# PHASE-INCOHERENT FEEDBACK COMMUNICATION

Thesis by

Jerry Butman

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1969

(Submitted July 1, 1968)

ACKNOWLEDGMENT

The benefit I have derived from Professor Thomas L. Grettenberg, my research advisor, has resulted not only from suggestions and comments during the many discussions, but from acquiring some of his unusual insight into communication problems. His guidance of my post-graduate studies has been genuinely appreciated. The appreciation (and pride) I have for my wife's competent typing of the thesis is second only to an appreciation of the encouragement and understanding she has shown. I wish to thank Dr. S. M. Farber for his work on some of the early details in the paper, and Dr. S. S. Lavenberg for the many useful discussions. I am, of course, grateful for the financial assistance afforded me by a Royal W. Sorensen Fellowship and a Ford Foundation Fellowship.

## ABSTRACT

Nonlinear feedback communication schemes proposed up to now have been restricted to coherent channels. This paper describes a scheme which not only lacks this restriction but yields performance better than the coherent ones that have been analyzed. Block coded orthogonal signals together with incoherent receiver forms are used on both forward and feedback links. It is assumed that the transmitter has a high peak power capability. However, it is shown that by increasing the code length the duty cycle of this mode can be made sufficiently small so that the contribution to average power is negligible. It is found that the probability of error for every message is upper-bounded by $\exp[-E(R)T]$ , where $T$ represents the code length and $E(R)$ is a function of the transmission rate, $R$ , the capacities of the forward and feedback channels, and a parameter that determines the rate with which the duty cycle decreases with increasing code length. When the forward and feedback capacities are equal, say $C$ , the $E(R)$ versus $R$ curve can be made to approach a curve that starts at $2C$ and decreases monotonically to the value $C$ at $R = C$ . This contrasts with the corresponding curve without feedback which starts at $C/2$ and decreases to zero at $R = C$ . The main advantage of information feedback over no feedback, namely $E(R = C) > 0$ , where $C$ is the forward channel capacity, can also be obtained even when the capacity of the feedback channel is less than that of the forward. To obtain such behavior, existing schemes require that the feedback channel capacity be at least as great as the forward.

# TABLE OF CONTENTS

## LIST OF ILLUSTRATIONS

# INTRODUCTION

It is not uncommon in engineering practice to find communication systems that use information feedback to provide highly reliable communication over an otherwise mildly reliable channel. The method is most often restricted to a very simple procedure. For example, in teletype systems the receiving equipment may ask for a re-transmission whenever certain types of errors are detected in the incoming stream of data. The fact that simple procedures such as this yield significant reductions in the number of errors raises the question of just how much improvement feedback can be made to effect.

Many authors have investigated this problem. The important question of whether or not feedback can increase the capacity of the forward channel was first answered by Shannon [1]. He showed that even a noiseless feedback link cannot be made to increase the capacity of the zero-memory binary symmetric channel. Lavenberg [6] has extended this result to the zero-memory time-discrete amplitude-continuous channel. When the forward channel has memory, however, Butman [2] has shown that its capacity can be increased. When the signal processing in the forward and feedback channels is restricted to linear operations (Linear Feedback Communication) efficient schemes have been designed by Butman [2], Schalkwijk and Kailath [3], Omura [5] and others. Accordingly, when the feedback channel is noiseless it is found that linear feedback can yield dramatically improved performance over the best non-feedback schemes. Unfortunately, the slightest amount of noise in the feedback link serves to degenerate every linear feedback communication scheme that is known. This difficulty is not encountered with nonlinear feedback schemes. In particular, Kramer [7] has analyzed a scheme for the phase-coherent channel which also yields striking improvement over communication without feedback, but which tolerates a small amount of feedback noise. Lavenberg [6] has improved upon this by devising schemes that tolerate almost as much noise in the feedback channel as there is in the forward.

In this paper a scheme will be given which (1) yields improvement over non-feedback communication for any finite amount of feedback noise, (2) outperforms all existing schemes for any given average forward and feedback power constraints, and (3) operates on the phase-incoherent as well as on the phase-coherent channel.

# CHAPTER I. SYSTEM DESCRIPTION

## 1.1 Introduction

The model usually used for the phase-incoherent channel is employed throughout the paper. The results in chapters 2 and 3, however, are also valid if the channel is phase-coherent, provided all coding and decoding operations remain unchanged. Thus a comparison is possible between the feedback communication scheme presented here and previous schemes designed only for the coherent channel.

A description of the scheme is given first in terms of the sequencing of events in time. In this way the peak power, average power, bandwidth and synchronization requirements can readily be deduced. A more complete description is provided by resorting to vector representation.

## 1.2 Channel Model

It is assumed that the forward and feedback channels available for communication can each be modeled in the following way.

Input: $s$

Output: $se^{i\theta} + n$

where the random phase $\theta$ is uniformly distributed on $[0, 2\pi]$ and $n$ is a zero-mean complex Gaussian [10] random vector with covariance matrix $N_1 I$ on the forward channel and $N_2 I$ on the feedback channel. In general the signal vector $s$ may have complex components.

Actual channels which can be modeled in this way abound in engineering practice. Notable examples are systems based on ionospheric or tropospheric scatter, but any system in which the phase of the carrier is not regulated or tracked falls equally well into the category. It is assumed that the noise in these systems is additive, zero-mean, white and Gaussian. Reference [8] gives an excellent account of the details in deriving the model for DSB - SC modulation on the ionospheric scatter channel. In this case it turns out that $s$ is real. However, if quadrature multiplexing is used, the signal vector becomes complex.

## 1.3 Time-Waveform Description

To facilitate in the description to be given it is assumed that information is continually being generated by a source at the rate of R bits/sec. As for reliable communication without feedback the information is first accumulated for $T$ seconds. This results in a sequence of $RT$ bits which might be any one of $M_1 = 2^{RT}$ possible sequences. In a non-feedback scheme the communication of this sequence is then accomplished by sending one of $M_1$ waveforms over the next $T$ - second interval. However, in order to utilize a feedback channel the following scheme is proposed:

The $T$ - second interval available for communicating the sequence is divided into three equal parts. Initial communication is performed by transmitting one of $M_1$ orthogonal waveforms over the interval $[0, T/3]$ with power $P_1$ watts. Upon reception of the signal perturbed by channel noise, the receiver decides on the most likely waveform and decodes it into one of the $M_1$ possible sequences. To inform the transmitter of its decision the receiver then communicates the first $aRT$ bits of this sequence, $a \le 1$, by sending one of $M_2 = 2^{aRT}$ orthogonal waveforms over the interval $[T/3, 2T/3]$. Based on the signal it observes and knowledge of the sequence it is trying to communicate the transmitter makes an estimate of the receiver's decision. If its estimate is the same as the sequence it is trying to communicate, it assumes the receiver's decision was correct, and transmits zero power on the remaining interval $[2T/3, T]$. Otherwise it makes use of this interval to re-transmit the original waveform, but with power $\alpha^2 P_1$ watts, $\alpha \ge 1$. The initial decision at the receiver remains unchanged if no signal is detected on the last interval. Otherwise the old decision is discarded and a new one made, based entirely on the signal that is observed.

## 1.4 Peak Power, Average Power, and Bandwidth

The average energy $\bar{E}$ expended in the forward direction is

$$\bar{E} = P_1 T/3 + \alpha^2 P_1 T/3 \cdot p$$

where $p$ is the probability that a re-transmission is necessary. It will later be shown that $p$ can be made sufficiently small so that $\bar{E} \cong P_1 T/3$. Thus while the peak power required is $\alpha^2 P_1$ the average power requirement $\bar{P}_1$ is only $P_1/3$. On the feedback channel $\bar{E} = P_2 T/3$ so that while the peak power is $P_2$ the average power $\bar{P}_2$ is only $P_2/3$.

Extension of the scheme to more than one feedback transmission is relatively straightforward. Figure 1 in the next section illustrates the sequence of events when N feedback transmissions are used. In this case the T - second interval must be divided into 2N + 1 equal parts. Consequently, the average forward and feedback powers become $\bar{P}_1 = P_1/2N+1$ and $\bar{P}_2 = NP_2/2N+1$, respectively.

At first glance the bandwidth requirement seems to be prohibitive for large N . It is bad enough that it grows as $2^{RT}/T$ when block coded orthogonal signaling is used without feedback. With feedback the requirement becomes $(2N + 1)2^{RT}/T$ since the same number, $2^{RT}$, of orthogonal waveforms must now be constructed on an interval $T/2N+1$. As will later be shown, however, using a coding delay $\delta T$, $(\delta \ll 1)$, with feedback is as good if not better, as far as probability of error is concerned, than a coding delay T without feedback. Thus to achieve the same performance a feedback scheme requires a bandwidth of only $\frac{(2N + 1)}{\delta T} 2^{R\delta T}$. This may easily be less than that required by the non-feedback system.

## 1.5 Vector Description

The vector channel model given in section 1.2 will now be the basis for a complete description of the feedback communication scheme. The scheme begins with the transmission of the vector $x_o$ and sequences through the events shown in figure 1.

The $M_1$ - dimensional vector $s_m$ is one vector in the set, $\mathcal{S}$ of $M_1$ orthogonal vectors with length $\sqrt{E_1} = \sqrt{P_1 T/2N+1}$,

$$\mathcal{S} = \left\{ s_i : (s_i, s_j) = E_1 \delta_{ij} , \quad i,j = 1,\ldots,M_1 \right\} ,$$

where $(s_i,s_j)$ denotes the Euclidean inner product of $s_i$ and $s_j$. The initial estimate of $s_m$ is $\hat{s}_o$, obtained from the observation $y_o$ by the decision rule:

$$D_o : \text{ pick } \hat{s}_o = s_k \text{ where } |(y_o, s_k)|^2 = \max_i |(y_o, s_i)|^2 .$$

The set $\mathcal{S}$ is partitioned into $M_2$ disjoint subsets, $Q_1,\ldots,Q_{M_2}$, each containing the same number of components. Associated with each subset is a unique $M_2$ - dimensional vector chosen from the set of orthogonal vectors

$$\mathcal{T} = \left\{ \sigma_i : (\sigma_i, \sigma_j) = E_2 \delta_{ij} , \ i,j = 1,\ldots,M_2 \right\}$$

where $E_2 = P_2 T/2N+1$ .

