NOISELESS LINEAR FEEDBACK AND ANALOG DATA TRANSMISSION*

By S. BUTMAN

Summary. It is well known that noiseless linear feedback achieves channel capacity for the additive Gaussian channel. It has also been shown that it can be used to achieve the rate-distortion bound on the mean squared error for an arbitrary Gaussian source sent over the infinite bandwidth white Gaussian channel. However, it is shown here that noiseless linear feedback by itself does not suffice when the channel is bandlimited. It is shown that, out of the more than countable variety of Gaussian sources that ordinarily exist, only a countable subset can be transmitted via the bandlimited noiseless feedback link at the theoretical efficiency predicted by Shannon's rate-distortion bound. Thus, some nonlinear operations are necessary in almost all cases even with feedback.

Introduction. Elias [1] and others [2-8] have shown that noiseless linear feedback (See Fig. 1) can be used to transmit a Gaussian random variable θ of variance λ over the additive Gaussian white noise (AGWN) channel so as to achieve the theoretically minimum mean-square error (MMSE), also called the rate-distortion-bound (RDB) on the mean-square error,

$$\epsilon^2_{min} = \lambda e^{-2CT} \tag{1}$$

predicted by Shannon's rate distortion theory [9-11], where $C = W\ln(1 + P/N_oW)$ nats/sec is the capacity of an AGWN channel of bandwidth W Hz, average transmitter power P watts, and one-sided noise spectral power density N_o watts/Hz, and where T in seconds is the duration of the transmission.

This article shows when the above result (1) of Elias can be applied to achieve the RDB on the MMSE for an arbitrary Gaussian source emitting the Gaussian random process $\theta(t)$, $0 \leq t \leq T$, over the AGWN channel. Specifically, it is shown here that noiseless linear feedback can achieve the RDB for a given source and bandlimited channel pair if and only if the spectrum if eigenvalues $\{\lambda_i\}$ of the Karhunen-Loeve (K-L) series expansion of the source satisfies

$$\lambda_i = \phi \left(1 + \frac{P}{N_oW}\right)^{N_i} \tag{2}$$

for all $\lambda_i > \phi$, where ϕ is a constant for the given source-channel pair

---

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

Dr. Butman is at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103.

(given by equations (11) in the sequel) and N_i is a positive integer for all $i \epsilon I \triangleq \{i: \lambda_i > \phi\}$ such that

$$\sum_{i \epsilon I} N_i = N \triangleq 2TW \tag{3}$$

In this case

$$\epsilon^2_{min} = \sum_{i \epsilon I} \phi + \sum_{i \epsilon I'} \lambda_i \tag{4}$$

where $I' = \{i: \lambda_i \leq \phi\} \cdot N_i = 0$ for $i \epsilon I'$.

Preliminaries. Before proceeding with the discussion it is appropriate to recall Shannon's result that the minimum amount of information that must be provided about a Gaussian random variable θ of known mean and known variance λ so as to reproduce it to within a mean squared error ϵ^2 is

$$H(\epsilon) = \tfrac{1}{2} \ln(\lambda/\epsilon^2) \tag{5}$$

The corresponding minimum transmission rate $R(\epsilon)$ (called the rate relative to the distortion ϵ) is

$$R(\epsilon) = \frac{1}{2T} \ln (\lambda/\epsilon^2). \tag{6}$$

We can invert this and write

$$\epsilon^2(R) = \lambda e^{-2RT} \tag{7}$$

or

$$\epsilon^2(H) = \lambda e^{-2H} \tag{8}$$

as the MMSE that can be achieved in any facsimile $\hat{\theta}$ of θ when H nats of information (sent at the rate R nats/sec for T seconds) are provided about θ. This simple result can be immediately applied to a sequence $\{\theta_i\}$ of independent Gaussian random variables of corresponding variances $\{\lambda_i\}$.

Let H be the total information that can be supplied about the sequence $\{\theta_i\}$ and let H_i be the information supplied about an individual element of this sequence θ_i. Then (8) implies a MMSE per θ_i of

$$\epsilon^2 = \sum \epsilon_i^2 = \sum \lambda_i e^{-2H_i} \tag{10}$$

Now, if we are free to allocate information as we please, there is an optimum allocation given by minimizing ϵ^2 over H_i subject to the constraints $H_i \geq 0$ and $\Sigma H_i = H$.

