XVII. Communications Systems Research: Coding and Synchronization Studies

TELECOMMUNICATIONS DIVISION

A. Performance of a Low-Rate Command Data Link, *S. Farber*

1. Introduction

This article gives the performance of an orthogonal signal frequency-shift-keyed command link. This ground-to-spacecraft link will code low-rate binary information into one of two frequency-modulated tones modulated onto the carrier. The purpose of this article is to examine the error performance capabilities of such a scheme under the assumptions (1) that the phase error out of the spacecraft tracking loop does not vary significantly over a bit time and (2) that the phase error out of the spacecraft tracking loop does vary significantly over a bit time.

2. System Model

The transmitted signal is assumed to be of the form

$$s(t) = (2P)^{1/2} \cos [\omega_c t + k \sin (\omega t + \theta) + \Psi], \qquad \omega = \omega_0, \omega_1$$

where ω_c is the carrier frequency, k the index of modulation, and $\omega = \omega_0$ represents a *zero* being transmitted while $\omega = \omega_1$ represents a *one* being transmitted. The angles Ψ and θ are assumed to be uniformly distributed random variables defined on the interval $-\pi, \pi$ radians.

The signal $s(t)$ can be expanded in a Fourier series about ω_c to yield

$$\begin{aligned} s(t) = {} & (2P)^{1/2} J_0(k) \cos(\omega_c t + \Psi) \\ & - (2P)^{1/2} J_1(k) \{\cos[\omega_c t - (\omega t + \theta) + \Psi] - \cos[\omega_c t + (\omega t + \theta) + \Psi]\} \\ & + (2P)^{1/2} J_2(k) \{\cos[\omega_c t - 2(\omega t + \theta) + \Psi] + \cos[\omega_c t + 2(\omega t + \theta) + \Psi]\} \\ & - (2P)^{1/2} J_3(k) \{\cos[\omega_c t - 3(\omega t + \theta) + \Psi] - \cos[\omega_c t + 3(\omega t + \theta) + \Psi]\} \\ & + (2P)^{1/2} J_4(k) \{\cos[\omega_c t - 4(\omega t + \theta) + \Psi] + \cos[\omega_c t + 4(\omega t + \theta) + \Psi]\} \\ & - \cdots \end{aligned}$$

where J_k is the Bessel function of order k. An indication of the behavior of $J_i(k)$, $i = 0, 1, 2$, can be seen in Fig. 1 for values of k satisfying $0 \leq k \leq 2$.

If the tracking loop works on the fundamental component, it will form an estimate $\hat{\Psi}(t)$ of $\Psi(t)$ so that the received signal mixed with $(2)^{1/2} \sin[\omega_c t + \hat{\Psi}(t)]$ and filtered will yield

$$r_1(t) = (P)^{1/2} \sin[k \sin(\omega t + \theta) + \phi(t)] + n_1(t)$$

while the received signal mixed with $(2)^{1/2} \cos[\omega_c t + \hat{\Psi}(t)]$ and filtered will yield

$$r_2(t) = (P)^{1/2} \cos[k \sin(\omega t + \theta) + \phi(t)] + n_2(t)$$

where $\phi(t) = \hat{\Psi}(t) - \Psi$ and $n_1(t)$ and $n_2(t)$ represent independent white gaussian noise of single-sided spectral density N_0 (Ref. 1).

If the tracking loop is a phase-locked loop preceded by a bandpass limiter, then the distribution on ϕ as given by Lindsey (Ref. 2) using DSN parameters is

$$p(\phi) = \frac{\exp(\rho_L \cos\phi)}{2\pi I_0(\rho_L)}, \qquad -\pi < \phi \leq \pi$$

where

$$\rho_L = \frac{3z}{\Gamma\left(1 + \frac{2}{\mu}\right)}, \qquad \Gamma = \frac{1 + 0.345\,zy}{0.862 + 0.690\,zy}$$