The vector for the first transmission over the feedback channel is determined from $\hat{s}_o$ by the rule:

$$G : \text{ send } w_1 = \sigma_i \text{ if } \hat{s}_o \in Q_i .$$

The estimate of $w_1$ is $\hat{w}_1$, determined from $s_m$ and the observation $z_1$ by the decision rule:

$D$ : find $m'$ such that $s_m \in Q_{m'}$, and pick

$\hat{w}_1 = \sigma_{m'}$ only if $|(z_1,\sigma_{m'})|^2 \geq \max_{i \neq m'} \left\{ |(z_1,\sigma_i)|^2 + \gamma^2 E_2^2 \right\}$. Otherwise choose $\hat{w}_1 = \sigma_k$ where

$$|(z_1,\sigma_k)|^2 = \max_{i \neq m'} |(z_1,\sigma_i)|^2 .$$

The operation denoted by $F$ determines $x_1$ from $\hat{w}_1$ and $s_m$ according to

$$x_1 = \begin{cases} 0 & \text{if} \quad s_m \in Q_i \\ s_m & \text{if} \quad s_m \notin Q_i \end{cases} , \quad \text{where} \quad \hat{w}_1 = \sigma_i .$$

Decision rules $D_1,\ldots,D_N$ are modified forms of $D_o$ which account for the possibility of a null decision. For each $n = 1,\ldots,N$ the rule is

$D_n$ : pick $\hat{x}_n = 0$ only if $|(y_n, s_i)|^2 < \beta_n^2 E_1^2$ for all $i = 1,\ldots,M_1$. Otherwise choose $\hat{x}_n = s_k$ where

$$|(y_n, s_k)|^2 = \max_i |(y_n, s_i)|^2 .$$

Operation $H$ determines $\hat{s}_1$ from $\hat{x}_1$ and $\hat{s}_o$ by the simple assignment

$$\hat{s}_1 = \begin{cases} \hat{s}_o & \text{if} \quad \hat{x}_1 = 0 \\ \hat{x}_1 & \text{if} \quad \hat{x}_1 \neq 0 \end{cases} .$$

$\hat{s}_2$ is similarly derived from $\hat{x}_2$ and $\hat{s}_1$, the sequence of forward and feedback transmissions continuing until $\hat{s}_N$, the final estimate of $s_m$, is obtained.

The statistics of the random quantities $\theta_k$, $n_k$ $(k = 0,1,\ldots,N)$ and $\varphi_k$, $\eta_k$ $(k = 1,\ldots,N)$ have already been described in section 1.2. The only additional information necessary before analysis begins is that they are all statistically independent. The parameters $\gamma$, $\alpha_1,\ldots, \alpha_N$, $\beta_1,\ldots,\beta_N$ are real numbers whose specification is delayed until the performance is derived in terms of them.

## 1.6 Full and Partial Information Feedback

The number of code words in library $\mathcal{S}$ is $M_1 = 2^{RT}$ while in $\mathcal{T}$ it is $M_2 = 2^{aRT}$. When $a = 1$ the receiver is able to communicate to the transmitter its exact initial decision $\hat{s}_o$ because, $M_2$ being equal to $M_1$, there is a one-to-one correspondence between the elements of $\mathcal{T}$ and those of $\mathcal{S}$. Certainly the receiver cannot communicate more information about $\hat{s}_o$. Thus this case is called full information feedback. When $a < 1$ the receiver communicates only the subset of $\mathcal{S}$ to which $\hat{s}_o$ belongs. This case is appropriately called partial information feedback.

It is natural, of course, to expect that partial feedback cannot yield as good a performance as full feedback. Its attractiveness, however, is more than academic since, as will later be shown, the main advantage of full feedback over no feedback, namely, a positive value for the reliability curve at $R = C$, can also be obtained with partial feedback. Thus highly reliable communication without undue coding complexity is possible even when the capacity of the feedback channel is much less than that of the forward. Moreover, partial feedback may be necessary even when the feedback capacity is greater than the forward. This can happen if the receiver has information of its own to communicate. If it is being generated at a high rate the residual feedback capacity available for improving the performance of the forward channel can easily be much less than the forward channel capacity.

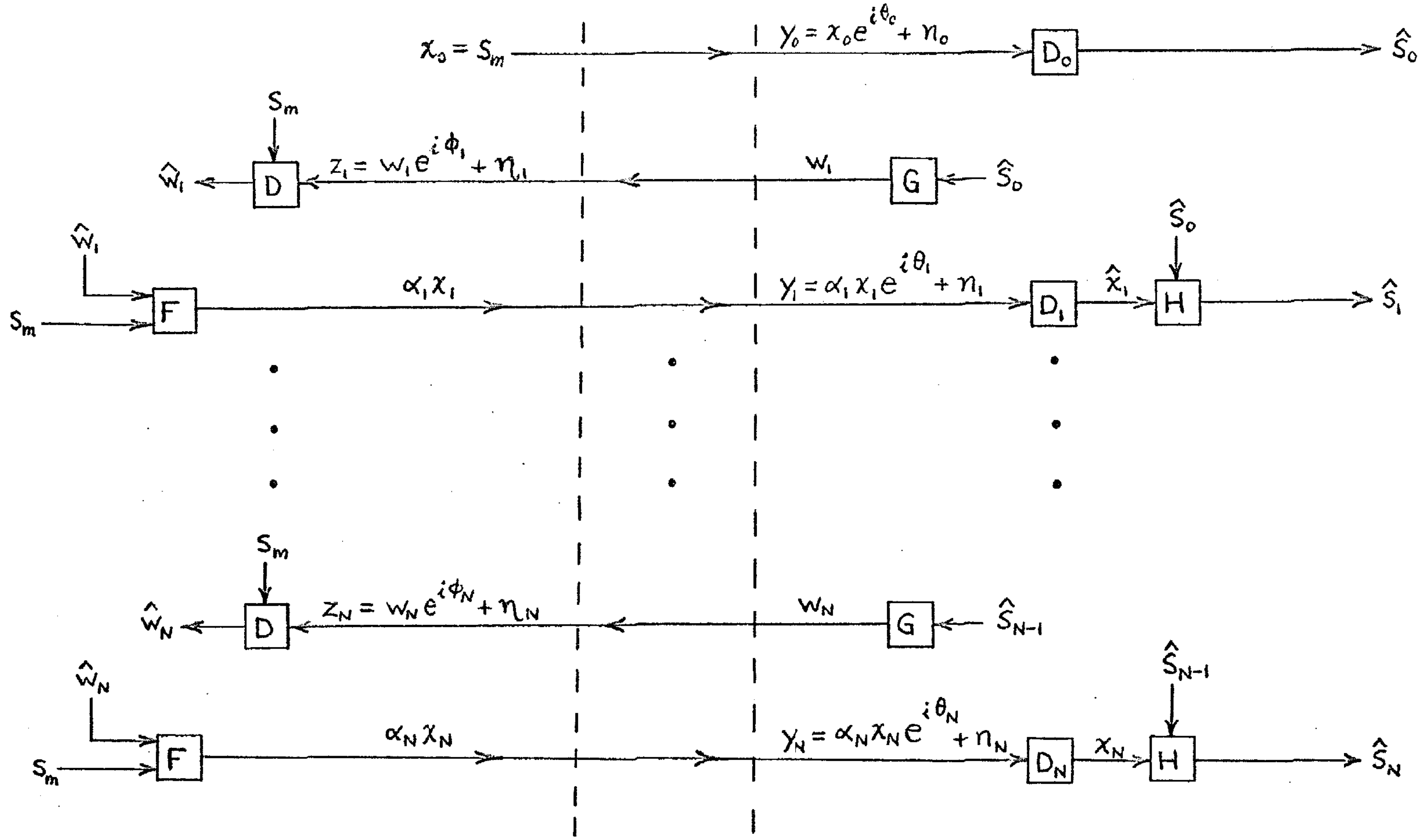


Figure 1 : The Feedback Communication Scheme

# CHAPTER II. FULL INFORMATION FEEDBACK

## 2.1 Introduction

The main object of this chapter will be the derivation of system performance when only one use of the feedback channel is permitted, it being used to provide maximum feedback information. Extension to an arbitrary number of feedback transmissions can be accomplished in a similar, almost identical, way. Rather than actually going through the details, however, it will be shown that the best performance attainable with more than one feedback transmission is worse than what is attainable with only one feedback transmission. That this result should not be taken seriously will be demonstrated by giving a modified scheme with improved performance. Surprisingly enough, however, it turns out that even this system cannot outdo the single feedback scheme. It just matches it. This can only mean one of two things. Either there is a still better way of using the feedback channel, or there is no point in ever using it more than once per message. Indeed, it is advantageous to make only a single feedback transmission since this requires less bandwidth (see section 1.4). The existence of a better feedback communication scheme is, of course, an open question, but it is noteworthy to point out that thus far this one yields the best probability of error.

## 2.2 Calculation of $p(e)$

From here on the information rate $R$ will be taken in units of nats per second, permitting the size of the sets $\mathcal{S}$ and $\mathcal{T}$ to be expressed as $M_1 = e^{RT} = M_2$ . The symbol $p(\cdot)$ will denote the probabil-

..ty of the event in brackets given that $s_m$ has been chosen for communication.

Analysis is begun by writing the probability of error

$$p(e) \triangleq p(\hat{s}_1 \neq s_m)$$

$$= \sum_{n=1}^{4} p(e|A_n)p(A_n) \qquad (2.1)$$

where

| | | | |
|---|---|---|---|
| | $A_1$ | denotes the event that | $\hat{s}_o = s_m$ and $\hat{w}_1 = \sigma_m$ , |
| | $A_2$ | " " " " | $\hat{s}_o = s_m$ but $\hat{w}_1 \neq \sigma_m$ , |
| | $A_3$ | " " " " | $\hat{s}_o \neq s_m$ but $\hat{w}_1 = \sigma_m$ , |
| and | $A_4$ | " " " " | $\hat{s}_o \neq s_m$ and $\hat{w}_1 \neq \sigma_m$ , |

$\sigma_m$ being the vector in $\mathcal{T}$ which corresponds to the vector $s_m$ in $\mathcal{S}$ . Now if $A_1$ occurs it can be seen that since $\hat{w}_1 = \sigma_m$ the device F will set $x_1 = 0$ . Also since $\hat{s}_o = s_m$ the device H will produce an error, $\hat{s}_1 \neq s_m$ , if and only if the estimate $\hat{x}_1$ is neither 0 nor $s_m$ . Thus

$$p(c|A_1) = p(\hat{x}_1 \neq 0 \text{ or } s_m \mid x_1 = 0) .$$

The decision rule $D_1$ indicates furthermore that

$$p(\hat{x}_1 \neq 0 \text{ or } s_m \mid x_1 = 0) = p\Big( |(y_1, s_i)|^2 > \beta_1^2 E_1^2 \text{ for some } i, \text{ and } |(y_1, s_j)|^2 > |(y_1, s_m)|^2 \text{ for some } j \neq m \;\Big|\; x_1 = 0 \Big) .$$

But for any two events X and Y the probability $p(X \cap Y)$ is less than either of $p(X)$ or $p(Y)$. So letting X be the event "$|(y_1, s_i)|^2 > \beta_1^2 E_1^2$ for some $i$" and Y the event "$|(y_1, s_j)|^2 > |(y_1, s_m)|^2$ for some $j \neq m$" it follows that

$$p(e|A_1) \leq p\Big( |(y_1, s_i)|^2 > \beta_1^2 E_1^2 \text{ for some } i \;\Big|\; x_1 = 0 \Big) . \quad (2.2)$$

Application of the union bounding technique [8] yields

$$p(e|A_1) < M_1 \, p\Big( |(y_1, s_i)|^2 > \beta_1^2 E_1^2 \;\Big|\; x_1 = 0 \Big)$$
$$= M_1 \, p\Big( |(n_1, s_i)|^2 > \beta_1^2 E_1^2 \Big)$$

where the equality follows from $y_1 = \alpha_1 x_1 e^{i\theta_1} + n_1$ and the given condition $x_1 = 0$. Since $n_1$ is complex Gaussian with zero mean and covariance $N_1 I$, and $\|s_i\|^2 = E_1$ it follows that $|(n_1, s_i)|^2$ has the Exponential probability density with mean $N_1 E_1$. Thus

$$p(e|A_1) < M_1 \, e^{-\beta_1^2 E_1 / N_1} .$$

Recalling that $M_1 = e^{RT}$, $E_1 = P_1 T/3$ and defining $C_1 \triangleq \dfrac{P_1/3}{N_1}$

yields

$$p(e|A_1) < e^{-T(\rho_1^2 C_1 - R)} \quad . \tag{2.3}$$

The considerations which led to (2.2) also indicate that

$$p(e|A_2) = p(\ \hat{x}_1 \neq 0 \text{ or } s_m \mid x_1 = s_m\ )$$

$$\leq p\left(\ |(y_1, s_j)|^2 > |(y_1, s_m)|^2 \quad \text{for some } j \neq m \ \middle|\ x_1 = s_m\right)$$

