noiseless feedback and it may well be possible to find an extension that is in some sense optimum for the noisy feedback case.

REFERENCES

1. P. E. Green, "Feedback Communication Systems," Lectures on Communication System Theory, Baghdady (ed.) McGraw-Hill Book Company, 1961.

2. A. J. Viterbi, "The Effect of Sequential Decision Feedback on Communication over the Gaussian Channel," Info. and Contr., 8, 1, Feb 1965.

3. M. Horstein, "Sequential Transmission Using Noiseless Feedback," IEEE Trans. on Information Theory, Jul 1963.

4. G. L. Turin, "Signal Design for Sequential Detection Systems," Int. Tech. Memo. M-69 University of California, Berkeley, 6 May 1964.

5. P. Elias, "Channel Capacity Without Coding," Lectures on Communication System Theory, Baghdady (ed.) McGraw-Hill Book Company, New York, 1961.

6. C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," Bell Sys. Tech. J., 38, 3, 1959, pp. 611-656.

7. J. Wolfowitz, "Coding Theorems of Information Theory," Springer-Verlag, Berlin, 1961.

8. C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," IRE Trans. on Information Theory, IT-2, 3, Sep 1956, pp. 8-19.

9. R. M. Fano, "Transmission of Information," John Wiley and Sons, Inc., New York, 1961.

10. D. Slepian, "Bounds on Communication," Bell Sys. Tech. J., 42, 1963, pp. 681-707.

11. H. Robbins and S. Munro, "A Stochastic Approximation Method," Ann. Math. Stat., 22, 1951, pp. 400-407.

12. J. Wolfowitz, "On Stochastic Approximation Methods," Ann. Math. Stat., 27, 1956, pp. 1151-56.

13. A. Dvoretsky, "On Stochastic Approximation," Proc. of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1956.

14. J. H. Venter, "On Stochastic Approximation Methods," Ph.D. Thesis, University of Chicago, 1963.

15. J. Sacks, "Asymptotic Distributions of Stochastic Approximation Procedures," Ann. Math. Stat., 29, 1958, pp. 373-405.

16. C. E. Shannon, "A Mathematical Theory of Communication," <u>Bell Sys. Tech. J.</u>, <u>27</u>, 1948, pp. 379-424 and 623-657.