$$z = \frac{P_c}{N_0 b_{L0}}, \qquad y = \frac{1}{800}$$

$$\mu = \frac{(\gamma_0)^{1/2} \exp\left(-\frac{\gamma_0 y}{2}\right)\left[I_0\left(\frac{\gamma_0 y}{2}\right) + I_1\left(\frac{\gamma_0 y}{2}\right)\right]}{(z)^{1/2} \exp\left(-\frac{zy}{2}\right)\left[I_0\left(\frac{zy}{2}\right) + I_1\left(\frac{zy}{2}\right)\right]}$$

and $\gamma_0 = 4$. (It should be noted that the usual DSN parameters are $y = 1/400$ and $\gamma_0 = 2$.) I_k is the modified Bessel function of order k. P_c represents the power in the carrier, b_{L0} the loop design bandwidth, and N_0 the single-sided spectral density of the noise. For the above signal, we find $P_c = PJ_0^2(k)$.

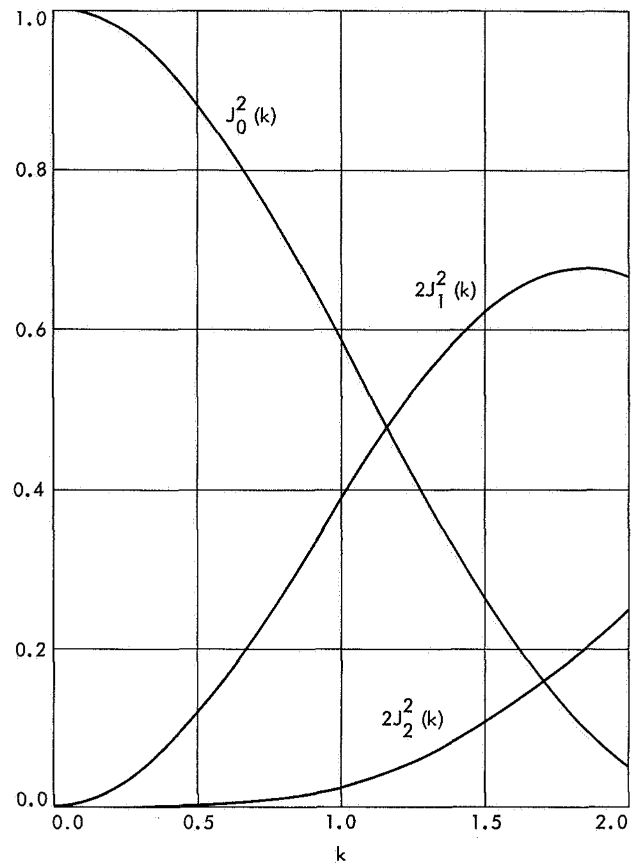

Fig. 1. Plot of J_0^2, $2J_1^2$, and $2J_2^2$, showing division of power between fundamental and other components

3. Error Rates for Various Detectors

For convenience, let us define the random variables

$$k_\phi = \frac{1}{T_b}\int_0^{T_b} \cos\phi(t)\,dt$$

$$\lambda_\phi = \frac{1}{T_b}\int_0^{T_b} \sin\phi(t)\,dt$$

where T_b is the time per bit.

If the data is extracted using only the component of $r_1(t)$ at frequency ω, namely,

$$(P)^{1/2} \cos\phi(t)\, 2J_1(k) \sin(\omega t + \theta) + n_1(t), \qquad 0 < t \leq T_b;\ \omega = \omega_0, \omega_1$$

then the problem is essentially to decide which of two signals is present. Hence, an incoherent phase receiver using orthogonal signals can be used to obtain a bit probability of error (Ref. 3) of

$$P_E^I = E\left\{\frac{1}{2}\exp\left[-\frac{1}{2}k_\phi^2 R\right]\right\}$$

where E is the expectation operation, $R = ST_b/N_0$, and $S = 2J_1^2(k)P$ is the power in the data.