But the term on the right is just the probability of error in communicating one of $M_1 = e^{RT}$ messages over an incoherent channel by using equal energy orthogonal signals and a type $D_o$ decision rule. It is known [9] that this is bounded above by $\exp[-TCE(R/C)]$, where $C$ is the channel capacity defined as the signal power-to-noise ratio, and

$$E(x) = \begin{cases} 1/2 - x \quad , & 0 < x \leq 1/4 \\ (1 - \sqrt{x})^2 \, , & 1/4 < x \leq 1 \\ 0 \quad , & x > 1 \end{cases} .$$

It should be recalled that the signal energy in the observation $y_1$ is $\alpha_1^2 E_1 = \alpha_1^2 P_1 T/3$ . Thus $C = \alpha_1^2 P_1/3N_1 = \alpha_1^2 C_1$ and it follows that

$$p(e|A_2) < e^{-T\alpha_1^2 C_1 E(R/\alpha_1^2 C_1)} \quad . \tag{2.4}$$

Event $A_3$ represents the case when the transmitter fails to detect an error in the receiver's initial estimate. Since it is almost certain that the receiver will persist in its error, the trivial bound

$$p(e|A_3) \leq 1 \tag{2.5}$$

is appropriate. Event $A_4$, on the other hand, represents the case when the transmitter does detect an error in the receiver's decision. It will thus re-transmit $s_m$ with energy $\alpha_1^2 E_1$. Now

$$\begin{aligned} p(e|A_4) &= p(\hat{x}_1 \neq s_m \mid x_1 = s_m) \\ &= p(\hat{x}_1 = 0 \mid x_1 = s_m) + p(\hat{x}_1 \neq 0 \text{ or } s_m \mid x_1 = s_m) \end{aligned} \tag{2.6}$$

where it can be seen that the second term on the right is just $p(e|A_2)$. For the first term, reference to decision rule $D_1$ indicates that

$$\begin{aligned} p(\hat{x}_1 = 0 \mid x_1 = s_m) &= p\left( |(y_1, s_i)|^2 < \beta_1^2 E_1^2 \text{ for all } i \;\middle|\; x_1 = s_m \right) \\ &\leq p\left( |(y_1, s_m)|^2 < \beta_1^2 E_1^2 \;\middle|\; x_1 = s_m \right) \\ &= p\left( |\alpha_1 E_1 e^{i\theta_1} + (n_1, s_m)|^2 < \beta_1^2 E_1^2 \right) \end{aligned}$$

where the last equality follows from the fact that $y_1 = \alpha_1 x_1 e^{i\theta_1} + n_1$. In the complex plane that represents the domain of the random variable $(n_1, s_m)$, the region $|\alpha_1 E_1 e^{i\theta_1} + (n_1, s_m)|^2 < \beta_1^2 E_1^2$ is a disc of radius $\beta_1 E_1$ centered at $-\alpha_1 E_1 e^{i\theta_1}$. See figure 2. Since $(n_1, s_m)$ is a zero mean complex Gaussian random variable with variance $N_1 E_1$, the probability that it falls within this disc is

$$\int \frac{1}{\pi N_1 E_1} e^{-|z|^2/N_1 E_1} \, dV(z)$$

$$< \pi\beta_1^2 E_1^2 \cdot \frac{1}{\pi N_1 E_1} e^{-(\alpha_1 E_1 - \beta_1 E_1)^2 / N_1 E_1}$$

$$= \beta_1^2 C_1 T \ e^{-TC_1(\alpha_1 - \beta_1)^2}$$

where the integral is taken over the disc.

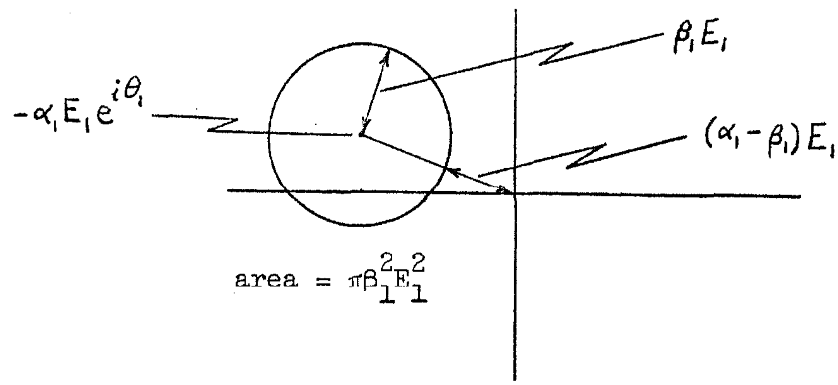


Figure 2: Disc of integration

Substituting this result together with (2.4) into (2.6) yields

$$p(e|A_4) < \beta_1^2 C_1 T \ e^{-TC_1(\alpha_1-\beta_1)^2} + e^{-T\alpha_1^2 C_1 E(R/\alpha_1^2 C_1)} \quad . \qquad (2.7)$$

Calculation of the probabilities $p(A_n)$ for $n = 1, 2, 3,$ and 4 follows next. The trivial bound

$$p(A_1) \leq 1 \qquad (2.8)$$

is chosen for $A_1$ since this event occurs most of the time if communication is reasonably reliable in both forward and feedback directions. For event $A_2$,

$$
\begin{aligned}
p(A_2) &= p(\ \hat{w}_1 \neq \sigma_m \text{ but } \hat{s}_o = s_m\ ) \\
&= p(\ \hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m\ )\ p(\ \hat{s}_o = s_m\ ) \\
&\leq p(\ \hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m).
\end{aligned}
$$

According to decision rule D ,

$$
\begin{aligned}
p(\ \hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m\ ) &= p\Big(\ |(z_1, \sigma_j)|^2 + \gamma^2 E_2^2 > |(z_1, \sigma_m)|^2 \text{ for some } j \neq m \ \Big|\ w_1 = \sigma_m \Big) \\
&= p\Big(\ |(\eta_1, \sigma_j)|^2 + \gamma^2 E_2^2 > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_m)|^2 \text{ for some } j \neq m \Big) \\
&= EX\Big\{ p\Big(\ |(\eta_1, \sigma_j)|^2 > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_m)|^2 - \gamma^2 E_2^2 \\
&\qquad \text{for some } j \neq m \ \Big|\ (\eta_1, \sigma_m)\ ,\ \varphi_1 \Big) \Big\} \qquad (2.9a)
\end{aligned}
$$

where EX stands for the expectation operator, taken here over the random variables $(\eta_1, \sigma_m)$ and $\varphi_1$ . By the union bound the quantity in brackets is less than

$$
M_1 p\Big(\ |(\eta_1, \sigma_j)|^2 > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_m)|^2 - \gamma^2 E_2^2 \ \Big|\ (\eta_1, \sigma_m)\ ,\ \varphi_1 \Big)
$$

and, since $|(\eta_1 \sigma_j)|^2$ has the Exponential probability density with mean $N_2 E_2$ , this in turn equals

$$M_1 \exp\left\{ - \frac{1}{N_2E_2} \cdot \left[ |E_2 e^{i\varphi_1} + (\eta_1, \sigma_m)|^2 - \gamma^2 E_2^2 \right] \right\} .$$

But the quantity in brackets, in (2.9a), being a probability, is also less than 1, and hence less than

$$\min \left\{ M_1 \exp\left( - \frac{1}{N_2E_2} \cdot \left[ |E_2 e^{i\varphi_1} + (\eta_1, \sigma_m)|^2 - \gamma^2 E_2^2 \right] \right), 1 \right\} .$$

Now for any positive numbers $X, Y$, if $Y < \min\{X, 1\}$ then $Y < X^\rho$ where $\rho$ can be any number in $[0,1]$ . Making the obvious choice for $X$, it follows that

$$p(A_2) < EX \left\{ M_1^\rho \exp\left( - \frac{\rho}{N_2E_2} \left[ |E_2 e^{i\varphi_1} + (\eta_1, \sigma_m)|^2 - \gamma^2 E_2^2 \right] \right) \right\}$$
$$= M_1^\rho e^{\rho\gamma^2 C_2 T} \cdot \frac{1}{1+\rho} e^{-\rho C_2 T/1+\rho} , \qquad 0 \le \rho \le 1$$

where $C_2 \triangleq \dfrac{P_2/3}{N_2}$ . The expectation has been performed by first taking it over $(\eta_1, \sigma_m)$ whose density is zero mean complex Gaussian with variance $N_2E_2$ and noting that the result is independent of $\varphi_1$ . Since the bound is valid for any $0 \le \rho \le 1$ it follows that

$$p(A_2) < \exp\left( -TC_2 \cdot \max_{0\le\rho\le1} \left[ \frac{\rho}{1+\rho} - \rho(\gamma^2 + R/C_2) \right] \right) .$$

It can be shown [9] that

$$\max_{0\le\rho\le1} \left[ \frac{\rho}{1+\rho} - \rho x \right] = E(x) .$$

Hence

$$p(A_2) < e^{-TC_2E(\gamma^2 + R/C_2)} \quad . \tag{2.9}$$

As mentioned before, event $A_3$ represents the case when an error occurs on the initial forward transmission but the transmitter fails to detect it on the subsequent feedback transmission. It is not difficult to see that the overall error probability must be at least as great as the probability of this event. That it need not be any greater is a fact which will be demonstrated after driving an upper bound for $p(e)$ .

$$\begin{aligned} p(A_3) &= p(\ \hat{w}_1=\sigma_m \text{ but } \hat{s}_o \neq s_m\ ) \\ &= p(\ \hat{w}_1=\sigma_m \mid w_1 \neq \sigma_m\ )\ p(\ \hat{s}_o \neq s_m) \end{aligned} \tag{2.10}$$

The second factor on the right is just the probability of error in communicating one of $M_1 = e^{RT}$ orthogonal signals over an incoherent channel with capacity $C_1$ . As previously mentioned this is bounded above by $\exp[-TC_1E(R/C_1)]$ . It can also be shown, incidentally, [9] that it is bounded below by $\exp[-T\{C_1E(R/C_1) + o(1)\}]$ . For the first factor, reference to decision rule D indicates that

$$\begin{aligned} p(\ \hat{w}_1=\sigma_m \mid w_1 \neq \sigma_m\ ) &= p(\ \hat{w}_1=\sigma_m \mid w_1=\sigma_k\ ,\ \text{say}\ ) \\ &= p\Big(\ |(z_1,\sigma_m)|^2 > |(z_1,\sigma_j)|^2 + \gamma^2E_2^2 \ \text{ for all } \\ &\qquad j \neq m \ \Big|\ w_1=\sigma_k\Big) \end{aligned}$$

$$= p\Big( |(\eta_1, \sigma_m)|^2 > |(\eta_1, \sigma_j)|^2 + \gamma^2 E_2^2 \quad \text{for all} \quad j \neq m, k$$

$$\text{and} \quad |(\eta_1, \sigma_m)|^2 > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_k)|^2 + \gamma^2 E_2^2 \Big) .$$

