On the Probability of Error Using a Repeat-Request Strategy on the Additive White Gaussian Noise Channel

By A. D. WYNER

(Manuscript received July 5, 1968)

An upper bound on the error probability is obtained for digital communication (with average power P_o and no bandwidth constraint) in the presence of additive white gaussian noise (with one-sided spectral density N_o) with the use of a noiseless feedback link. A repeat-request strategy is used: the receiver decodes a signal only when it is relatively sure that one particular message was actually transmitted, otherwise it requests (via the feedback channel) a retransmission. We show that as the coding delay T becomes large, we can transmit at an effective rate $\bar{R} < C = P_o/N_o$, the channel capacity, with error probability P_e approximately exp $\{-T[(\sqrt{C} - \sqrt{\bar{R}})^2 + C - \bar{R}]\}$, which is a considerable improvement over the reliability attainable with a one-way channel. These results parallel those obtained earlier by Forney for the discrete memoryless channel.

I. INTRODUCTION

In a recent paper, Forney studied a repeat-request strategy for communication of digital information over a discrete memoryless channel when a fedback channel is available.[3] In this system the receiver decodes a received message only when it is relatively "sure" that one particular message was actually transmitted. If the receiver is not confident that one particular message was actually transmitted, then it requests (via the feedback channel) that the transmitter repeat the message. Forney showed that considerable improvement in the resulting error probability (over the best one-way scheme) was obtainable with a negligible degradation in the effective rate of transmission. In this paper we apply Forney's ideas to the additive white Gaussian noise channel (with no bandwidth constraint) and obtain analogous results. Furthermore, our coding scheme is constructive—the codes being orthogonal codes.

We will consider the following channel. The channel input signal is a real-valued function $s(t)$, defined on the interval $[0, T]$, which satisfies the "energy" constraint

$$\int_0^T s^2(t)\, dt = P_0 T. \tag{1}$$

The average signal "power" is therefore P_o. The channel output $r(t)$ is the sum of $s(t)$ and a sample $n(t)$ from a white Gaussian noise process with one-sided spectral density N_o (and with mean zero). By expanding $s(t)$, $r(t)$ and $n(t)$ on any orthonormal basis of $\mathfrak{L}_2[0, T]$, it is easy to show that an equivalent channel model is as follows.[8,9] (This equivalent channel model is the one we use in this paper.) The input signals are are (semi-infinite) vectors $\mathbf{x} = (x_1, x_2, \cdots)$ which satisfy

$$\sum_{k=1}^{\infty} x_k^2 = AT. \tag{2}$$

The channel output is a vector $\mathbf{y} = (y_1, y_2, \cdots)$, where

$$y_k = x_k + z_k, \qquad k = 1, 2, \cdots,$$

and the $z_k (k = 1, 2, \cdots)$ are independent Gaussian variates with zero mean and unit variance. The parameter A is equal to $2P_o/N_o$, and we assume that A is held fixed throughout the paper. We also assume that it takes T seconds for the channel to process $\mathbf{x}$, and that successive T-second transmissions are independent.

A *code* with parameters M and T is a set of M signals (called "code vectors" or "code words") $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots)$, $i = 1, 2, \cdots, M$, which satisfy equation (2), that is

$$\sum_{k=1}^{\infty} x_{ik}^2 = AT, \qquad i = 1, 2, \cdots, M. \tag{3}$$

We assume that each of the M code words is equally likely to be transmitted, so that the *transmission rate* is $R = 1/T \ln M$ nats (natural units) per second, and $M = e^{RT}$. It is the task of the receiver to examine the channel output $\mathbf{y}$ and to announce the code word, say $D(\mathbf{y})$, which it believes was actually transmitted. Let P_{ei} be the probability that $D(\mathbf{y}) \neq \mathbf{x}_i$ given that $\mathbf{x}_i$ is transmitted. The overall error probability is therefore

$$P_e = \frac{1}{M} \sum_{i=1}^{M} P_{ei}.$$

It is easy to show that for a given code, the "optimal" decoding rule D

(which minimizes P_e) selects for $D(\mathbf{y})$ that code word $\mathbf{x}_i$ which maximizes (with respect to i) the inner product

$$\langle \mathbf{x}_i , \mathbf{y} \rangle = \sum_{k=1}^{\infty} x_{ik} y_k$$

Define $P_e^*(M, T)$ as the smallest attainable error probability P_e for a code with parameters M and T. Set $M = [e^{RT}]$, and let $T \to \infty$ with the rate R held fixed. Then it is well known that if $R < A/2 = P_o/N_o \triangleq C$, the "channel capacity,"

$$P_e^*([e^{RT}], T) = \exp \{-E_0(R)T[1 + \epsilon_0(T)]\}, \tag{4}$$

where $E_0(R) > 0$, and $\epsilon_0(T) \to 0$ as $T \to \infty$.[1,8,9] Thus at rates $R < C$, the error probability tends to zero exponentially in T. Further, for rates $R > C$, $P_e^*([e^{RT}], T) \to 1$, so that the capacity C is the supremum of the rates for which "error-free" coding is possible.