If the data is extracted from the fundamental components of both $r_1(t)$ and $r_2(t)$, namely,

$$(P)^{1/2}\cos\phi(t)\, 2J_1(k)\sin(\omega t + \theta) + n_1(t), \qquad 0 < t < T_b$$

and

$$(P)^{1/2}\sin\phi(t)\, 2J_1(k)\sin(\omega t + \theta) + n_2(t), \qquad \omega = \omega_0, \omega_1$$

then by using the doubly incoherent receiver discussed in *Subsection 8*, it is possible to obtain a probability of error of

$$P_E^D = \min_{0 \leq \beta \leq 1} E\{\tfrac{1}{2}c(\beta)\exp[-\tfrac{1}{2}(k_\phi^2 + \lambda_\phi^2)R]\}$$

where

$$c(\beta) = \frac{\exp\left[\frac{1}{2}\left(\frac{1-\beta}{1+\beta}\right)\lambda_\phi^2 R\right] - \beta^2 \exp\left[-\frac{1}{2}\left(\frac{1-\beta}{1+\beta}\right)k_\phi^2 R\right]}{1-\beta^2}$$

and β is an arbitrary gain factor, $0 \leq \beta \leq 1$.

We note that when $\beta = 0$, the doubly incoherent receiver degenerates to the incoherent receiver so that we always have $P_E^D \leq P_E^I$ for a given index of modulation k. Since, as the index of modulation increases from zero, the amount of power in the data will increase, causing R to increase, while the amount of power in the carrier will decrease, causing ρ_L to decrease, there will be an optimum value of k, corresponding to an optimum division of power. In particular, ρ_L depends on

$$z = \frac{P_c}{N_0 b_{L0}} = \frac{PT_b}{N_0} \cdot \frac{1}{b_{L0}T_b} \cdot J_0^2(k)$$

where

$$R = \frac{ST_b}{N_0} = \frac{PT_b}{N_0} \cdot 2J_1^2(k)$$

By letting $\delta = 1/2b_{L0}T_b$ and $\mathscr{R} = PT_b/N_0$, we can write

$$z = 2\mathscr{R}\,\delta\, J_0^2(k)$$

$$R = 2\mathscr{R}\, J_1^2(k)$$

[It should be noted that Lindsey (Ref. 2) uses $\delta = 1/b_{L0}T_b$.]

4. Extremely Low Data Rates

When the data rate is extremely low, corresponding to $\delta << 1$, it is appropriate (Ref. 1) to use the approximations

$$k_\phi = \frac{1}{T_b}\int_0^{T_b} \cos\phi(t)\, dt \simeq E\{\cos\phi\}$$

and

$$\lambda_\phi = \frac{1}{T_b}\int_0^{T_b} \sin\phi(t)\, dt \simeq E\{\sin\phi\}$$

Using Lindsey's model for the density on ϕ as given above, we find

$$k_\phi \simeq \eta = \frac{I_1(\rho_L)}{I_0(\rho_L)}$$

$$\lambda_\phi \simeq 0$$

The optimum value of β for the doubly incoherent receiver then occurs at $\beta = 0$ so that the doubly incoherent receiver reduces to the incoherent receiver with probability of error given by

$$P_E^D = P_E^I = \tfrac{1}{2}\exp\left(-\tfrac{1}{2}\eta^2 R\right)$$

The resulting minimum value of the probability of error is plotted in Fig. 2a versus $\mathcal{R}$ for several values of δ, while the optimum values of k are plotted in Fig. 2b and the resulting values of ρ_L are plotted in Fig. 2c. In order that the tracking loop acquire frequency lock, it is necessary to require that $\rho_L \geqq 6$.

5. Moderate Data Rates

Moderate data rates occur when the phase does not vary significantly over a bit time so that $\delta \cong 1$ and the approximations