Since $|(\eta_1, \sigma_m)|^2$ has the Exponential probability density with mean $N_2 E_2$, conditioning on it yields

$$p(\hat{w}_1 = \sigma_m \mid w_1 = \sigma_k) = \int_0^\infty p\Big( x > |(\eta_1, \sigma_j)|^2 + \gamma^2 E_2^2 \quad \text{for all } j \neq m, k$$

$$\text{and } x > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_k)|^2 + \gamma^2 E_2^2 \ \Big|$$

$$|(\eta_1, \sigma_m)|^2 = x \Big) \cdot \frac{1}{N_2 E_2} e^{-x/N_2 E_2} \, dx .$$

The conditioning $|(\eta_1, \sigma_m)|^2 = x$ may be dropped since the random variables $(\eta_1, \sigma_j)$, $j = 1, \ldots, M_1$, are statistically independent. Inspection of the integrand reveals that it is zero for $x < \gamma^2 E_2^2$, so the lower limit of integration may just as well be $\gamma^2 E_2^2$. Making the substitution $y = x - \gamma^2 E_2^2$ yields

$$e^{-\gamma^2 E_2^2 / N_2 E_2} \int_0^\infty p\Big( y > |(\eta_1, \sigma_j)|^2 \quad \text{for all} \quad j \neq m, k \quad \text{and}$$

$$y > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_k)|^2 \Big) \cdot \frac{1}{N_2 E_2} e^{-y/N_2 E_2} \, dy .$$

The integral factor may be written as

$$p\Big( |(\eta_1, \sigma_m)|^2 > |(\eta_1, \sigma_j)|^2 \quad \text{for all} \quad j \neq m, k$$

$$\text{and} \quad |(\eta_1, \sigma_m)|^2 > |E_2 e^{i\varphi_1} + (\eta_1, \sigma_k)|^2 \Big) ,$$

which is just the probability of making the particular error $\sigma_m$ in communicating one of $M_1$ orthogonal signals over a $C_2$ capacity incoherent channel terminating in a type $D_o$ decision rule. Since the probability of making some error in such a scheme is bounded above by $\exp[-TC_2E(R/C_2)]$, (and below by $\exp\left(-T[C_2E(R/C_2) + o(1)]\right)$), the probability of making any particular error is $1/M_1$ - th of this (actually $\frac{1}{M_1 - 1}$), or $\exp\left(-T[C_2E(R/C_2) + R]\right)$. Thus

$$p(\hat{w}_1 = \sigma_m \mid w_1 \neq \sigma_m) < e^{-\gamma^2 C_2 T} \cdot e^{-T[C_2E(R/C_2) + R]} . \quad (2.11a)$$

Substituting this result into (2.10) yields

$$p(A_3) < e^{-T[C_1E(R/C_1) + C_2E(R/C_2) + R + \gamma^2 C_2]} . \quad (2.11b)$$

From the lower bounds mentioned above it also follows that

$$p(A_3) > e^{-T[C_1E(R/C_1) + C_2E(R/C_2) + R + \gamma^2 C_2 + o(1)]} . \quad (2.11c)$$

Finally,

$$\begin{aligned} p(A_4) &= p(\hat{s}_o \neq s_m \text{ and } \hat{w}_1 \neq \sigma_m) \\ &\leq p(\hat{s}_o \neq s_m) \\ &< e^{-TC_1E(R/C_1)} . \end{aligned} \quad (2.12)$$

This completes the calculations necessary for using formula (2.1) in obtaining an upper bound for $p(e)$. The result is

$$p(e) < e^{-T[\beta_1^2 C_1 - R]} + e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + C_2 E(\gamma^2 + R/C_2)]}$$
$$+ e^{-T[C_1 E(R/C_1) + C_2 E(R/C_2) + R + \gamma^2 C_2]}$$
$$+ \beta_1^2 C_1 T \, e^{-T[C_1(\alpha_1 - \beta_1)^2 + C_1 E(R/C_1)]}$$
$$+ e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + C_1 E(R/C_1)]} \quad . \quad (2.13)$$

## 2.3 Specification of System Parameters

A lower bound to the probability of error is readily obtained from the results of the previous section. Since $p(e) \geq p(e|A_3)\, p(A_3) \cong p(A_3)$, it follows from (2.11c) that

$$p(e) > e^{-T[C_1 E(R/C_1) + C_2 E(R/C_2) + R + \gamma^2 C_2 + o(1)]} \quad . \quad (2.14)$$

To show that $p(e)$ is also bounded above by the same expression one need only observe how $\alpha_1$ and $\beta_1$ control the bound in (2.13). It is clear that the third term on the right can be made the dominating component by choosing these parameters sufficiently large. To subordinate the first term it is only necessary that

$$\beta_1^2 > E(R/C_1) + \frac{C_2}{C_1} E(R/C_2) + 2\frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \quad .$$

For the second and last terms it suffices to set

$$\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) > C_1 E(R/C_1) + C_2 E(R/C_2) + R + \gamma^2 C_2 \quad .$$

But

$$\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) \geq \frac{\alpha_1^2 C_1}{2} - R \quad .$$

Thus choose

$$\frac{\alpha_1^2}{2} - \frac{R}{C_1} > E(R/C_1) + \frac{C_2}{C_1} E(R/C_2) + \frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \quad .$$

This will be satisfied if $\alpha_1^2 > 2\beta_1^2$ . For the fourth term, on the other hand, it is necessary that

$$(\alpha_1 - \beta_1)^2 > \frac{C_2}{C_1} E(R/C_2) + \frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \quad .$$

Letting $\alpha_1 = \Lambda_1 \beta_1$ yields

$$(\Lambda_1 - 1)^2 \beta_1^2 > \frac{C_2}{C_1} E(R/C_2) + \frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \quad ,$$

but since

$$\beta_1^2 > E(R/C_1) + \frac{C_2}{C_1} E(R/C_2) + 2 \frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \quad ,$$

it is enough to choose

$$(\Lambda_1 - 1)^2 \; [E \quad C_1) + \frac{C_2}{C_1} E(R/C_2) \; + 2 \frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \; ]$$

$$\frac{C_2}{C_1} E(R/C_2) + \frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1} \quad ,$$

or,

$$\Lambda_1 \; > \; 1 + \sqrt{\frac{C_2 E(R/C_2) + R + \gamma^2 C_2}{C_1 E(R/C_1) + C_2 E(R/C_2) + 2R + \gamma^2 C_2}} \quad .$$

Since $\Lambda_1$ must also be greater than $\sqrt{2}$, $(\alpha_1^2 > 2\beta_1^2)$, it follows that sufficient conditions for eliminating all but the third term in (2.13) are

$$\left.\begin{aligned} \beta_1 \; &> \; \sqrt{E(R/C_1) + \frac{C_2}{C_1} E(R/C_2) + 2\frac{R}{C_1} + \gamma^2 \frac{C_2}{C_1}} \\ &\text{and} \\ \alpha_1 \; &= \; \Lambda_1 \beta_1 \quad , \end{aligned}\right\} \quad (2.15)$$

where

$$\Lambda_1 \; > \; \max \left\{ \sqrt{2} \; , \; 1 + \sqrt{\frac{C_2 E(R/C_2) + R + \gamma^2 C_2}{C_1 E(R/C_1) + C_2 E(R/C_2) + 2R + \gamma^2 C_2}} \right\} .$$

Thus $p(e)$ can be bounded above by

$$p(e) \; < \; A \, e^{-T[C_1 E(R/C_1) + C_2 E(R/C_2) + R + \gamma^2 C_2]} \quad , \qquad (2.16)$$

where $A$ is greater than 1 but decreases to 1 exponentially in $T$.

The selection of $\gamma^2$ is restricted by the requirement that the peak power $\alpha_1^2 P_1$ be used infrequently. If $p(re - tr)$ denotes the probability that a re-transmission is required, then

$$\begin{aligned} p(re - tr) &= p(\hat{w}_1 \neq \sigma_m) \\ &= p(\hat{w}_1 \neq \sigma_m , \hat{s}_o = s_m) + p(\hat{w}_1 \neq \sigma_m , \hat{s}_o \neq s_m) \\ &= p(A_2) + p(A_4) \\ &< e^{-TC_2E(\gamma^2 + R/C_2)} + e^{-TC_1E(R/C_1)} \quad , \end{aligned}$$

where the bound follows from (2.9) and (2.12). Thus $p(re - tr)$ can be made arbitrarily small by increasing the coding delay, $T$, only if $\gamma^2 + R/C_2 < 1$ and $R < C_1$. The first condition will automatically be met by introducing a positive number $\nu < 1$ and setting

$$\gamma^2 + R/C_2 = \nu \ .$$

This, of course, will require that $R < \nu C_2$.

## 2.4 Limiting Performance

In section 1.4 it was pointed out that the average power, $\bar{P}_1$, required in the forward direction is given by

$$\bar{P}_1 = \frac{P_1}{3} + \alpha_1^2 \frac{P_1}{3} p(re - tr) \quad .$$

Clearly, the approximation $\bar{P}_1 \cong P_1/3$ follows if $R < \min[\nu C_2, C_1]$ and $T$ is large enough. Under these circumstances $\bar{C}_1 = C_1$, where $\bar{C}_1 \triangleq \bar{P}_1/N_1$ is the capacity of the forward channel. Also $\bar{C}_2 = C_2$ since $\bar{P}_2 = P_2/3$. In terms of $\bar{C}_1, \bar{C}_2$, and $\nu$, then, the performance that can be obtained with full information feedback and only a single use of the feedback channel is given by

$$\boxed{e^{-T[E^*(R) + o(1)]} < p(e) \le e^{-T[E^*(R) + o(1)]}} \tag{2.17}$$

and

$$p(re - tr) < e^{-T\bar{C}_2 E(\nu)} + e^{-T\bar{C}_1 E(R/\bar{C}_1)} \tag{2.18}$$

where

$$\boxed{E^*(R) = \bar{C}_1 E(R/\bar{C}_1) + \bar{C}_2 E(R/\bar{C}_2) + \nu\bar{C}_2} \tag{2.19}$$

provided that $R < \min[\nu\bar{C}_2 , \bar{C}_1]$ .