Although this type of behavior of P_e^* is typical of a large class of channels, the present channel is unique in two ways. First the exponent $E_0(R)$ is known exactly, namely

$$E_0(R) = \begin{cases} C/2 - R, & 0 \leqq R \leqq C/4, \\ [C^{\frac{1}{2}} - R^{\frac{1}{2}}]^2, & C/4 \leqq R \leqq C. \end{cases} \tag{5}$$

Second, an explicit construction of codes which achieve error probability as in equation (4) is known. In fact, P_e as in equations (4) and (5) can be achieved when the code is any set of M orthogonal vectors. The simplest such code is that for which x_{ik} (the kth coordinate of $\mathbf{x}_i$) is given by

$$x_{ik} = \begin{cases} (AT)^{\frac{1}{2}}, & k = i, \\ 0, & k \neq i, \end{cases} \qquad i = 1, 2, \cdots, M, \quad k = 1, 2, \cdots . \tag{6}$$

For this orthogonal code, the inner product of $\mathbf{y}$ and the ith code word is

$$\langle \mathbf{x}_i , \mathbf{y} \rangle = y_i (AT)^{\frac{1}{2}}, \qquad i = 1, 2, \cdots, M;$$

so that the optimal decoding rule is

$$D(\mathbf{y}) = \mathbf{x}_i \quad \text{if} \quad y_i > y_j \qquad \text{for all } j \neq i, \qquad 1 \leqq j \leqq M. \tag{7}$$

With probability one, (7) is satisfied for exactly one i. Notice that the coordinates $y_j (j > M)$ are irrelevant to the receiver. Further, from the symmetry of the orthogonal code (6), we can without loss of generality, assume that code word $\mathbf{x}_1$ is transmitted. Hence, the error probability is

$$P_e = P_{e1} = \Pr \bigcup_{j=2}^{M} \{y_1 \leqq y_j\}, \tag{8}$$

where the probability is computed with $\{y_i\}_1^M$ independent unit variance Gaussian random variables with $Ey_1 = (AT)^{\frac{1}{2}}$ and $Ey_j = 0$ $(2 \leqq j \leqq M)$.

Now suppose we can use a noiseless feedback link. As before, we transmit one of a set of $M = e^{RT}$ orthogonal signals $\{\mathbf{x}_i\}_1^M$, where $\mathbf{x}_i$ is given by (6). Instead of the decoding rule (7), let us use the rule

$$D(\mathbf{y}) = \mathbf{x}_i \text{ if } y_i > y_j + \Delta \quad \text{for all } j \neq i, \quad 1 \leqq j \leqq M, \tag{9}$$

where $\Delta > 0$ will be chosen later. If no y_i satisfies (9) then we request a retransmission via the feedback channel, and use (9) on the second received vector, and so on. The probability of error decreases as Δ increases. The price which we pay for this increased reliability is an increase in the length of time which it will take to complete the transmission of the M-ary message, and the consequential reduction in the effective rate of transmission. In fact, let E_R be the event that we ask for a retransmission, and let $P(E_R)$ be its probability. Then from the assumption that successive transmissions are independent, the expected number of T-second transmissions required to accept a message is

$$\sum_{j=1}^{\infty} j \Pr \{j \text{ transmissions are required}\}$$

$$= \sum_{j=1}^{\infty} j[1 - P(E_R)][P(E_R)]^{j-1} = [1 - P(E_R)] \sum_{j=1}^{\infty} jP(E_R)^{j-1}$$

$$= [1 - P(E_R)] \cdot \frac{1}{[1 - P(E_R)]^2} = \frac{1}{1 - P(E_R)}.$$

Thus the average length of time required to transmit the M-ary message is $\bar{T} = T/(1 - P(E_R))$. If $P(E_R)$ is small, then $\bar{T}$ is not much greater than T.

Suppose that we use this repeat-request strategy repeatedly—that is, if the receiver does not call for a retransmission, then the transmitter sends a new M-ary message. For $k = 1, 2, \cdots$, let the random variable N_k be the number of M-ary messages which the receiver accepts (that is, it does not call for a retransmission) in kT seconds. Then we can write

$$N_k = \sum_{j=1}^{k} \xi_j ,$$

where the random variables $\xi_j = 1$ if the receiver accepts a message on the jth T-second interval, and $\xi_j = 0$ otherwise. Note that $Pr\{\xi_j = 0\} = P(E_R)$, and that the $\{\xi_j\}_1{}^k$, are independent (since we have assumed that successive T-second transmissions are independent). Thus

$$
\begin{aligned}
&(i) \quad E(N_k) = kE(\xi_i) = k(1 - P(E_R)) \\
&(ii) \quad N_k/k \to 1 - P(E_R), \quad \text{as } k \to \infty, \qquad (10) \\
&\qquad \text{with probability 1.}
\end{aligned}
$$

Statement (*ii*) follows from the strong law of large numbers (see Ref. 3, p. 190). Since each M-ary message contains $\ln M = RT$ nats, the effective rate of transmission $\bar{R}$, in the light of (10),

$$
\begin{aligned}
\bar{R} &= \frac{[E(N_k)]RT}{kT} \text{ nats/sec} \\
&= R[1 - P(E_R)] = R(T/\bar{T}).
\end{aligned} \qquad (11)
$$