This minimum can be found simply by setting the derivative of $\epsilon^2 + \phi H$, where ϕ is a Lagrange multiplier, equal to zero. The result is $H_i = \frac{1}{2} \ln (\lambda_i/\phi)$ provided $\lambda_i > \phi$; otherwise $H_i = 0$ because H_i must be positive or zero. Consequently, we have the well-known relationships

$$\epsilon_{min}(H) = \sum_{i \epsilon I} \phi + \sum_{I \epsilon I'} \lambda_i$$

$$H = \frac{1}{2} \sum_{i \epsilon I} \ln(\lambda_i/\phi) \tag{11}$$

as first stated in Kolmogorov's paper [10]. Moreover, the proceeding minimization also exhibits the optimum allocation of information, per θ_i, namely,

$$H_i = \begin{cases} \frac{1}{2}\ln\left(\lambda_i/\phi\right) & \text{if } \lambda_i > \phi \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where ϕ is calculated for the given H and $\{\lambda_i\}$ from [11].

Finally, let us recall the K-L series expansion for a Gaussian random process $\theta(t)$ on $0 \le t \le T$. The K-L series is

$$\theta(t) = \underset{K \to \infty}{\text{l.i.m.}} \sum_{i=1}^{K} \theta_i \psi_i(t) \tag{13}$$

where $\{\psi_i(t)\}$ is the set of deterministic orthogonal eigenfunctions of the kernel $R(t,u) = \underset{\sim}{E}[\theta(t)\theta(u)]$ which satisfy

$$\lambda_i \psi_i(t) = \int_0^T R(t,u)\, \psi_i(u)du, \tag{14}$$

and the sequence of coefficients $\{\theta_i\}$ is a sequence of independent Gaussian random variables of corresponding variances $\{\lambda_i\}$. Here $\underset{\sim}{E}$ denotes expectation and l.i.m denotes limit in the mean.

The Limitation of Linear Feedback Over the Bandlimited Channel. So far we have provided nothing that has been unknown before. It seems that if we knew of a scheme achieving capacity for the AGWN channel, we could use it to transmit H = CT nats of information about $\theta(t)$. By sending information $H_i = CT_i$ about θ_i, where $H_i = \frac{1}{2}\ln(\lambda_i/\phi)$, we would receive θ_i with an error of not more than ϕ and thus achieve the RDB. Alternatively, we could subdivide the channel into subchannels of capacity C_i in such a way that $H_i = C_i T$. Unfortunately, however, although linear feedback achieves capacity, it cannot by itself provide such an arbitrary division of information when the channel bandwidth is finite, for an integral number of iterations must be used per message, hence per θ_i. Thus, we can transmit θ_i using N_i feedback iterations (N_i forward transmissions and N_i-1 feedback transmissions). The minimum error that can be achieved per θ_i in this way is

$$\epsilon_i^2 = \lambda_i \left(1 + \frac{2E_i}{N_o N_i}\right)^{-N_i}, \tag{15}$$

where E_i is the total average energy transmitted per θ_i for the entire N_i feedback iterations. We note that the information transmitted per θ_i when using this approach is

$$H_i = \tfrac{1}{2}N_i \ln\left(1 + \frac{2E_i}{N_o N_i}\right); \tag{16}$$

the maximum amount of information that can be sent is $CT = \frac{1}{2}N\ln(1+P/N_oN)$. This is possible if and only if $E_i/N_oN_i = PT/N_oN$, where the N_i must sum to N and E_i's sum to E = PT. Any other choice will result in the transmission of less than CT nats of information. Consequently

$$H_i = \tfrac{1}{2}N_i \ln\left(1 + \frac{P}{N_o W}\right) \tag{17}$$

is adjustable only in steps of $\ln(1 + P/N_oW)$ for a fixed P, W, and N_o. Therefore, we cannot provide the optimum information allocation for an arbitrary source spectrum. The optimum allocation is possible if and only if the spectrum of eigenvalues satisfies

$$\tfrac{1}{2}\ln \lambda_i/\phi = \tfrac{1}{2}N_i \ln\left(1 + \frac{P}{N_oW}\right)$$

when $\lambda_i > \phi$; which is equivalent to the condition in (2).