$$k_\phi = \frac{1}{T_b}\int_0^{T_b} \cos\phi(t)\,dt \cong \cos\phi$$

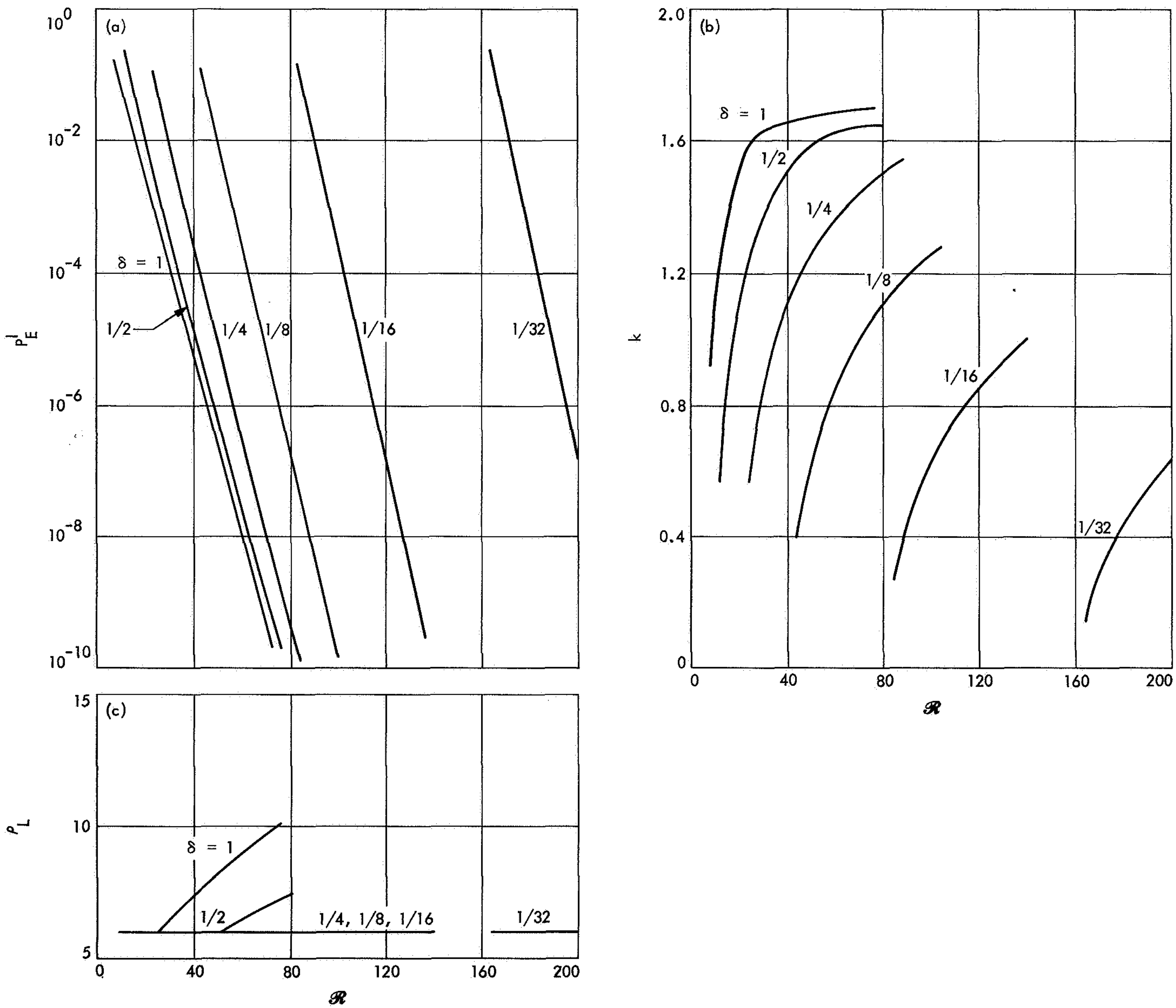

Fig. 2. Plots of behavior of incoherent receiver under the assumption of non-constant phase, showing (a) probability of error, (b) optimal value of modulation index k, and (c) resulting value of ρ_L

and

$$\lambda_\phi = \frac{1}{T_b}\int_0^{T_b} \sin\phi(t)\,dt \cong \sin\phi$$

are valid.

Under these circumstances, we find the probability of error for the incoherent receiver is

$$P_E^I = \int_{-\pi}^{\pi} \tfrac{1}{2}\exp\left(-\tfrac{1}{2}R\cos^2\phi\right)\exp\left(\rho_L\cos\phi\right)\frac{d\phi}{2\pi I_0(\rho_L)}$$

The minimum value of P_E^I is plotted in Fig. 3a, the optimum value of k to give this value of P_E^I is plotted in Fig. 3b, and the resulting value of ρ_L is plotted in Fig. 3c.