Let us turn our attention to the probability of error. Since we are using the orthogonal code of (6), we can, as above, without loss of generality, assume that code word $\mathbf{x}_1$ is transmitted. Using the decoding rule of equation (1.9) we make an error only when for some $j > 1$, $y_j > y_i + \Delta$ for *all* $i = 1, 2, \cdots, M$ and $i \neq j$. (In this case $D(\mathbf{y}) = \mathbf{x}_j$.) Thus the error probability is

$$P_e = \Pr \bigcup_{j=2}^{M} \bigcap_{i \neq j} \{y_j > y_i + \Delta\}. \qquad (12)$$

As in (8), the probability in equation (12) is computed with $Ey_1 = (AT)^{\frac{1}{2}}$ and $Ey_j = 0 \ (2 \leqq j \leqq M)$.

Let us further define E_1 as the event that either an error occurs or a repeat-request occurs. If $\mathbf{x}_1$ is transmitted, E_1 has probability

$$\Pr(E_1) = \Pr \bigcup_{j=2}^{M} \{y_1 \leqq y_j + \Delta\}, \qquad (13)$$

where as above, the probability in (13) is computed with $Ey_1 = (AT)^{\frac{1}{2}}$ and $Ey_j = 0$, $j > 1$. Clearly the probability of a repeat-request is

$$P(E_R) = P(E_1) - P_e \leqq P(E_1). \qquad (14)$$

Consider the parameter Δ. In the interest of minimizing P_e, we want to make Δ large. However, in the interest of minimizing $P(E_R)$ and therefore making $\bar{R}$ as close to R as possible, we want to make Δ small. The approach which we will take is to choose Δ just small enough so that as the parameter $T \to \infty$ (R is held fixed), $P(E_1) \to 0$; so that by (14), $P(E_R) \to 0$. Thus the effective transmission rate $\bar{R} \approx R$. We will see that this results in a considerable improvement in P_e over that of equations (4) and (5). Roughly speaking, we will show that the resulting exponent is increased from that in equation (5) to approximately

$$E_F(R) = [C^{\frac{1}{2}} - R^{\frac{1}{2}}]^2 + C - R = 2C^{\frac{1}{2}}(C^{\frac{1}{2}} - R^{\frac{1}{2}}). \tag{15}$$

The exponents $E_0(R)$ and $E_F(R)$ are plotted in Fig. 1. Notice that the improvement is greatest in the neighborhood of capacity where (as $R \to C)E_F(R) \approx (C - R)$ and $E_0(R) \sim (C - R)^2/4C$.

II. SUMMARY AND DISCUSSION OF RESULTS

The main result is given as a corollary to the following two theorems which provide information on the trade-off between P_e and $P(E_1)$ as Δ is varied. The proofs are given in Section III.

Theorem 1: Let $\{y_i\}_1^M$, *be independent Gaussian random variables with unit variance and expectation*

$$\begin{aligned} Ey_1 &= (AT)^{\frac{1}{2}}, \\ Ey_j &= 0, \qquad 2 \leqq j \leqq M. \end{aligned} \tag{16}$$

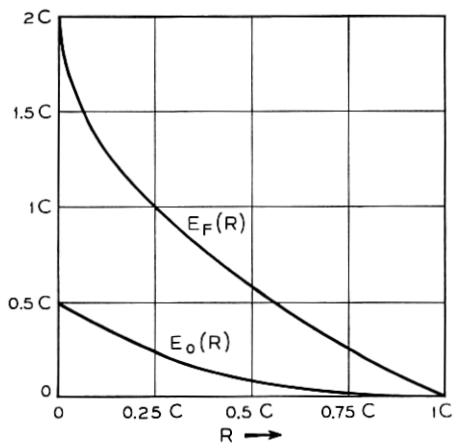

Fig. 1 — Exponents for white Gaussian noise channel: $E_0(R)$-one way exponent, $E_F(R)$-repeat-request exponent.

Let $M = e^{RT}$, where $0 < R < A/2 = C$, and let $\Delta = \delta(2T)^{\frac{1}{2}}$, where

$$C^{\frac{1}{2}} - (4R)^{\frac{1}{2}} \leqq \delta < C^{\frac{1}{2}} - R^{\frac{1}{2}}. \tag{17}$$

Then

$$P(E_1) = \Pr \bigcup_{j=2}^{M} \{y_1 \leqq y_j + \Delta\} \leqq 2 \exp \{-[C^{\frac{1}{2}} - R^{\frac{1}{2}} - \delta]^2 T\}. \tag{18}$$

Notice that $\delta = 0$ will satisfy (17) if $R \geqq C/4$. In this case $P(E_1) = P_e$ (see (8)), and (18) yields $E_0(R) \geqq [C^{\frac{1}{2}} - R^{\frac{1}{2}}]^2 (C/4 \leqq R \leqq C)$, a fact which is contained in (5). In fact, the proof of Theorem 1 closely parallels the derivation of P_e for orthogonal codes (for a one-way channel).