The Infinite Bandwidth Case. This case has been previously solved by Cruise [6] and Fang [8], however, we include it here for completeness. We note that the preceeding limitation on the bandlimited channel disappears when the bandwidth is infinite. For $N = 2WT \to \infty$ as $W \to \infty$, and we can use an infinite number of iterations $N_i = \infty$ per θ_i in order to transmit

$$H_i = \lim_{N_i \to \infty} \frac{1}{2N_i} \ln\left(1 + 2E_i/N_oN_i\right) \quad (18)$$

$$= \frac{E_i}{N_o} \text{ nats per } \theta_i \quad (19)$$

Since $H = E/N_o$ when $N = \infty$, and since $E = \Sigma E_i$, we can adjust the energies E_i in a continuous manner in order to allocate

$$E_i/N_o = \tfrac{1}{2}\ln(\lambda_i/\phi) \text{ nats} \quad (20)$$

for θ_i and thus achieve the RDB for any Gaussian source over the infinite bandwidth channel.

Equivalence of Frequency and Time Multiplexed Schemes. We are led to consider only two equivalent feedback communication schemes, or more generally, only two variants of any capacity-achieving code, namely, time-division multiplex or frequency-division multiplex. Both reduce to one general representation when expressed in terms of degrees of freedom $N = 2WT$. In the case of feedback, degrees of freedom are equivalent to feedback iterations. Thus the N_i degrees of freedom or iterations necessary to achieve an error $\epsilon_i^2 = \lambda_i(1 + 2E_i/N_oN_i)^{N_i}$ can be obtained by transmitting N_i pulses, each of duration $\Delta = 1/2W$, for a total time per θ_i of $T_i = N_i\Delta = N_i/2W$; or N_i pulses each of duration $\Delta_i = T/N_i = 1/2W_i$ can be transmitted per θ_i, where W_i is the bandwidth allocated for the transmission of θ_i and $C_i = W_i\ln(1 + P_i/N_oW_i) = W_i\ln(1 + P/N_oW)$ is the capacity allotment per θ_i. The latter approach is the one originally used by Elias. The main point is that $E_i/N_iN_o = P/N_oW = \rho$, which is the signal-to-noise ratio per channel signal, is a constant. Thus, the input to the channel appears white and expanded in bandwidth.

The MMSE Achievable with Linear Feedback if (2) is not Satisfied. It is instructive to consider an example where (2) is not satisfied. The simplest case is provided by any bandlimited Gaussian source whose spectrum is flat over a bandwidth W_s. However, since the usual meaning of a bandlimited spectrum implies signals of infinite duration ($\theta(t)$, $\infty \le t \le \infty$), our time-limited process, $\theta(t)$, $0 \le t \le T$, will have infinite bandwidth and hence an infinite K-L expansion. Nevertheless, if we order the sequence of coefficients $\{\theta_i\}$ according to $\lambda_1 \ge \lambda_2 \ge \ldots$, we will need to retain only N_s terms, where N_s is determined from $\lambda_{N_s} > \phi \ge \lambda_{N_s} + 1$. We assume, therefore, that $\lambda_i = \lambda > \phi$ for $i \le N_s$. Thus, the spectrum of eigenvalues is flat in the transmission band

$W_s \triangleq N_s/2T$.

Next, suppose that $N = 2WT$ is not an integral multiple of N_s; in fact, let $N = KN_s + L$, where K and L are integers, but $1 \leq L < N_s$, so that $0 < L/N_s = \alpha < 1$. The RDB is

$$\epsilon^2_{min} = N_s\phi = N_s\lambda(1 + \rho)^{-(K + \alpha)} \qquad (21)$$

where $\rho = P/N_oW = E/N_oN$. If this is to be achieved we must transmit $H_i = \frac{1}{2}(K + \alpha)\ln(1 + \rho)$ nats per θ_i. Since α is not an integer, we see that noiseless linear feedback alone cannot provide the best result. With noiseless linear feedback we have the choice of making some N_i's greater than K, whereupon others must be less than K because they all must sum to N. It turns out that the best linear noiseless feedback can do is to achieve an error of

$$\epsilon^2 = L\,\lambda(1 + \rho)^{-(K + 1)} + (N_s - L)\lambda(1 + \rho)^{-K} \qquad (22)$$

This is obtained by allocating K feedback interations per θ_i, i = 1, ..., N_s. The remaining L degrees of freedom are not sufficient for an additional iteration of all the θ_i's. However, they can be used to make one fore iteration of L out of N_s θ's or, L iterations of any linear combination of all $N_s\theta$'s, or something in between. It turns out that (22) is independent of how the remaining L degrees are used if the coding is linear and the source spectrum is flat over W_s.