The probability of error for the doubly incoherent receiver is

$$P_E^D = \min_{0\leq\beta\leq 1}\int_{-\pi}^{\pi} c(\beta)\frac{d\phi}{2\pi I_0(\rho_L)}\tfrac{1}{2}\exp\left(-\tfrac{1}{2}R\right)$$

where

$$c(\beta) = \frac{\exp\left[\frac{1}{2}\left(\frac{1-\beta}{1+\beta}\right)R\sin^2\phi\right] - \beta^2\exp\left[-\frac{1}{2}\left(\frac{1-\beta}{1+\beta}\right)R\cos^2\phi\right]}{1-\beta^2}$$

We note that, when $\beta\rightarrow 1$, we can evaluate $c(\beta)$ by L'Hospital's rule to find

$$c(1) = 1 + \frac{R}{8}$$

which is independent of ϕ. Combining this with the fact that the performance of the doubly incoherent detector cannot be better than the performance of the incoherent detector with $\phi = 0$, we find

$$\frac{1}{2}\exp\left(-\frac{1}{2}R\right)\leq P_E^D \leq \frac{1}{2}\left(1+\frac{R}{8}\right)\exp\left(-\frac{1}{2}R\right)$$

so that P_E^D must be exponentially asymptotic to the optimum receiver performance for the given signaling scheme.

The minimum value of P_E^D is plotted in Fig. 4a, the optimum value of k to give this value of P_E^D is plotted in Fig. 4b, the resulting value of ρ_L is plotted in Fig. 4c, and the optimum value of β is plotted in Fig. 4d.

6. In Between Rates

For rates between those discussed in *Subsections* 3, 4, and 5, we expect the probabilities of error to fall somewhere in between the probabilities of error obtained above.

This would imply that as the rate increases from extremely low to moderate, the probability of error for the incoherent detector would increase from the values in Fig. 2a to the much larger values in Fig. 3a. The probability of error for the doubly incoherent detector, however, would decrease from the values in Fig. 2a to the slightly lower values in Fig. 4a. The desirability of using a doubly incoherent receiver, which is somewhat more complicated to implement as opposed to the simpler incoherent receiver, would, of course, depend on the exact probability of error for the rate under consideration.