Theorem 2: Let $\{y_i\}_1^M$, be independent gaussian random variables with unit variance and expectation

$$Ey_1 = (AT)^{\frac{1}{2}}, \qquad Ey_j = 0, \quad 2 \leqq j \leqq M. \tag{19}$$

Let $M = e^{RT}$, where $0 \leqq R < A/2 = C$, and let $\Delta = \delta(2T)^{\frac{1}{2}}$, where

$$\delta > C^{\frac{1}{2}} - (4R)^{\frac{1}{2}}. \tag{20}$$

With R and δ held fixed, and θ_1, θ_2 arbitrary but satisfying

$$\theta_1 > 0, \tag{21a}$$

$$0 < \theta_2 < \begin{cases} R^{\frac{1}{2}} \\ \dfrac{\delta - [C^{\frac{1}{2}} - (4R)^{\frac{1}{2}}]}{2}, \end{cases} \tag{21b}$$

then for T sufficiently large,

$$\begin{aligned} P_e &= \Pr \bigcup_{j=2}^{M} \bigcap_{i \neq j} \{y_j > y_i + \Delta\} \\ &\leqq 2(1 + \theta_1) \exp \{-[(R^{\frac{1}{2}} + \delta - \theta_2)^2 + (C^{\frac{1}{2}} - R^{\frac{1}{2}} + \theta_2)^2 - R]T\}. \end{aligned} \tag{22}$$

Again notice that $\delta = 0$ will satisfy (20) if $R > C/4$. In this case also, (22) yields $E_0(R) \geqq [C^{\frac{1}{2}} - R^{\frac{1}{2}}]^2$, when $R > C/4$ (since θ_2 can be made arbitrarily small).

Let us now use these theorems to find the value of $\Delta = \delta(2T)^{\frac{1}{2}}$ which gives the smallest upper bound on P_e without substantially changing the effective rate $\bar{R} \geqq R[1 - P(E_1)]$. Since P_e is a decreasing function of δ, we choose δ as large as possible with the proviso that $P(E_1) \rightarrow 0$.

From Theorem 1, this value of δ is

$$\delta = C^{\frac{1}{2}} - R^{\frac{1}{2}} - \gamma_1 , \qquad (23)$$

where $\gamma_1 > 0$. If γ_1 is sufficiently small, this choice of δ satisfies (17) and (20). With δ so chosen, for any $\gamma_2 > 0$ we can find a T sufficiently large so that $\bar{R} \geqq R(1 - \gamma_2)$. Further, substitution of equation (23) into equation (22) yields an exponent

$$-[C^{\frac{1}{2}} - \gamma_1 - \theta_2)^2 + (C^{\frac{1}{2}} - R^{\frac{1}{2}} + \theta_2)^2 - R]T.$$

Finally, since γ_1, γ_2, θ_1 and θ_2 can be made arbitrarily small we have our main result:

Corollary: Let $\theta_1 > 0$, $\epsilon > 0$ be arbitrary. Let $\bar{R} < C$. Then for T sufficiently large, there is a repeat-request communication system using orthogonal codes with an effective rate of $\bar{R}$ and error probability

$$P_e \leqq 2(1 + \theta_1) \exp \{-[(C^{\frac{1}{2}} - \bar{R}^{\frac{1}{2}})^2 + C - \bar{R} - \epsilon]T\}.$$

Let us turn our attention to (4) and (5) which give the error probability for the one-way Gaussian channel. The fact that $E_0(R) \leqq (C^{\frac{1}{2}} - R^{\frac{1}{2}})^2$ can be demonstrated by a "sphere-packing" argument.[9] This argument states that $P_e^*(M, T) \geqq Q$, where Q is the probability of error which would result if it were possible to subdivide Euclidean M-space into M congruent cones (each with apex at the origin), one for each code word, and each code word were placed on the axis of its cone at a distance $(AT)^{\frac{1}{2}}$ from the origin. Setting the "sphere-packing exponent"

$$E_{SP}(R) = (C^{\frac{1}{2}} - R^{\frac{1}{2}})^2,$$

we have from the above corollary that for effective transmission rates $\bar{R} < C$ we can obtain an error exponent arbitrarily close to

$$E_F(\bar{R}) = E_{SP}(\bar{R}) + C - \bar{R}. \qquad (24)$$

For discrete memoryless channels it is possible to find a lower bound to the optimal (one-way) error probability using an analogous sphere-packing argument.[7] Forney showed that using a repeat-request strategy similar to the one used here, one can obtain an error exponent arbitrarily close to that of equation (24) [with the appropriate $E_{SP}(\bar{R})$].[3] Forney also studied the so called (discrete) "very noisy channel," which is closely related to our Gaussian channel* and obtained results similar

* Our Gaussian channel may be thought of as a "very noisy channel" since the signal-to-noise ratio per coordinate is zero.

to our results. Thus, in the light of Forney's results, the above corollary is not surprising.

Let us also remark that Kramer has found a scheme for our white noise channel with a feedback link that attains an error exponent of $C - R$, which is less than that in equation (24).[4] In Kramer's scheme, the receiver observes the signal until it is sufficiently confident that one particular message was actually transmitted. It then informs the transmitter, via the feedback channel, to start the next M-ary transmission, thereby using the feedback channel only once per M-ary message. In the repeat-request scheme studied here, the number of uses of the feedback channel per M-ary transmission is an unbounded random variable. Thus the two schemes, while similar (in that the feedback channel is used only to convey a "decision"), are not directly comparable. On the other hand, there are schemes which use the feedback channel considerably more heavily (so called "information feedback") which in some cases attain somewhat better performance than the repeat-request strategy. (See for example Refs. 5, 6, and 10).