A plot of $E^*(R)$ versus $R$ when $\bar{C}_2 = \bar{C}_1$ and $\nu = 1$ is shown

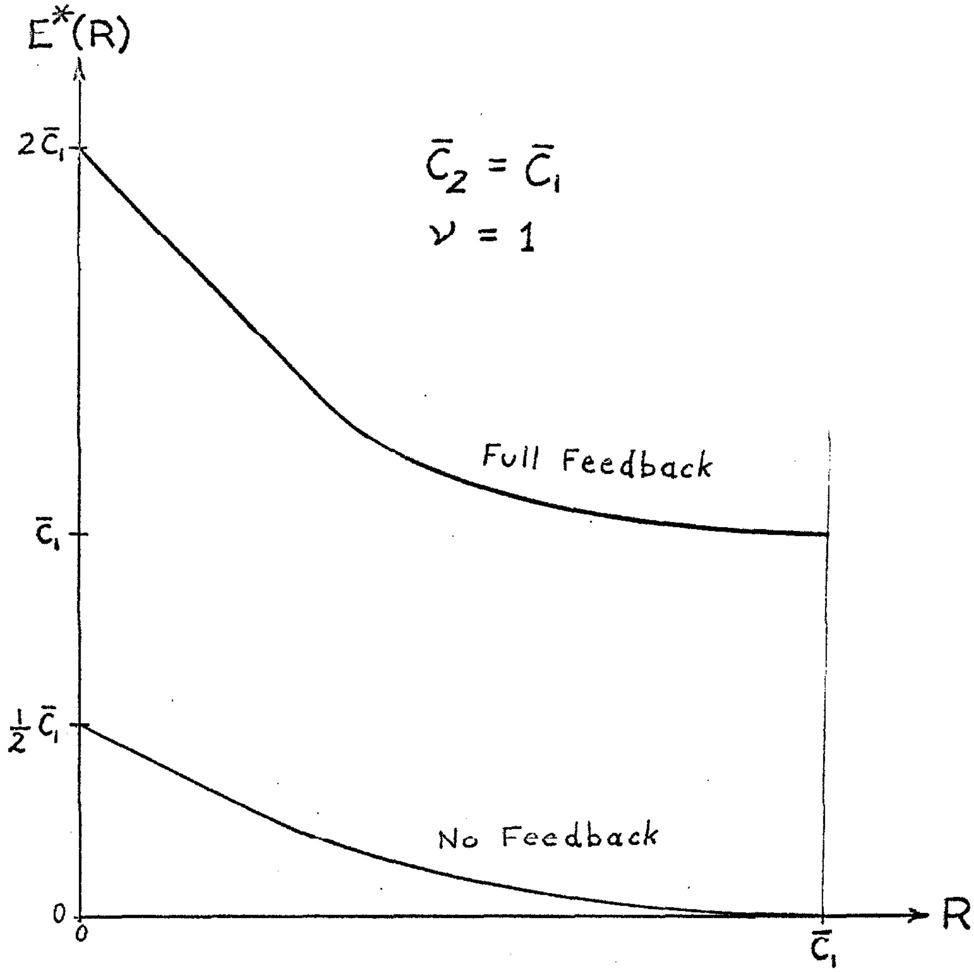


Figure 3 : Full Information Feedback Performance

in figure 3. As far as the probability of error is concerned the best performance is obtained if $\nu = 1$ . However, because $E(1) = 0$ the bound for $p(re - tr)$ degenerates at $\nu = 1$ . If $\nu < 1$ then $E(\nu) > 0$ and increasing $T$ reduces $p(re - tr)$ to any desired value. Thus while the limiting case $\nu = 1$ is not allowed, any value arbitrarily close to 1 is permissible.

## 2.5 Extension To N Feedback Transmissions

It will now be shown that for given average forward and feedback powers the performance of the scheme with $N > 1$ is not better than with $N = 1$ .

### (a) The scheme of figure 1

By analogy with the method in section 2.2 the probability of error for arbitrary $N$ can be found by writing

$$p_N(e) \stackrel{\Delta}{=} p(\hat{s}_N \neq s_m)$$

$$= \sum_{n=1}^{4} p(e|A_n^{(N)})\, p(A_n^{(N)})$$

where

$A_1^{(N)}$ denotes the event that $\hat{s}_{N-1} = s_m$ and $\hat{w}_N = \sigma_m$ ,

$A_2^{(N)}$ " " " " $\hat{s}_{N-1} = s_m$ but $\hat{w}_N \neq \sigma_m$ ,

$A_3^{(N)}$ denotes the event that $\hat{s}_{N-1} \neq s_m$ but $\hat{w}_N = \sigma_m$ ,

and $A_4^{(N)}$ " " " " $\hat{s}_{N-1} \neq s_m$ and $\hat{w}_N \neq \sigma_m$ ,

$\sigma_m$ being the vector in $\mathcal{T}$ which corresponds to the vector $s_m$ in $\mathcal{S}$ .

Event $A_3^{(N)}$ represents the worst possible situation. The receiver has made an error, $\hat{s}_{N-1} \neq s_m$ , but the transmitter fails to detect it, $\hat{w}_N = \sigma_m$ . Thus, since the transmitter will not repeat the message, the receiver will persist in its error. In other words, $p(e|A_3^{(N)}) \simeq 1$ . It follows that $p_N(e) \gtrsim p(A_3^{(N)})$ . But

$$p(A_3^{(N)}) = p(\hat{w}_N = \sigma_m , \hat{s}_{N-1} \neq s_m)$$

$$= p(\hat{w}_N = \sigma_m \mid w_N \neq \sigma_m)\, p(\hat{s}_{N-1} \neq s_m) \quad .$$

A probability similar to the first factor on the right has already been calculated in section 2.2. There it was shown that when $N = 1$ ,

$$p(\hat{w}_1 = \sigma_m \mid w_1 \neq \sigma_m) > e^{-T[C_2 E(R/C_2) + R + \gamma^2 C_2 + o(1)]}$$

(see the paragraph preceeding (2.11a) ). It should be recalled that $C_2 = \frac{E_2/T}{N_2}$ where $E_2$ is the energy in the feedback signal. When $N = 1$ this energy is derived from expending $P_2$ watts for $T/3$ seconds, hence the definition $C_2 = \frac{P_2/3}{N_2}$ . When $N > 1$ , however, only $T/2N+1$ seconds are available for each feedback transmission. Thus

$E_2 = P_2T/2N + 1$ and $C_2 = \frac{P_2/2N + 1}{N_2}$ . Since $p(\hat{w}_N = \sigma_m \mid w_N \neq \sigma_m)$ depends on $N$ only through $E_2$ it follows that by defining $C_2$ in this way,

$$p(\hat{w}_N = \sigma_m \mid w_N \neq \sigma_m) > e^{-T[C_2E(R/C_2) + R + \gamma^2 C_2 + o(1)]}$$

Thus

$$p(\hat{s}_N = s_m) > p(\hat{s}_{N-1} \neq s_m)\, e^{-T[C_2E(R/C_2) + R + \gamma^2 C_2 + o(1)]}$$

Now $p(\hat{s}_o \neq s_m)$ is the probability of error in communicating one of $M_1 = e^{RT}$ orthogonal signals with energy $E_1 = P_1T/2N + 1$ over the forward channel. Letting $C_1 = \frac{P_1/2N + 1}{N_1}$ it follows [9] that

$$p(\hat{s}_o \neq s_m) > e^{-T[C_1E(R/C_1) + o(1)]}$$

By induction on $N$ it can thus be concluded that

$$p_N(e) > e^{-T[C_1E(R/C_1) + NC_2E(R/C_2) + NC_2(\gamma^2 + R/C_2) + o(1)]}$$

But in order to keep the probability of an $N$ - th re-transmission small, that is $p(x_N = s_m) \ll 1$ , $\gamma^2 + R/C_2$ must be kept less than 1. The reason for this is that

$$p(x_N = s_m) = p(\hat{w}_N \neq \sigma_m)$$

$$\leq \; p( \hat{w}_N \neq \sigma_m \mid w_N = \sigma_m ) + p( \hat{s}_{N-1} \neq s_m )$$

$$< \; e^{-TC_2E(\gamma^2 + R/C_2)} + p( \hat{s}_{N-1} \neq s_m )$$

where the last inequality follows from (2.9) with $C_2$ redefined as in this section. Thus

$$p_N(e) \;>\; e^{-T[C_1E(R/C_1) + NC_2E(R/C_2) + NC_2 + o(1)]} \quad .$$

Before a comparison with $p_1(e)$ can be made it remains to express this bound in terms of $\bar{C}_1 \triangleq \bar{P}_1/N_1$ and $\bar{C}_2 \triangleq \bar{P}_2/N_2$, where $\bar{P}_1$ and $\bar{P}_2$ are the average forward and feedback power expenditures required. Well, since with high probability no power is used in the forward direction after the initial transmission, $\bar{P}_1 \cong P_1/2N + 1$ . In the feedback direction $P_2$ watts are expended on each of the N feedback transmissions. Consequently, $\bar{P}_2 = NP_2/2N + 1$ . Therefore $\bar{C}_1 = C_1$ , $\bar{C}_2 = NC_2$ , and

$$p_N(e) \;>\; e^{-T[\bar{C}_1E(R/\bar{C}_1) + \bar{C}_2E(NR/\bar{C}_2) + \bar{C}_2 + o(1)]} \quad .$$

This can now be compared to the limiting performance that can be approached when $N = 1$ , ( (2.17) and (2.19) with $\nu = 1$ ) ,

$$p_1(e) \;\leq\; e^{-T[\bar{C}_1E(R/\bar{C}_1) + \bar{C}_2E(R/\bar{C}_2) + \bar{C}_2 + o(1)]} \quad .$$

Since $E(R/\bar{C}_2) > E(NR/\bar{C}_2)$ it follows that $p_1(e)$ is exponentially smaller than $p_N(e)$ .

## (b) A modified scheme with improved $p_N(e)$

With the scheme in figure 1 the receiver always feeds back its current estimate, $\hat{s}_n$ . This, of course, is why it uses power $P_2$ watts N times. Suppose, instead, that the scheme is modified as follows:

Change G so that the receiver transmits its current estimate only if it differs from the previous estimate. That is,

$$G : \quad \text{send } w_1 = \sigma_i , \quad \text{where } \hat{s}_o = s_i ,$$

but for $n = 2,3,\ldots,N$ send

$$w_n = \begin{cases} 0 & \text{if } \hat{s}_{n-1} = \hat{s}_{n-2} \\ \lambda_n \sigma_j & \text{if } \hat{s}_{n-1} \neq \hat{s}_{n-2} \end{cases} ,$$

where $\hat{s}_{n-1} = s_j$ .

Without going into the details of what other modifications must be made in order to accommodate this one, the event which determines a lower bound to the probability of error can immediately be determined. For suppose that an error occurs on the initial forward transmission and on the first feedback transmission in such a way that the transmitter fails to detect the former error. That is, suppose event $A_3^{(1)}$ occurs. In this case the transmitter will send $x_1 = 0$ . The receiver will then imagine its initial estimate to be correct and the error will persist. According to G the second feedback transmission will be $w_2 = 0$ so that from here on 0 will be pointlessly transmitted back and forth N times. The initial error will never be corrected. Thus $p_N(e|A_3^{(1)}) \simeq 1$ and $p_N(e) \gtrsim p(A_3^{(1)})$ . But just as in subsection 2.5 (a) ,

$$p(A_3^{(1)}) = p( \hat{w}_1 = \sigma_m \mid w_1 \neq \sigma_m ) \, p( \hat{s}_o \neq s_m )$$

$$> e^{-T[C_1E(R/C_1) + C_2E(R/C_2) + C_2 + o(1)]} ,$$

where $C_1 = \frac{P_1/2N + 1}{N_1}$ and $C_2 = \frac{P_2/2N + 1}{N_2}$ . Also, as before, $\bar{C}_1 \simeq C_1$ . $\bar{C}_2$ , however, is approximately $C_2$ in this case, rather than $NC_2$ , because it is very likely that no power will be transmitted on all feedback iterations after the first. Therefore,

$$p_N(e) > e^{-T[\bar{C}_1E(R/\bar{C}_1) + \bar{C}_2E(R/\bar{C}_2) + \bar{C}_2 + o(1)]} .$$

It is possible, at the expense of laborious calculations, to show that $p_N(e)$ is also bounded above by the same expression. This is simply a consequence of the fact that all other events that contribute to $p_N(e)$ have probabilities which are monotonically decreasing functions of $\alpha_n$ and $\lambda_n$ , $n = 1,\ldots,N$ . Consequently these terms can be subordinated to $p(A_3^{(1)})$ by choosing the parameters sufficiently large.