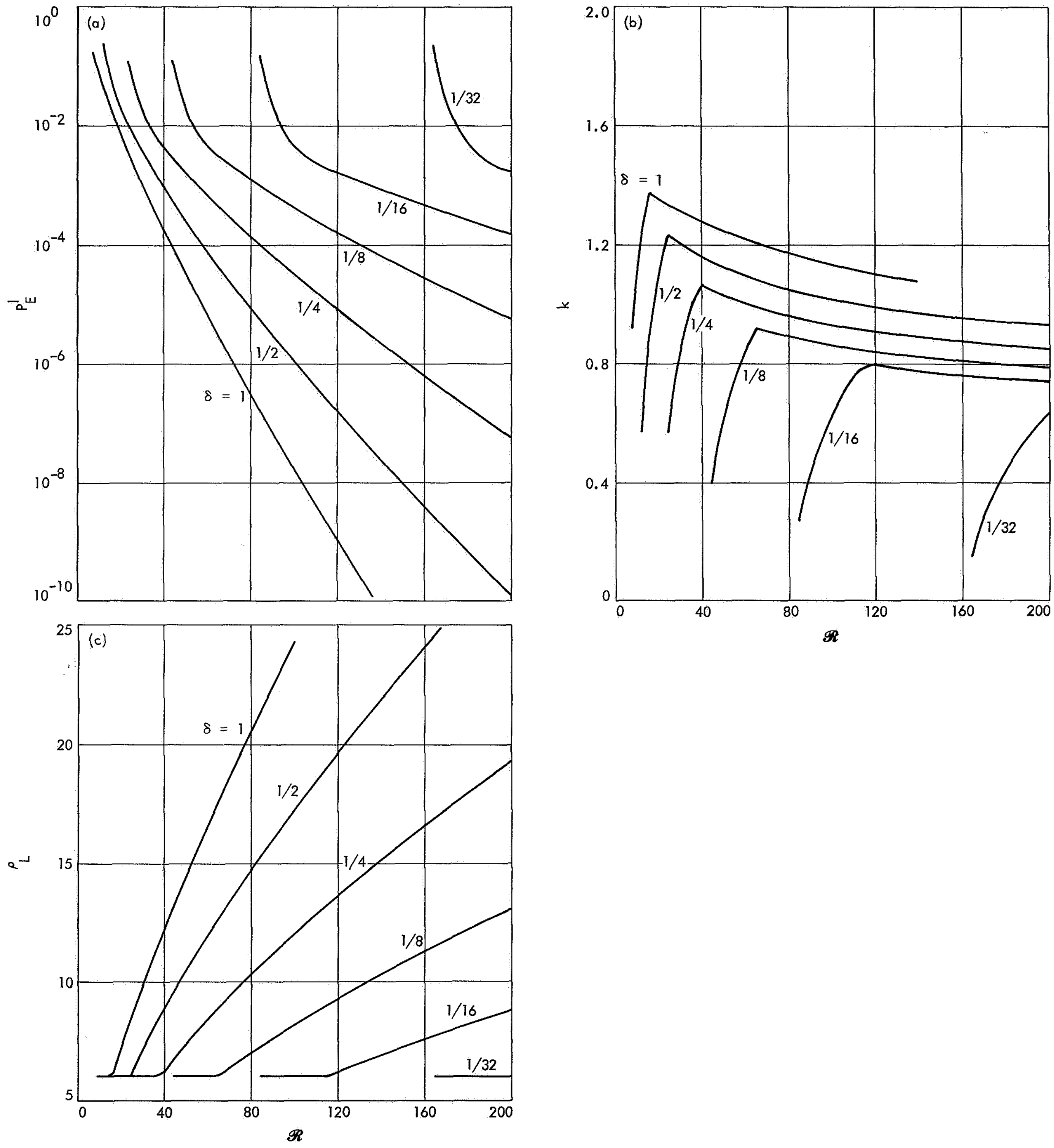

Fig. 3. Plots of behavior of incoherent receiver under the assumption of constant phase, showing (a) probability of error, (b) optimal value of modulation index *k*, and (c) resulting value of ρ_L