Finally, an important problem which has been completely ignored here is the requirement that the transmitter have a buffer in which it can store data which will accumulate at the transmitter at times when the receiver asks for retransmissions. If the buffer has finite capacity, it will occasionally overflow, introducing a further source of errors. Some quantitative results on this problem have been obtained by the author, and will be reported in a future paper.

III. PROOFS OF THEOREMS

We begin with some definitions. Let

$$g(\alpha) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp(-\alpha^2/2), \quad -\infty < \alpha < \infty,$$

be the standard Gaussian density, and let

$$\Phi(u) = \int_{-\infty}^{u} g(\alpha)\, d\alpha, \qquad -\infty < u < \infty,$$

be the cumulative error function, and let

$$\Phi_c(u) = \int_{u}^{\infty} g(\alpha)\, d\alpha = 1 - \Phi(u), \qquad -\infty < u < \infty,$$

be the complementary error function. Let $b = (AT)^{\frac{1}{2}} = (2CT)^{\frac{1}{2}}$ so that y_1 has density $g(\alpha - b)$ and y_j $(2 \leqq j \leqq M)$ has density $g(\alpha)$. We will use the following

Lemma 1: For $u \geqq 0, \Phi_c(u) \leqq \exp(-u^2/2)$; *and for* $u \leqq 0, \Phi_c(u) \leqq \exp(-u^2/2)$ (Wozencraft and Jacobs Ref. 8):

Proof: For $u \geqq 0$,

$$[\Phi_c(u)]^2 = \int_u^\infty \int_u^\infty g(\alpha)g(\beta)\, d\alpha\, d\beta \leqq \int_{\mathcal{R}} \int g(\alpha)g(\beta)\, d\alpha\, d\beta = \frac{\exp(-u^2)}{4},$$

where $\mathcal{R} = \{(\alpha, \beta): \alpha^2 + \beta^2 \geqq 2u^2, \alpha \geqq 0, \beta \geqq 0\}$. Taking square roots, we have

$$\Phi_c(u) \leqq \frac{\exp(-u^2/2)}{2} \leqq \exp(-u^2/2).$$

The rest of Lemma 1 follows on noting that $\Phi(u) = \Phi_c(-u)$.

Proof of Theorem 1: Let R $(0 < R < C)$ and δ satisfying (17) be given. Since y_1 has density $g(\alpha - b)$, and the $\{y_i\}_1^M$ are independent,

$$\begin{aligned} P(E_1) &= \Pr \bigcup_{j=2}^{M} \{y_j \geqq y_1 - \Delta\} \\ &= \int_{-\infty}^{\infty} d\alpha\, g(\alpha - b) \Pr \left\{ \begin{matrix} \text{at least one} \\ y_j \geqq y_1 - \Delta \end{matrix} \,\middle|\, y_1 = \alpha \right\} \\ &= \int_{-\infty}^{\infty} d\alpha\, g(\alpha - b) \Pr \bigcup_{j=2}^{M} \{y_j \geqq \alpha - \Delta\}. \end{aligned} \tag{25}$$

Now since the y_j $(j > 1)$ have density $g(\alpha)$,

$$\Pr \bigcup_{j=2}^{M} \{y_j \geqq \alpha - \Delta\} \leqq \begin{cases} 1 \\ (M-1) \Pr \{y_j \geqq \alpha - \Delta\} \leqq M\Phi_c(\alpha - \Delta). \end{cases} \tag{26}$$

Letting a be a parameter to be specified later, we break the integral of equation (25) into two parts, $\alpha \leqq a$ and $\alpha \geqq a$. We then apply the first upper bound of (26) in the first part, and the second bound of (26) in the second part. Thus

$$P(E_1) \leqq \int_{-\infty}^{a} g(\alpha - b)\, d\alpha + M \int_a^\infty g(\alpha - b)\Phi_c(\alpha - \Delta)\, d\alpha.$$

If we assume that

$$a \geqq \Delta, \tag{27}$$

we can use the bound of Lemma 1 on $\Phi_c(\alpha - \Delta)$ and obtain

$$P(E_1) \leqq \int_{-\infty}^{a} g(\alpha - b)\, d\alpha + M \int_{a}^{\infty} g(\alpha - b) \exp[-(\alpha - \Delta)^2/2]\, d\alpha$$
$$= P_1 + MP_2. \tag{28}$$

We now overbound P_1 and P_2. First,

$$P_1 = \int_{-\infty}^{a} g(\alpha - b)\, d\alpha = \int_{-\infty}^{a-b} g(\alpha)\, d\alpha = \Phi(a - b).$$