In any case it can be seen, at least, that this scheme has the potential for better performance than the scheme of figure 1. The reason, apparently, is that it is better to expend all the available feedback power on the first iteration than to distribute it over many iterations. This supports the idea that a single iteration scheme is best of all.

Comparison of $p_N(e)$ with $p_1(e)$ shows that the modified scheme can, at best, only match the performance possible with the single feedback scheme.

# CHAPTER III. PARTIAL INFORMATION FEEDBACK

## 3.1 Introduction

Equation (2.19), it will later be shown, indicates an advantage of communication with full information feedback over communication without feedback. As long as $\bar{C}_2 \geq \bar{C}_1$ this superiority prevails at all rates below the forward channel capacity. However, if $\bar{C}_2 < \bar{C}_1$, then for rates above the feedback channel capacity, (and below the forward), full information feedback is inferior to no feedback (see the Appendix). But no matter what quality feedback channel is available, so long as it can carry some reliable information, it seems natural to expect that its use should improve communication over the forward channel. In this chapter it is shown how partial information feedback bears out this expectation. Fortunately, many useful calculations have already been performed in the previous chapter, so the analysis here will be fairly straightforward.

It should be recalled that with partial information feedback the receiver communicates only which of the $M_2 = e^{aRT}$ subsets of $\mathcal{S}$ $\hat{s}_o$ belongs to, rather than $\hat{s}_o$ itself. Thus the feedback channel need only handle a rate of $aR$ nats/sec, rather than $R$. However, this procedure gives rise to an event that is impossible with full information feedback, yet which is the determining factor in the probability of error for partial feedback. This event is denoted by $B_3$, below.

3.2 Calculation of p(e)

As in section 2.1 let $s_m$ be the vector chosen by the transmitter for communication. Let $Q_{m'}$ be the subset of $\mathcal{Q}$ to which $s_m$ belongs, and define

$B_1$, the event that $\hat{s}_o = s_m$ and $\hat{w}_1 = \sigma_{m'}$ ,

$B_2$, " " " $\hat{s}_o = s_m$ and $\hat{w}_1 \neq \sigma_{m'}$ ,

$B_3$, " " " $\hat{s}_o \neq s_m$ , $w_1 = \sigma_{m'}$ , and $\hat{w}_1 = \sigma_{m'}$ ,

$B_4$, " " " $\hat{s}_o \neq s_m$ , $w_1 \neq \sigma_{m'}$ , and $\hat{w}_1 = \sigma_{m'}$ ,

$B_5$, " " " $\hat{s}_o \neq s_m$ and $\hat{w}_1 \neq \sigma_{m'}$ .

It is readily verified that these events are mutually exclusive, their union exhausting the space of possible outcomes. Thus p(e) can be expressed as

$$p(e) = \sum_{n=1}^{5} p(e|B_n)\, p(B_n) \quad . \tag{3.1}$$

The considerations leading to a bound for $p(e|B_1)\, p(B_1)$ are identical to those used for $p(e|A_1)\, p(A_1)$ in section 2.2. In fact,

$$p(e|B_1)\, p(B_1) \leq p(e|B_1)$$

$$= p(\hat{x}_1 \neq 0 \text{ or } s_m \mid x_1 = 0)$$

$$= p(e|A_1)$$

$$< e^{-T[\beta_1^2 C_1 - R]} , \qquad (3.2)$$

where the last inequality follows that (2.3). Similarly,

$$p(e|B_2) = p(\hat{x}_1 \neq 0 \text{ or } s_m \mid x_1 = s_m)$$

$$= p(e|A_2)$$

$$< e^{-T\alpha_1^2 C_1 E(R/\alpha_1^2 C_1)} .$$

On the other hand, $p(B_2)$ is not quite $p(A_2)$ .

$$p(B_2) = p(\hat{s}_o = s_m , \hat{w}_1 \neq \sigma_{m'} )$$

$$\leq p(\hat{w}_1 \neq \sigma_{m'} \mid \hat{s}_o = s_m)$$

$$= p(\hat{w}_1 \neq \sigma_{m'} \mid w_1 = \sigma_{m'}) ,$$

which leads to the same bound given by (2.9) except for $aR$ replacing $R$ . Thus

$$p(B_2) < e^{-TC_2 E(\gamma^2 + aR/C_2)} ,$$

and

$$p(e|B_2)\,p(B_2) < e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + C_2 E(\gamma^2 + aR/C_2)]} \quad . (3.3)$$

Now

$$p(B_3) \cong p(\,\hat{s}_o \neq s_{m'}\,,\; w_1 = \sigma_{m'}\,)$$

$$= p(\,w_1 = \sigma_{m'} \mid \hat{s}_o \neq s_m\,)\; p(\,\hat{s}_o \neq s_m\,)$$

$$< p(\,w_1 = \sigma_{m'} \mid \hat{s}_o \neq s_m\,)\; e^{-TC_1 E(R/C_1)} \quad .$$

But $p(\,w_1 = \sigma_{m'} \mid \hat{s}_o \neq s_m\,)$ is just the probability that of the $M_1 - 1$ equiprobable errors, $\hat{s}_o = s_j$, $j = 1, \ldots, m-1, m+1, \ldots, M_1$, the receiver chooses one belonging to the subset $Q_{m'}$. Since there are $M_1/M_2 - 1$* possible candidates in this set it follows that

$$p(\,w_1 = \sigma_{m'} \mid \hat{s}_o \neq s_m\,) = \frac{1}{M_1 - 1} \cdot \left(\frac{M_1}{M_2} - 1\right)$$

$$\leq 1/M_2$$

$$= e^{-aRT} \quad .$$

*Each subset has $M_1/M_2$ elements, but the $-1$ is necessary because the particular subset $Q_{m'}$ contains the vector $s_m$.

Thus

$$p(e|B_3)\ p(B_3) \ < \ e^{-T[C_1E(R/C_1) + aR]} \ . \qquad (3.4)$$

Event $B_4$ is similar to event $A_3$ :

$$p(B_4) = p(\ \hat{s}_o \neq s_m\ ,\ \hat{w}_1 \neq \sigma_{m'}\ ,\ \hat{w}_1 = \sigma_{m'}\ )$$

$$\leq p(\ \hat{w}_1 = \sigma_{m'}\ |\ w_1 \neq \sigma_{m'}\ )\ p(\ \hat{s}_o \neq s_m\ )$$

$$< p(\ \hat{w}_1 = \sigma_{m'}\ |\ w_1 \neq \sigma_{m'}\ )\ e^{-TC_1E(R/C_1)} \ .$$

A bound for $p(\ \hat{w}_1 = \sigma_{m'}\ |\ w_1 \neq \sigma_{m'}\ )$ can be obtained by observing that if the calculations leading to (2.11a) were carried out with aR replacing R , as is appropriate for partial feedback, the result would be

$$p(\ \hat{w}_1 = \sigma_{m'}\ |\ w_1 \neq \sigma_{m'}\ ) \ < \ e^{-T[C_2E(aR/C_2) + aR + \gamma^2C_2]} \ .$$

Thus

$$p(e|B_4)\ p(B_4) \ < \ e^{-T[C_1E(R/C_1) + C_2E(aR/C_2) + aR + \gamma^2C_2]} \ . \qquad (3.5)$$

Finally, $p(e|B_5)\ p(B_5)$ can be bounded by the same expression used for $p(e|A_4)\ p(A_4)$ as given by (2.7) and (2.12), namely,

$$p(e|B_5)\,p(B_5) \;<\; \beta_1^2 C_1 T \;\; e^{-T[C_1(\alpha_1-\beta_1)^2 + C_1E(R/C_1)]} + e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + C_1E(R/C_1)]} \quad . \tag{3.6}$$

Substituting the results of (3.2) - (3.6) into (3.1) yields

$$\begin{aligned} p(e) \;<\; & e^{-T[\beta_1^2 C_1 - R]} + e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + C_2E(\gamma^2 + aR/C_2)]} \\ & + e^{-T[C_1E(R/C_1) + aR]} + e^{-T[C_1E(R/C_1) + C_2E(aR/C_2) + aR + \gamma^2 C_2]} \\ & + \beta_1^2 C_1 T \; e^{-T[C_1(\alpha_1-\beta_1)^2 + C_1E(R/C_1)]} \\ & + e^{-T[C_1E(R/C_1) + \alpha_1^2 E(R/\alpha_1^2 C_1)]} \quad . \end{aligned} \tag{3.7}$$

## 3.3 Selection of Parameters

In (3.7) the third term can be made the dominant component by choosing $\alpha_1$ and $\beta_1$ sufficiently large. The derivation of sufficient conditions is quite analagous to the method in section 2.3. The result is,

$$\beta_1 \;>\; \sqrt{E(R/C_1) + (1+a)\,R/C_1} \tag{3.8a}$$

and

$$\alpha_1 \;=\; \Lambda_1 \beta_1 \;\; , \tag{3.8b}$$

where

$$\Lambda_1 > \max\left\{\sqrt{2}\ ,\quad 1 + \sqrt{\frac{aR/C_1}{E(R/C_1) + (1+a)\ R/C_1}}\right\} . \qquad (3.8c)$$

Thus,

$$p(e) \leq e^{-T[C_1E(R/C_1) + aR + o(1)]} . \qquad (3.9)$$

Selection of $a$ is constrained by the requirement that $p(\text{re-tr})$ be very small.

$$p(\text{re-tr}) = p(\hat{w}_1 \neq \sigma_{m'})$$

$$\leq p(\hat{w}_1 \neq \sigma_{m'}, \hat{s}_o = s_m) + p(\hat{s}_o \neq s_m)$$

$$= p(B_2) + p(\hat{s}_o \neq s_m)$$

$$< e^{-TC_2E(\gamma^2 + aR/C_2)} + e^{-TC_1E(R/C_1)} .$$

Thus it is necessary that $\gamma^2 + aR/C_2 < 1$ and $R < C_1$. In the first condition the choice $\gamma^2 = 0$ may as well be made since the $p(e)$ performance is independent of $\gamma$. Then to satisfy $aR/C_2 < 1$, introduce a positive number $\nu < 1$ and let

$$a = \min[1\ ,\ \nu C_2/R] . \qquad (3.10)$$

Note that this restricts $R$ to be greater than $\nu C_2$, for otherwise $a = 1$ and full information feedback becomes the mode of operation.

Assuming that $p(\text{re-tr})$ is sufficiently small so that $\bar{P}_1 \cong P_1/3$, and therefore $\bar{C}_1 \cong C_1$, it follows that

$$\boxed{p(e) \leq e^{-T[\bar{C}_1 E(R/C_1) + \nu\bar{C}_2 + o(1)]}} , \tag{3.11}$$

and

$$p(\text{re-tr}) < e^{-T\bar{C}_2 E(\nu)} + e^{-T\bar{C}_1 E(R/\bar{C}_1)} , \tag{3.12}$$

provided that $\nu C_2 < R < C_1$ and $0 < \nu < 1$.