Similar results prevail when the λ's are not all the same. The only difference is that we initially take N_i as the smallest integer in $\ln(\lambda_i/\phi)/\ln(1 + \rho)$. This leaves a remainder of

$$L = N - \sum_{i=1}^{N_s} N_i < N_s \qquad (23)$$

iterations still to be used. Again, if only linear operations are to be used, we can use these L degrees of freedom for one more iteration of some of the θ's, but not all. Hence, we can achieve

$$\epsilon^2 = \sum_{i=1} \lambda_i(1 + \rho)^{-(N_i + 1)} + \sum_{i=L+1} \lambda_i(1 + \rho)^{-N_i} + \sum_{i=N_s+1} \lambda_i, \qquad (24)$$

where the last term reminds us of the error in that portion of $\theta(t)$ about which no information is sent.

The best allocation of the L remaining iterations can be found. Define

$$\alpha_i = \ln(\lambda/\phi)/\ln(1 + t) - N_i\ , \quad 0 \leq \alpha_i < 1, \qquad (25)$$

and note that

$$L \triangleq \sum_{i=1}^{N_i} \alpha_i\ . \qquad (26)$$

Then, after using N_i iterations per θ_i, the error is

$$\epsilon^2 = \sum_{i=1}^{N_s} \lambda_i (1+\rho)^{-N_i} + \sum_{i=N_s+1}^{\infty} \lambda_i , \tag{27}$$

but $\lambda_i = \phi(1+\rho)^{N_i+\alpha_i}$; hence

$$\epsilon^2 = \phi \sum_{i=1}^{N_s} (1+\rho)^{\alpha_i} + \sum_{i=N_s+1} \lambda_i \geq \epsilon^2_{min} \tag{28}$$

unless $\alpha_i = 0$ for all $i = 1, \ldots, N_s$.

Now, it is clear that to reduce ϵ^2 farther we should reduce the largest terms in the sum (28). Thus, we allocate one more iteration for those L θ's that have the largest α's.

For notational convenience assume $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_{N_s}$. Then the MMSE resulting from linear feedback alone is

$$\epsilon^2 = \phi \sum_{i=1}^{N_s} (1+\rho)^{\beta_i} + \sum_{i=N_s+1}^{\infty} \lambda_i \tag{29}$$

where $\beta_i = -1 + \alpha_i$ for $i \leq L$ and $\beta_i = \alpha_i$ for $i > L$. Since $\Sigma\beta_i = L-\Sigma\alpha_i=0$, we see that the first sum in (29) is lower bounded by N_s. This proves that the error given by (29) satisfies

$$\epsilon^2 \geq N_s\phi + \sum_{i=N_s+1} \lambda_i = \epsilon^2_{min} \tag{30}$$

with equality if and only if $\beta_i = 0$ for all $i = 1, \ldots, N_s$; thus, if and only if $L = 0$, for if $L \neq 0$ $\alpha_1 = 1 - \beta_1 = 1$ is impossible.

Discussion. We have stated that the effect of linear feedback on the source relative to the channel is to expand the N_s source symbols into N channel symbols, hence to expand the source bandwidth W_s to match the channel bandwidth W, and to make the expanded source appear white. Another interpretation is to view feedback as compressing the channel bandwidth and transforming the originally white noise spectrum into a colored spectrum (if the source spectrum is not flat) so as to match the channel to the given source. Both viewpoints tie in very well with some results of Ovseyevich and Pinsker [12] for one-way linear communication. They showed that a necessary and sufficient condition for achieving the RDB under average power constraints on the transmitted signals is that the product of the source spectrum and the noise spectrum in the transmission band to be constant.

To see how this result is obtained it is necessary only to realize that the total entropy H over a one-way colored channel with two-sided noise spectrum $\{N_{oi}/2\}$ and average signal energies $\{E_i\}$ is a maximum over $\{E_i\}$ under a total average energy constraint on the transmitter if and only if $2E_i + N_{oi} = \mu$, where μ is a Lagrange multiplier. Then the entropy per E_i is $H_i = \frac{1}{2}\ln(\mu/N_{oi}) = \frac{1}{2}\ln(1 + 2E_i/N_{oi})$. At the same time we know that to achieve the RDB H_i must equal $\frac{1}{2}\ln(\lambda_i/\phi)$. Therefore, $\lambda_i N_{oi} = \mu\phi$ must be a constant for all $\lambda_i > \phi$. Since $\lambda_i = \phi(1+\rho)^{N_i}$ we see that $N_{oi} = \mu(1+\rho)^{-N_i}$. This agrees with the general consequences of feedback, namely, that the equivalent noise in a single transmission of energy E_i must give a signal-to-noise ratio $\rho_{oi} = 2E_i/N_{oi}$ equal to $(1+\rho)^{N_i} - 1$.