If we further assume that

$$a \leqq b, \tag{29}$$

we can use Lemma 1 and obtain

$$P_1 \leqq \exp[-(b - a)^2/2]. \tag{30}$$

Second,

$$P_2 = \int_{a}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp[-\tfrac{1}{2}(\alpha - b)^2] \exp[-\tfrac{1}{2}(\alpha - \Delta)^2]\, d\alpha$$
$$= \int_{a}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\left(\alpha - \frac{b + \Delta}{2}\right)^2\right] \exp[-(b - \Delta)^2/4]\, d\alpha$$
$$= \frac{\exp[-(b - \Delta)^2/4]}{\sqrt{2}} \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\sqrt{2}[a-(b+\Delta)/2]}^{\infty} \exp(-v^2/2)\, dv$$
$$= \frac{\exp[-(b - \Delta)^2/4]}{\sqrt{2}} \Phi_c\left\{\sqrt{2}\left[a - \left(\frac{b + \Delta}{2}\right)\right]\right\}.$$

If we now make a third assumption that

$$a \geqq \frac{b + \Delta}{2}, \tag{31}$$

we can use Lemma 1 again (and $2^{-\frac{1}{2}} \leqq 1$) to bound P_2:

$$P_2 \leqq \exp[-(b - \Delta)^2/4] \exp\left\{-\left[a - \left(\frac{b + \Delta}{2}\right)\right]^2\right\}$$
$$= \exp[-(b - a)^2/2] \exp[-(a - \Delta)^2/2]. \tag{32}$$

Inserting the bounds on P_1 and P_2 into (28), we obtain

$$P(E_1) \leqq \exp[-(b - a)^2/2]\{1 + M \exp[-(a - \Delta)^2/2]\}, \tag{33a}$$

where from (27), (29), and (31),

$$\left.\begin{matrix} \Delta \\ \dfrac{b+\Delta}{2} \end{matrix}\right\} \leqq a \leqq b. \tag{33b}$$

It remains to choose the parameter a. A good choice will probably result when the upper bound of (28) is differentiated with respect to a and the result set equal to zero:

$$g(a - b) - Mg(a - b) \exp [-(a - \Delta)^2/2] = 0,$$

or

$$M \exp [-(a - \Delta)^2/2] = 1, \tag{34a}$$

or since $M = \exp (RT)$ and $\Delta = \delta(2T)^{\frac{1}{2}}$,

$$a = (R^{\frac{1}{2}} + \delta)(2T)^{\frac{1}{2}}. \tag{34b}$$

Let us now verify that when $0 < R < C$, constraints (33b) are satisfied for this choice of a. Since $R > 0$, $a \geqq \Delta$. Further, since $b = (2CT)^{\frac{1}{2}}$,

$$a - \left(\frac{b + \Delta}{2}\right) = \{\delta - [C^{\frac{1}{2}} - (4R)^{\frac{1}{2}}]\}\left[\frac{(2T)^{\frac{1}{2}}}{2}\right] \geqq 0,$$

since δ satisfies (17). Finally, from (17),

$$b - a = [C^{\frac{1}{2}} - (R^{\frac{1}{2}} + \delta)](2T)^{\frac{1}{2}} \geqq 0.$$

Thus constraints (33b) are, in fact, satisfied. Thus from (34) and (33a)

$$P(E_1) \leqq 2 \exp [-(C^{\frac{1}{2}} - R^{\frac{1}{2}} - \delta)^2 T],$$

which is Theorem 1.

Proof of Theorem 2: Let R $(0 \leqq R < C)$, $\delta > C^{\frac{1}{2}} - (4R)^{\frac{1}{2}}$, and θ_1, θ_2 satisfying equation (21) be given. Then

$$P_e = \Pr \bigcup_{j=2}^{M} \bigcap_{i \neq j} \{y_i < y_j - \Delta\} \leqq \sum_{j=2}^{M} \Pr \bigcap_{i \neq j} \{y_i < y_j - \Delta\},$$

or

$$P_e \leqq M \Pr \bigcap_{i \neq j} \{y_i < y_j - \Delta\}, \qquad j \geqq 2.$$

The last inequality follows from the symmetry of the distributions of the y_j $(j \geqq 2)$. Recalling that the density for y_j $(j \geqq 2)$ is $g(\alpha)$, and that the $\{y_i\}_1^M$ are independent,

$$= M \int_{-\infty}^{\infty} g(\alpha)\, d\alpha \Pr \left\{ \begin{matrix} \text{for all} \quad i \neq j \\ y_i < y_j - \Delta \end{matrix} \,\middle|\, y_j = \alpha \right\}$$

$$= M \int_{-\infty}^{\infty} g(\alpha)\, d\alpha \Pr \bigcap_{i \neq j} \{y_i < \alpha - \Delta\}.$$

Again using the independence of the y_i and the fact that the density of y_1 is $g(\alpha - b)$ we have

$$\Pr \bigcap_{i \neq j} \{y_i < \alpha - \Delta\} = \left[\int_{-\infty}^{\alpha-\Delta} g(\alpha - b)\, d\alpha\right]\left[\int_{-\infty}^{\alpha-\Delta} g(\alpha)\, d\alpha\right]^{M-2}$$
$$= \Phi(\alpha - \Delta - b)[\Phi(\alpha - \Delta)]^{M-2}.$$

Substituting, we obtain

$$P_e \leqq M \int_{-\infty}^{\infty} g(\alpha)\Phi(\alpha - \Delta - b)[\Phi(\alpha - \Delta)]^{M-2}\, d\alpha. \tag{35}$$