That $p(e)$ is also bounded below by the right side of (3.11) can be seen as follows:

$$p(e) \geq p(e|B_3)\, p(B_3)$$

$$\gtrsim p(B_3) .$$

But
$$p(B_3) = p(\, w_1 = \sigma_{m'} \mid \hat{s}_o \neq s_m \,)\, p(\, \hat{s}_o \neq s_m \,)$$

$$> \left[\frac{1}{M_2} - \frac{1}{M_1}\right] \cdot e^{-T[C_1 E(R/C_1) + o(1)]}$$

$$> e^{-T[C_1 E(R/C_1) + aR + o(1)]} .$$

Since $\bar{C}_1 \cong C_1$ and $a = \min[1\ ,\ \nu C_2/R]$ , it follows that

$$p(e) \ > \ e^{-T[\bar{C}_1 E(R/\bar{C}_1) + \nu\bar{C}_2 + o(1)]} \tag{3.13}$$

if $\nu\bar{C}_2 < R < \bar{C}_1$ .

## 3.4 Overall System Performance

When $\nu\bar{C}_2 \geq \bar{C}_1$ the system is in the full feedback mode so the performance is given, for all $0 < R < \bar{C}_1$ , by (2.17) - (2.19) . When $\nu\bar{C}_2 < \bar{C}_1$ the performance is that of full feedback for $0 < R \leq \nu\bar{C}_2$ , and partial feedback for $\nu\bar{C}_2 < R < \bar{C}_1$ . By combining (2.17) - (2.19) with (3.11) - (3.13) the overall system performance can be written as,

$$\boxed{e^{-T[E^*(R) + o(1)]} \ < \ p(e) \ \leq \ e^{-T[E^*(R) + o(1)]}} \ ,$$

and

$$p(\text{re-tr}) \ < \ e^{-T\bar{C}_2 E(\nu)} + e^{-T\bar{C}_1 E(R/\bar{C}_1)} \ ,$$

provided $0 < R < \bar{C}_1$ and $0 < \nu < 1$ , where

$$\boxed{E^*(R) = \begin{cases} \bar{C}_1 E(R/\bar{C}_1) + \bar{C}_2 E(R/\bar{C}_2) + \nu\bar{C}_2 \ , & 0 \leq R \leq \min[\nu\bar{C}_2, \bar{C}_1] \\ \bar{C}_1 E(R/\bar{C}_1) + \nu\bar{C}_2 \ , & \min[\nu\bar{C}_2, \bar{C}_1] < R \leq \bar{C}_1 \end{cases}} \tag{3.14}$$

# CHAPTER IV. COMPARISON WITH EXISTING SCHEMES

## 4.1 Introduction

At present there exist two other feedback communication schemes which tolerate noise in the feedback channel, one due to Kramer [7], the other due to Lavenberg [6]. Kramer uses block coded orthogonal signals and a matched filter receiver in a scheme quite similar to the one presented in this paper. The receiver always sends back its current estimate of the message. The transmitter sends the actual message only on the initial transmission. Thereafter it transmits only the difference between the message and its estimate of the receiver's decision. Lavenberg has modified this scheme by employing a similar differencing technique on the feedback channel. That is, rather than always sending back its current estimate, the receiver sends only the difference between its last two estimates. The added complexity, it turns out, is rewarded by a greater immunity to feedback noise. In the following sections the scheme of this paper will be compared to each of these schemes. It should be pointed out, however, that the comparison is only valid for the phase-coherent channel.

## 4.2 Comparison with No Feedback

When block coded orthogonal signals are used to communicate over the band-infinite phase-incoherent channel without the use of feedback, the probability of error is given by

$$e^{-T[E_{NF}^{*}(R) + o(1)]} < p(e) \le e^{-T[E_{NF}^{*}(R) + o(1)]} ,$$

provided $0 < R < \bar{C}_1$ , where

$$E^*_{NF}(R) = \bar{C}_1 E(R/\bar{C}_1) \quad , \quad 0 \le R \le \bar{C}_1 \quad .$$

That this is the best possible performance that can be obtained without feedback would follow if the orthogonal signal set were optimum. For two complex dimensions this has recently been proved by Schaffner [11], but for an arbitrary dimensionality it remains an unproved conjecture.

Assuming the conjecture is valid, comparison of $E^*_{NF}(R)$ with $E^*(R)$ , (3.14), shows that the use of feedback substantially improves on the best performance possible without feedback. The improvement is particularly striking when $R$ is close to $\bar{C}_1$ . Here $E^*_{NF}(R)$ is close to zero, indicating the necessity for large $T$ to obtain small $p(e)$ . $E^*(R)$ , however, remains bounded away from zero as long as $\bar{C}_2 > 0$ .

## 4.3 Comparison with Kramer's Scheme

Kramer [7] showed that the performance of his scheme with $N$ feedback transmissions is given, for $0 < R < \bar{C}_1$ , by

$$E^*_K(R) = (N + 1)\, \bar{C}_1 E[R/(N+1)\bar{C}_1]$$

provided that $\bar{C}_2 \ge N(N + 1)\, \bar{C}_1$ , where

$$E^*_K(R) \triangleq \lim \left[- \frac{1}{T} \ln p(e)\right] \quad .$$

This can be compared to $E^*(R)$ . If $\bar{C}_2 \geq N(N+1)\bar{C}_1$ then

$$E^*(R) \geq \bar{C}_1 E(R/\bar{C}_1) + N(N+1)\bar{C}_1 E[R/N(N+1)\bar{C}_1] + \nu N(N+1)\bar{C}_1 .$$

Since the most striking comparison of $E^*(R)$ with $E^*_{NF}(R)$ occurs at $R = \bar{C}_1$, it is interesting to compare $E^*(R = \bar{C}_1)$ with $E^*_K(R = \bar{C}_1)$ for varying $\bar{C}_2$ . The curved portion of $E^*_K(R = \bar{C}_1)$ versus $\bar{C}_2$, in figure 4, is only a rough guess to the actual curve, made with the help of Lavenberg's upper bound to $E^*_K(R)$ , [6], in the range $N\bar{C}_1 \leq \bar{C}_2 \leq N(N+1)\bar{C}_1$ . For the purpose of comparison, however, it is clear that an exact knowledge of the curve is not important. Note that $E^*_K(R = \bar{C}_1) = 0$ when $\bar{C}_2 < N\bar{C}_1$ .

## 4.4 Comparison with Lavenberg's Scheme

With $N$ feedback transmissions the performance of Lavenberg's scheme for $0 < R < \bar{C}_1$ is given by

$$E^*_L(R) = \begin{cases} \bar{C}_1 E(R/\bar{C}_1 + \bar{C}_2 E(R/\bar{C}_2) & , \bar{C}_1 \leq \bar{C}_2 \leq N\bar{C}_1 \\ (N+1)\bar{C}_1 E[R/(N+1)\bar{C}_1] & , \bar{C}_2 \geq (N+1)\bar{C}_1 \end{cases}$$

where

$$E^*_L(R) \triangleq \lim [-\frac{1}{T} \ln p(e)] .$$

It can be seen that $E^*(R) > E^*_L(R)$ for all $0 < R < \bar{C}_1$ and for any $\bar{C}_2$ . Figure 4 shows $E^*_L(R = \bar{C}_1)$ versus $\bar{C}_2$ . Note that the curve vanishes when $\bar{C}_2 \leq \bar{C}_1$ .

Note also that while $E_L^*(R = \bar{C}_1)$ and $E_K^*(R = \bar{C}_1)$ saturate at $(N + 1)\bar{C}_1 E(1/N + 1)$ for large $\bar{C}_2$ , $E^*(R = \bar{C}_1)$ continues to grow. This is somewhat misleading. The peak-to-average power ratio in the schemes of Kramer and Lavenberg is $2(2N + 1)$ , independent of $\bar{C}_2$ . For the scheme in this paper, however, the ratio is $3\alpha_1^2$ , an increasing function of $\bar{C}_2$ , ( see (2.15) ) . If $3\alpha_1^2$ is constrained to be equal to $2(2N + 1)$ , so that the schemes have equal peak-to-average power ratios, it turns out that $E^*(R = \bar{C}_1)$ also saturates at large values of $\bar{C}_2$ . This behavior is depicted by the dashed curve in figure 4. Note that the saturation value is actually below that for the other schemes. It must be remembered, however, that since the present scheme uses only one feedback transmission its bandwidth requirement is less, by the factor $(2N + 1)/3$ , than that required by the schemes of Kramer and Lavenberg.

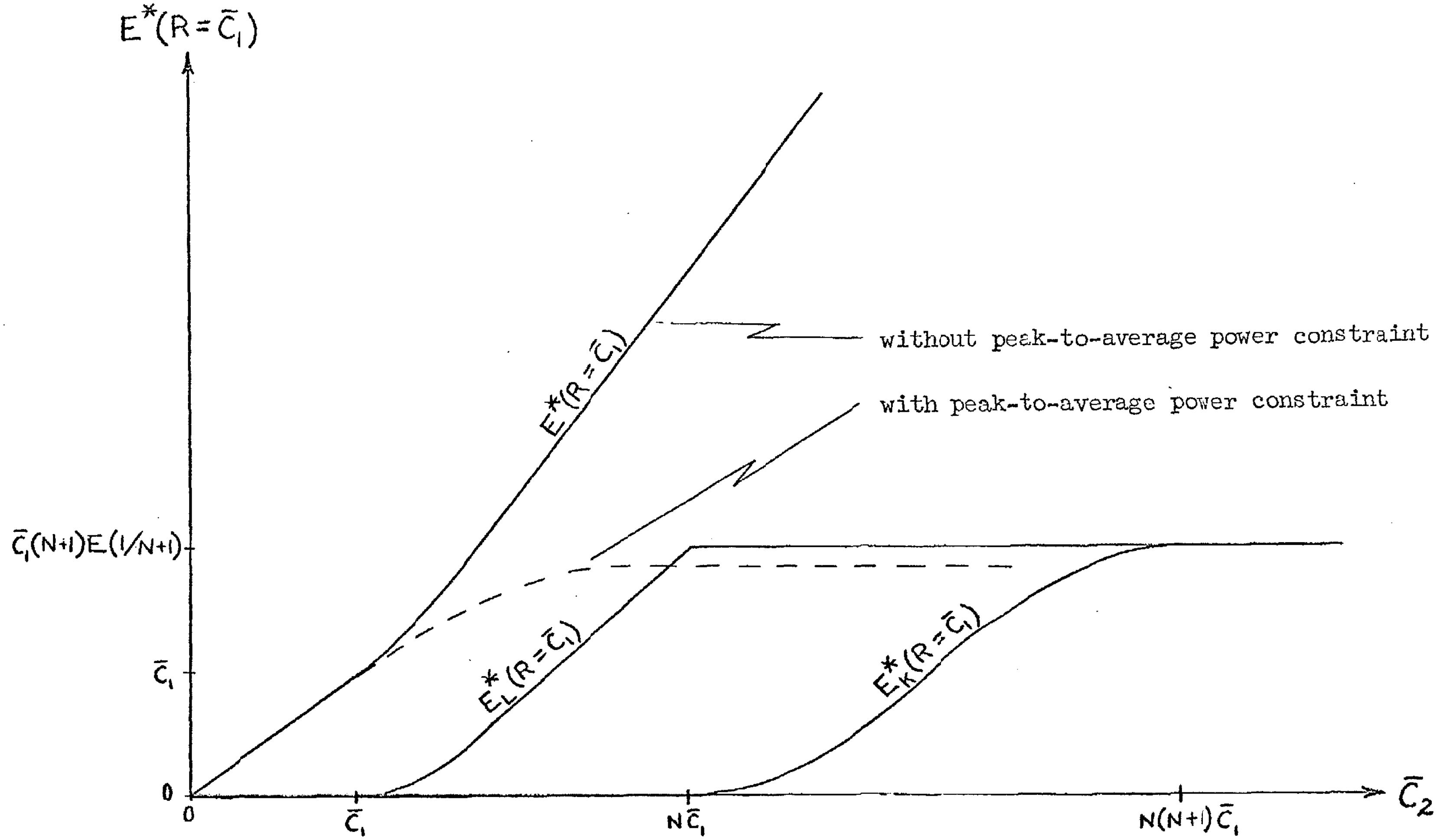


Figure 4 : Comparison of Schemes

## CONCLUSIONS

The dramatic effectiveness of feedback communication, up to now applied only to phase-coherent channels, is now extended to include phase-incoherent systems as well. Since such systems are less expensive to build this constitutes an important contribution to communication technology, especially in the field of relay communication where the maintenance of phase-coherency requires special equipment at each relay station.