We shall conclude this topic by observing that weighting coefficients $\{a_i\}$ of the pre-emphasis filter for sending the source over the compressed and "colored" channel are defined by

$$a_i^2 = E_i/\lambda_i = (\mu - N_{oi})/2\lambda_i = N_{oi}(\mu - N_{oi})/2\mu\phi.$$

Thus,

$$a_i^2 = \frac{\mu}{2\lambda_i}\left(1 - \frac{\phi}{\lambda_i}\right) = \frac{\mu\phi}{2\lambda_i}\left(\frac{1}{\phi} - \frac{1}{\lambda_i}\right) , \qquad (31)$$

indicating that the optimum filter suppresses those K-L coefficients that are either too strong or too weak. This follows from the observation that $|a_i|^2$ increases from zero as λ_i increases from ϕ to a maximum of $\mu/8\phi$ when $\lambda_i = 2\phi$, but eventually decreases back down to zero inversely with λ_i as $\lambda_i \to \infty$. However, the energy E_i increases steadily with λ_i as $E_i = \mu(1-\phi/\lambda_i)$. Also, although the signals out of the pre-emphasis filter have different energies, the N signals carrying information over the original AGWN channel all have the same average energy $e = E_i/N_i$. Thus, relative to the AGWN channel we have a bandwidth-expanding pre-emphasis filter with N_s inputs, $N = \Sigma N_i$ outputs, and N_i weighting coefficients b_{ij}, $j = 1, \ldots, N_i$ for each i, $i = 1, \ldots, N_s$ given by $b_{ij}^2 = e/\lambda_i$. It simply whitens the source spectrum to match it to the white noise spectrum of the channel, in accord with the theory. But this whitening is accompanied by a bandwidth expansion and is not the same as the "usual" whitening precess in which the bandwidth is preserved. The usual prewhitenning is not optimum.

The last subject appropriate for this paper is the question of what to do when condition (2) is not satisfied. We hasten to point out that feedback has nothing to do with the answer. The only thing that feedback permits is a bandwidth expansion of the source to match the channel bandwidth, and it does so only in integral multiples of the source bandwidth. Thus, feedback is out of the question when $W_s \geq W$. More precisely, feedback is a factor only when the N degrees of freedom in the channel exceed the N_s degrees of freedom of the source.

Now, the solution to our problem would evolve from a solution of the nonlinear source encoding problem that arises when we attempt to ram N_s independent source variables through a channel using only $N \leq N_s$ degrees of freedom. For example, the RDB for a source with, say, a flat spectrum $\lambda_i = \lambda > \epsilon_{min}^2/N_s$ should be

$$\epsilon_{min}^2 = N_s\lambda(1+\rho)^{-\beta} \qquad 0 < \beta = \frac{N}{N_s} < 1 \qquad (32)$$

We say a nonlinear coding problem because we know that a linear many-to-one transformation, aside from not having an inverse, will not work.

The simplest of all situations is when $\beta = \frac{1}{2}$. In this case we are faced with the problem of mapping the complex plane onto the real line. That is, a pair of source symbols θ_1 and θ_2, say, must be mapped onto a single channel signal, x, which can be transmitted only once. The receiver observes a noisy version of x, say, x + n, and must reconstitute both θ_1 and θ_2 from it in such a way that

$$\hat{\theta}_1 = \hat{\theta}(x + n), \quad \hat{\theta}_2 = \hat{\theta}_2(x + n)$$

has an error of

$$\begin{aligned} \epsilon^2 &= E[(\theta - \hat{\theta}\ (x + n))^2 + (\theta - \hat{\theta}\ (x + n))^2] \\ &= 2\lambda(1 + \rho)^{\frac{1}{2}} = \epsilon^2_{min} \end{aligned} \qquad (33)$$

We must somehow determine these mappings $\hat{\theta}_1(x + n)$, $\hat{\theta}_2(x + n)$ and the mapping $x - x(\theta_1, \theta_2)$, where ρ may be a parameter.