Also note that

$$\begin{aligned} [\Phi(\alpha - \Delta)]^{M-2} &= [1 - \Phi_c(\alpha - \Delta)]^{M-2} \\ &\leqq \exp\,[-(M - 2)\Phi_c(\alpha - \Delta)]. \end{aligned} \tag{36}$$

As in the proof of Theorem 1, we break the integral in (35) into two parts $\alpha \leqq a$ and $\alpha \geqq a$, where a will be specified later. In the range $\alpha \leqq a$ we overbound $\Phi(\alpha - \Delta - b)$ by unity, and $[\Phi(\alpha - \Delta)]^{M-2}$ by (36). In the range $\alpha \geqq a$, we overbound $[\Phi(\alpha - \Delta)]^{M-2}$ by unity. Thus

$$\begin{aligned} P_e \leqq{} & M \int_{-\infty}^{a} g(\alpha) \exp\,[-(M - 2)\Phi_c(\alpha - \Delta)]\, d\alpha \\ & + M \int_{a}^{\infty} g(\alpha)\Phi(\alpha - \Delta - b)\, d\alpha = MP_1 + MP_2\,. \end{aligned} \tag{37}$$

We now overbound P_1 and P_2. First,

$$\begin{aligned} P_1 &= \int_{-\infty}^{a} g(\alpha) \exp\,[-(M - 2)\Phi_c(\alpha - \Delta)]\, d\alpha \\ &\leqq \exp\,[-(M - 2)\Phi_c(a - \Delta)] \int_{-\infty}^{a} g(\alpha)\, d\alpha \\ &\leqq \exp\,[-(M - 2)\Phi_c(a - \Delta)]. \end{aligned} \tag{38}$$

Second, if we assume that

$$a \leqq b + \Delta \tag{39}$$

we can write

$$P_2 = \int_a^\infty g(\alpha)\Phi(\alpha - \Delta - b)\, d\alpha$$

$$= \int_a^{\Delta+b} g(\alpha)\Phi(\alpha - \Delta - b)\, d\alpha + \int_{\Delta+b}^\infty g(\alpha)\Phi(\alpha - \Delta - b)\, d\alpha.$$

In the first integral, $\alpha - \Delta - b \leqq 0$, so that we may use Lemma 1 to bound $\Phi(\alpha - \Delta - b)$. In the second integral, we overbound $\Phi(\alpha - \Delta - b)$ by unity. Thus

$$P_2 \leqq \int_a^{\Delta+b} g(\alpha) \exp\,[-(\alpha - \Delta - b)^2/2]\, d\alpha + \int_{\Delta+b}^\infty g(\alpha)\, d\alpha$$

$$\leqq \int_a^\infty g(\alpha) \exp\,[-(\alpha - \Delta - b)^2/2]\, d\alpha + \Phi_c(\Delta + b).$$

Since from (20) and the fact that $R < C$,

$$\Delta + b = (\delta + C^{\frac{1}{2}})(2T)^{\frac{1}{2}} > 2(C^{\frac{1}{2}} - R^{\frac{1}{2}})(2T)^{\frac{1}{2}} > 0,$$

we can again use Lemma 1 to overbound $\Phi_c(\Delta + b)$. Using the definition of $g(\alpha)$, we have

$$P_2 \leqq \int_a^\infty \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\,(-\alpha^2/2) \exp\,[-(\alpha - \Delta - b)^2/2]\, d\alpha$$
$$+ \exp\,[-(\Delta + b)^2/2]$$

$$= \exp\,[-(b + \Delta)^2/4] \frac{1}{(2\pi)^{\frac{1}{2}}} \int_a^\infty \exp\left[-\left(\alpha - \frac{b + \Delta}{2}\right)^2\right] d\alpha$$
$$+ \exp\,[-(\Delta + b)^2/2]$$

$$= \exp\,[-(b + \Delta)^2/4] \frac{2^{-1/2}}{(2\pi)^{\frac{1}{2}}} \int_{\sqrt{2}[a-(b+\Delta)/2]}^\infty \exp\,(-v^2/2)\, dv$$
$$+ \exp\,[-(\Delta + b)^2/2]$$

$$\leqq \exp\,[-(b + \Delta)^2/4]\Phi_c\left[\sqrt{2}\left(a - \frac{b + \Delta}{2}\right)\right] + \exp\,[-(\Delta + b)^2/2].$$

If we further assume that

$$a \geqq (b + \Delta)/2, \tag{40}$$

then we can again employ Lemma 1 to bound $\Phi_c[\sqrt{2}(a - (b + \Delta)/2)]$. Hence

$$P_2 \leqq \exp[-(b+\Delta)^2/4] \exp\left\{-\left[a - \left(\frac{b+\Delta}{2}\right)\right]^2\right\} + \exp[-(\Delta+b)^2/2] \tag{41}$$

$$= \exp\{-\tfrac{1}{2}[a^2 + (a-\Delta-b)^2]\} + \exp[-(\Delta+b)^2/2].$$

The difference between the second and first exponents in (41) is

$$-\tfrac{1}{2}\{(\Delta+b)^2 - [a^2 + (a-\Delta-b)^2]\} = a[a - (\Delta+b)] \leqq 0,$$