The practical scheme presented in this paper was found to yield better performance than what was possible with previous schemes. The advantage lies not only in improved reliability for any given forward and feedback average power constraints, but in the ability to tolerate any finite amount of noise in the feedback channel. Of course, increasing the noise progressively deteriorates the performance, but, unlike the previous schemes, complete degeneracy never occurs. This is important in designing systems that carry on two-way communication. Here each channel must transmit information of its own in addition to any feedback information for the other channel. Since the available power on each channel must be shared between these two, it is certainly advantageous, if not essential, that the amount of power required for feedback purposes be small.

An interesting question arises in this context. Perhaps the essential advantages of feedback can be obtained even with asymptotically zero feedback power. If this were true, it would mean that both channels in a two-way communication system could operate arbitrarily close to capacity, and still gain the more important advantages of feedback. The basis for this conjecture lies in the observation that while the rate of feedback transmission is R bits/sec. (for full feedback, say), the average information communicated to the transmitter is very small. With high probability the receiver always sends back a correct initial estimate. Since the average feedback information is small, it would seem that the required feedback channel capacity should also be small.

There is one more feature of the scheme presented in this paper that is worthy of notice. It was found that increasing the number of feedback transmissions serves only to increase the required bandwidth, without any corresponding improvement in reliability. Although it is strongly suspected, it is not known for sure whether this phenomenon can also be expected when the ultimate performance of feedback communication is determined.

# APPENDIX

## DEGENERACY OF FULL INFORMATION FEEDBACK AT RATES GREATER THAN THE FEEDBACK CHANNEL CAPACITY

Here it is shown that when $\bar{C}_2 < R < \bar{C}_1$, communication with full information feedback is inferior to communication with no feedback. Only the case $N = 1$ is considered:

$$p(e) = \sum_{n=1}^{4} p(e|A_n)\, p(A_n)$$

$$\geq p(e|A_4)\, p(A_4)$$

$$= p(\hat{s}_1 \neq s_m \mid \hat{s}_o \neq s_m, \hat{w}_1 \neq \sigma_m)\, p(\hat{s}_o \neq s_m, \hat{w}_1 \neq \sigma_m)$$

$$= p(\hat{x}_1 \neq s_m \mid x_1 = s_m)\, p(\hat{s}_o \neq s_m, \hat{w}_1 \neq \sigma_m)$$

$$= p(\hat{x}_1 \neq s_m \mid x_1 = s_m)\, p(\hat{w}_1 \neq \sigma_m \mid w_1 \neq s_m)\, p(\hat{s}_o \neq s_m)\,. \tag{A.1}$$

$p(\hat{x}_1 \neq s_m \mid x_1 = s_m)$ :

According to decision rule $D_1$ it is clear that if $|(y_1, s_i)|^2 > |(y_1, s_m)|^2$ for some $i \neq m$, then $\hat{x}_1 \neq s_m$. Thus

$$p(\hat{x}_1 \neq s_m \mid x_1 = s_m) \geq p\Big( |(y_1, s_i)|^2 > |(y_1, s_m)|^2 \text{ for some } i \neq m \;\Big|\; x_1 = s_m \Big)\,.$$

But the right hand side may be regarded [9] as the probability of error in communicating one of $M_1 = e^{RT}$ messages over the phase-incoherent channel using orthogonal signals and corresponding optimum receiver. This is bounded below by $\exp\{-T[CE(R/C) + o(1)]\}$, where $C$ is the capacity of the channel. In the situation of interest $C = \alpha_1^2 C_1$. Thus

$$p(\hat{x}_1 \neq s_m \mid x_1 = s_m) > e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + o(1)]} \quad . \qquad (A.2)$$

<u>$p(\hat{w}_1 \neq \sigma_m \mid w_1 \neq \sigma_m)$</u> :

$$p(\hat{w}_1 \neq \sigma_m \mid w_1 \neq \sigma_m) = 1 - p(\hat{w}_1 = \sigma_m \mid w_1 \neq \sigma_m)$$

$$\geq 1 - e^{-T[C_2 E(R/C_2) + R + \gamma^2 C_2]}$$

where the bound follows from (2.11a). Thus

$$p(\hat{w}_1 \neq \sigma_m \mid w_1 \neq \sigma_m) \cong 1 \quad . \qquad (A.3)$$

<u>$p(\hat{s}_o \neq s_m)$</u> :

The same argument as for $p(w_1 \neq s_m \mid x_1 = s_m)$ applies here too, except that $C = C_1$ :

$$p(\hat{s}_o \neq s_m) > e^{-T[C_1 E(R/C_1) + o(1)]} \quad . \qquad (A.4)$$

Substituting (A.2) - (A.4) into (A.1) yields

$$p(e) \; > \; e^{-T[\alpha_1^2 C_1 E(R/\alpha_1^2 C_1) + C_1 E(R/C_1) + o(1)]} \quad . \qquad (A.5)$$

It will now be shown that for large $T$

$$\bar{C}_1 = (1 + \alpha_1^2)\, C_1$$

This will follow from

$$\bar{P}_1 = \frac{P_1}{3} + \frac{\alpha_1^2 P_1}{3}\, p(\text{re-tr})$$

if it can be shown that $p(\text{re-tr}) \to 1$ as $T \to \infty$ .

In section 2.3 it was shown that

$$p(\text{re-tr}) \; = \; p(A_2) + p(A_4) \quad .$$

But

$$p(A_2) \; = \; p(\, \hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m \,)\; p(\, \hat{s}_o = s_m \,)$$

$$= \; p(\, \hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m \,)\; [1 - p(\, \hat{s}_o \neq s_m \,)]$$

$$> \; p(\, \hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m \,)\; [1 - e^{-TC_1 E(R/C_1)}] \quad .$$

Now

$$p(\hat{w}_1 \neq \sigma_m \mid w_1 = \sigma_m) = p\Big( |(z_1,\sigma_j)|^2 + \gamma^2 E_2^2 > |(z_1,\sigma_m)|^2 \text{ for some } j \neq m \;\Big|\; w_1 = \sigma_m \Big)$$

$$\geq p\Big( |(z_1,\sigma_j)|^2 > |(z_1,\sigma_m)|^2 \text{ for some } j \neq m \;\Big|\; w_1 = \sigma_m \Big)$$

which is the probability of error in communication over a $C_2$ - capacity phase-incoherent channel using block orthogonal signaling and corresponding optimum decision rule. Turin [12] has shown that when $R > C_2$ this approaches 1 as $T \to \infty$ . Thus if $\bar{C}_2 < R < C_1$ ,

$$p(A_2) \to 1 \text{ as } T \to \infty .$$

Also

$$p(A_4) = p(\hat{w}_1 \neq \sigma_m , \; \hat{s}_o \neq s_m)$$

$$\cong p(\hat{s}_o \neq s_m)$$

which approaches 1 as $T \to \infty$ if $R > C_1$ . Thus

$$p(\text{re-tr}) \to 1 \text{ as } T \to \infty \text{ if } R > \bar{C}_2 ,$$

and it follows that

$$\bar{P}_1 \to (1 + \alpha_1^2) \frac{P_1}{3} \quad \text{as} \quad T \to \infty \quad .$$

Thus making the substitution $\bar{C}_1 = (1 + \alpha_1^2) C_1$ in (A.5) yields

$$p(e) > \exp\left\{-T\left[\frac{\alpha_1^2}{1+\alpha_1^2} \bar{C}_1 E[R(1+\alpha_1^2)/\alpha_1^2\bar{C}_1] + \frac{1}{1+\alpha_1^2} \bar{C}_1 E[R(1+\alpha_1^2)/\bar{C}_1] + o(1)\right]\right\}. \tag{A.6}$$

But $\frac{1+\alpha_1^2}{\alpha_1^2} > 1$ and $1+\alpha_1^2 > 1$ for any $\alpha_1^2$ . Thus

$$E[R(1+\alpha_1^2)/\alpha_1^2\bar{C}_1] < E(R/\bar{C}_1)$$

and

$$E[R(1+\alpha_1^2)/\bar{C}_1] < E(R/\bar{C}_1) \quad .$$

It follows that

$$p(e) > e^{-T[\bar{C}_1 E(R/\bar{C}_1) + o(1)]} \quad .$$

On the other hand, the probability of error obtainable with the same average power expenditure, but without the use of feedback, is

$$p_{NF}(e) \le e^{-T\bar{C}_1 E(R/\bar{C}_1)} \quad .$$

## REFERENCES

[1] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel", I.R.E. Trans. on Information Theory, Vol. IT-2, pp. 8-19, September 1956.

[2] S. Butman, "Optimum Linear Coding for Additive Noise Systems Using Information Feedback", Tech. Report No. 1, Communications Theory Lab., C.I.T., May 1967. Forthcoming I.E.E.E. publication.

[3] J. P. M. Schalkwijk and T. Kailath, "A Coding Scheme for Additive Noise Channels with Feedback - Part I: No Bandwidth Constraint", I.E.E.E. Trans. on Information Theory, Vol. IT-12, pp. 172-182, April 1966.

[4] J. P. M. Schalkwijk, "A Coding Scheme for Additive Noise Channels with Feedback - Part II: Band-Limited Signals", I.E.E.E. Trans. on Information Theory, Vol. IT-12, pp. 183-189, April 1966.

[5] J. K. Omura, "Signal Optimization for Channels with Feedback", Report SEL-66-068, Stanford Electronics Labs., Stanford University, August 1966.

[6] S. S. Lavenberg, "Feedback Communication Using Orthogonal Signals", Tech. Report No. 2, Communications Theory Lab., C.I.T., December 1967.

[7] A. J. Kramer, "Analysis of Communication Schemes Using an Intermittent Feedback Link", Report SEL-67-014, Stanford Electronics Labs., Stanford University, March 1967.

[8] J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, John Wiley and Sons Inc., New York, 1965.

[9] T. L. Grettenberg, "Exponential Error Bounds for Incoherent Orthogonal Signals", I.E.E.E. Trans. on Information Theory, Vol. IT-14, pp. 163-164, January 1968.

[10] T. L. Grettenberg, "A Representation Theorem for Complex Normal Processes", I.E.E.E. Trans. on Information Theory, Vol. IT-11, pp. 305-306, April 1965.

[11] C. A. Schaffner and H. A. Krieger, "The Global Optimization of Two and Three Phase-Incoherent Signals", Tech. Report No. 3, Communications Theory Lab., C.I.T., January 1968.

[12] G. L. Turin, "The Asymptotic Behavior of Ideal M-ary Systems", Proc. I.R.E., Vol. 47, pp. 93-94, January 1959.