Initial thoughts seem to indicate that a partial quantization, i.e., quantizing θ_1, but allowing a continuous mapping for θ_2 is unavoidable. Complete quantization is probably not much worse. In any case, it is still to be established, even in this simple case, now close to the theoretical bound of $2\lambda(1 + \rho)^{\frac{1}{2}}$ the two-to-one mappings will go. Shannon proved that such mappings exist for all W and W_s. Their discovery would be of considerable practical use, quite apart from linear feedback theory.

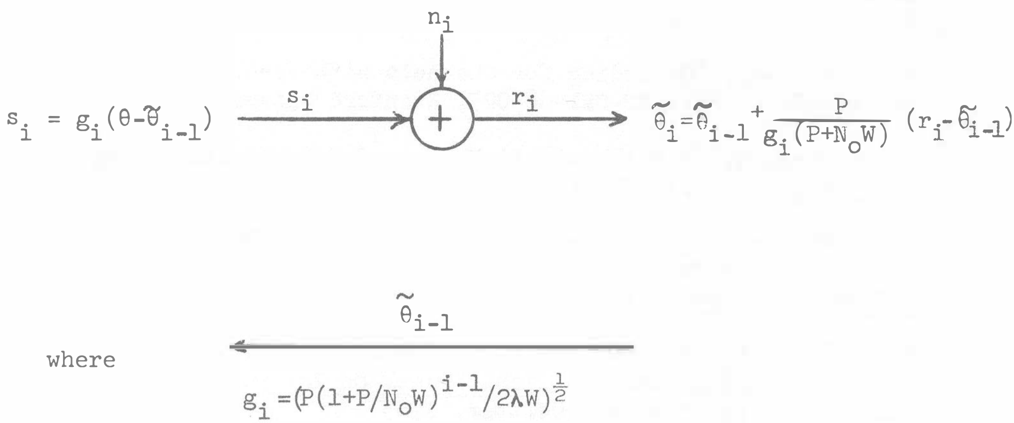

Figure 1. Optimal Noiseless Linear Feedback Scheme for Transmitting θ

REFERENCES

1. P. Elias, Channel Capacity Without Coding, Quarterly Progress Report, Res. Lab of Elec. MIT, 90-93, Oct. 1956. Also in Lectures on Communication System Theory, Ed., Baghdady, New York, McGraw-Hill Book Company, 363-366, 1961.

2. P. Elias, Networks of Gaussian channels with applications to feedback systems, IEEE Trans. on Information Theory, IT-13, 493-501, Jul. 1967.

3. J. K. Omura, Optimum linear Transmission of Analog Data for Channels With Feedback, IEEE Trans. on Information Theory, IT-14, No. 1, Jan. 1968.

4. J. P. M. Schalkwijk and L. I. Bluestein, Transmission of Analog Waveforms Through Channels with Feedback, IEEE Trans. on Information Theory, IT-13, No. 4, Oct. 1967.

5. T. Kailath, An Application of Shannon's Rate-distortion Theory to Analog Communication Over Feedback Channels, Presented at the First Annual Princeton Conference on Information Science and Systems, 30-31 Mar. 1967. Also in Proc. IEEE Letters, 55, No. 6, 1102-1103, Jan. 1967.

6. T. J. Cruise, "Achievement of Rate-Distortion Bound Over Additive White Noise Channel Utilizing a Noiseless Feedback Link," Proc. IEEE Letters, Vol. 55, No. 4, pp. 583-584.

7. Butman, S., "A General Formulation of Linear Feedback Communication Systems with Solutions," IEEE Trans. on Information Theory, Vol. IT-15, May, 1969.

8. R. J. F. Fang, "On Coding for Channels with Feedback," Ph.D. Dissertation, Rpt. SU-SEL-68-025, Stanford University, March 1968.

9. C. E. Shannon, "A Mathematical Theory of Communication," Bell Sys. Tech. J., Vol. 27, pp. 379-623, July and October, 1948.

10. A. N. Kolmogorov, "On the Shannon Theory of Information Transmission in the Case of Continuous Signals," IRE Trans. on Information Theory, pp. 102-108, December 1956.

11. T. J. Goblick, "Theoretical Limitations on the Transmission of Data From Analog Sources," IEEE Trans. on Information Theory, Vol. IT-11, pp. 558-567, October, 1965

12. I. A. Ovseyevich and M. S. Pinsker, "Optimum Linear Pre-emphasis and Correction," Engineering Cybernetics, No. 5, pp. 39-46, 1963.