by (39) and (40). Thus, the first term of (41) is not less than the second, and

$$P_2 \leqq 2\exp\{-\tfrac{1}{2}[a^2 + (a-\Delta-b)^2]\}. \tag{42}$$

Inserting the bounds on P_1 (38) and P_2 (42) into (37), we obtain

$$P_e \leqq M\exp[-(M-2)\Phi_c(a-\Delta)] + 2M\exp\{-\tfrac{1}{2}[a^2 + (a-\Delta-b)^2]\}, \tag{43a}$$

where from equations (39) and (40),

$$\frac{b+\Delta}{2} \leqq a \leqq b+\Delta. \tag{43b}$$

It remains to choose the parameter a, and here we will simply state a good choice of a without giving a motivating argument. Let

$$a = (R^{\frac{1}{2}} + \delta - \theta_2)(2T)^{\frac{1}{2}} \tag{44}$$

(where θ_2 is the arbitrary parameter which was selected at the beginning of the proof). We must verify that constraints (43b) are satisfied for this choice of a. First, since $R < C$ and $\theta_2 > 0$,

$$b + \Delta - a = (C^{\frac{1}{2}} - R^{\frac{1}{2}} + \theta_2)(2T)^{\frac{1}{2}} > 0.$$

Thus $a \leqq b + \Delta$. Second, from equation (21b),

$$a - \left(\frac{b+\Delta}{2}\right) = \tfrac{1}{2}\{\delta - [C^{\frac{1}{2}} - (4R)^{\frac{1}{2}}] - 2\theta_2\}(2T)^{\frac{1}{2}} \geqq 0,$$

so that $a \geqq (b+\Delta)/2$ and (43b) is satisfied.

Now consider the second term in (43a). Direct substitution of (44) shows that this term is

$$2M\exp\{-[(R^{\frac{1}{2}} + \delta - \theta_2)^2 + (C^{\frac{1}{2}} - R^{\frac{1}{2}} + \theta_2)^2]T\},$$

a single exponential decay in T (as $T \to \infty$). Finally consider the exponent of the first term of (43). Substituting (44), it is

$$-(M-2)\Phi_c(a-\Delta) = -(\exp(RT)-2)\Phi_c\{[R^{\frac{1}{2}}-\theta_2](2T)^{\frac{1}{2}}\}.$$

Making use of the asymptotic formula $\Phi_c(u) \approx (2\pi u)^{-\frac{1}{2}}e^{-u^2/2}$ as $u \to \infty$ (see p. 106 of Ref. 2), and letting $T \to \infty$ (and noting that from equation (21b), $R^{\frac{1}{2}} - \theta_2 > 0$), this exponent is asymptotic to

$$\frac{-1}{(2\pi)^{\frac{1}{2}}(R^{\frac{1}{2}}-\theta_2)(2T)^{\frac{1}{2}}}\exp(+KT),$$

where $K > 0$. Thus the first term of equation (43a) decays to zero as a double exponential in T, very much more rapidly than the second term of equation (43a). We can find a T sufficiently large so that the ratio of the first to second terms of equation (43a) $\leqq \theta_1$. With T so chosen

$$P_e \leqq (1+\theta_1)2\exp\{-[(R^{\frac{1}{2}}+\delta-\theta_2)^2+(C^{\frac{1}{2}}-R^{\frac{1}{2}}+\theta_2)^2-R]T\}$$

which is Theorem 2.

REFERENCES

1. Fano, R. M., *Transmission of Information,* Cambridge, Massachusetts: MIT Press, 1961.
2. Feller, W., *An Introduction to Probability Theory and Its Applications,* vol. I, New York: Wiley, 1950.
3. Forney, G. D., "Exponential Error Bounds for Erasure, List, and Decision Feedback Schemes," IEEE Trans. on Information Theory, *IT-14,* No. 2 (March 1968), pp. 206–220.
4. Kramer, A. J., "Use of Orthogonal Signaling in Sequential Decision Feedback," Information and Control, *10,* No. 5 (May 1967), pp. 509–521.
5. Kramer, A. J., "Analysis of Communication Schemes Using an Intermittent Feedback Link," Stanford University Center for Systems Research, Tech. Rept. No. 7050-11.
6. Schalkwijk, J. P. M. and Kailath, T., "A Coding Scheme for Additive Channels with Feedback, IEEE Trans. on Information Theory, *IT-12,* No. 2 (April 1966), pp. 172–188.
7. Shannon, C. E., Gallager, R. G., and Berlekamp, E. R., "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels," Information and Control, *10,* No. 1, 5 (January and May 1967), pp. 65–103 and 522–552.
8. Wozencraft, J. and Jacobs, I., *Principles of Communication Engineering,* New York: Wiley, 1965.
9. Wyner, A. D., "On the Probability of Error for Communication in White Gaussian Noise," IEEE Trans. on Information Theory, *IT-13,* No. 1 (January 1967), pp. 86–90.
10. Wyner, A. D., "On the Schalkwijk-Kailath Coding Scheme with a Peak Energy Constraint," IEEE Trans. on Information Theory, *IT-14,* No. 1 (January 1968), pp. 129–134.