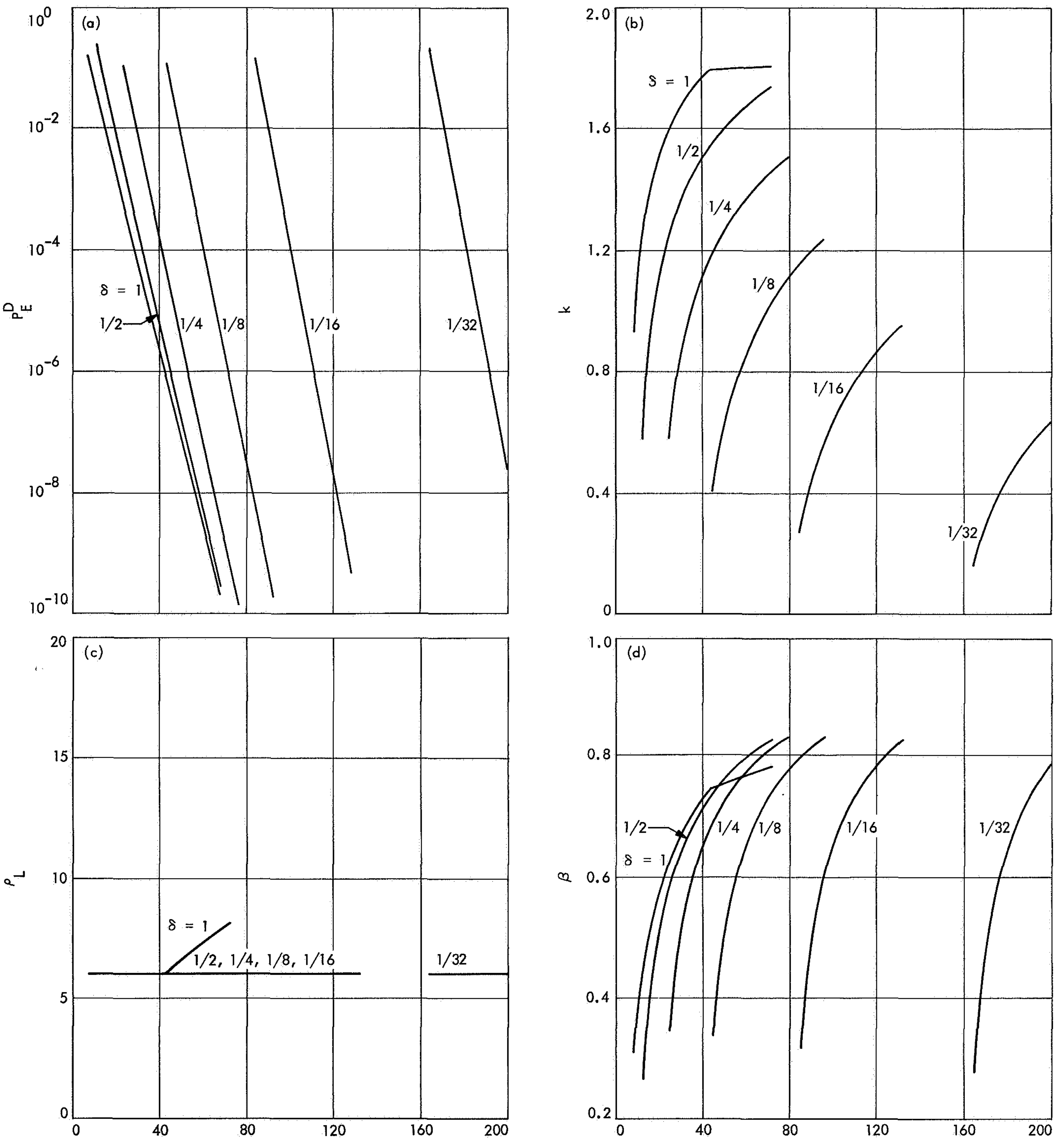

Fig. 4. Plots of behavior of doubly incoherent receiver under the assumption of constant phase, showing (a) probability of error, (b) optimal value of modulation under k, (c) resulting value of ρ_L, and (d) optimal value of β

7. Using the Second Harmonic

The doubly incoherent detector may also be used to extract information about the data from the first harmonic of $r_1(t)$ and the second harmonic of $r_2(t)$; namely,

$$(P)^{1/2} \cos\phi(t)\, 2J_1(k) \sin(\omega t + \theta) + n_1(t), \qquad \omega = \omega_0, \omega_1$$

and

$$(P)^{1/2} \cos\phi(t)\, J_2(k) \cos 2(\omega t + \theta) + n_2(t), \qquad 0 < t \leq T_b$$

This would yield a probability of error of

$$P_E^D = \min_{0 \leq \beta \leq 1} E\{\tfrac{1}{2} c(\beta) \exp[-\tfrac{1}{2} k_\phi^2 (R_1 + R_2)]\}$$

where

$$c(\beta) = \frac{\exp\left[\frac{1}{2}\left(\frac{1-\beta}{1+\beta}\right) k_\phi^2 R_2\right] - \beta^2 \exp\left[-\frac{1}{2}\left(\frac{1-\beta}{1+\beta}\right) k_\phi^2 R_1\right]}{1-\beta^2}$$

and

$$R_1 = 2J_1^2(k)\,\mathcal{R}$$

$$R_2 = 2J_2^2(k)\,\mathcal{R}$$

This yields an improvement in signal-to-noise ratio of about $J_2^2(k)/J_1^2(k)$ over the incoherent receiver. An indication of this ratio can be obtained from Fig. 1. For values of k near 1, the improvement is about 10%.

8. Description and Analysis of the Doubly Incoherent Receiver

We assume two signals of the form

$$r_1(t) = (P)^{1/2} \sin[k \sin(\omega t + \theta_1) + \phi] + n_1(t)$$

and

$$r_2(t) = (P)^{1/2} \cos[k \sin(\omega t + \theta_2) + \phi] + n_2(t), \qquad 0 < t \leq T;\ \omega = \omega_0, \omega_1$$

where the angles θ_1 and θ_2 are arbitrary and $n_1(t)$ and $n_2(t)$ are the independent white gaussian noise process of the one-sided spectral density N_0.

The doubly incoherent receiver then consists of two sections of the form shown in Fig. 5, one with $\omega = \omega_0$ and one with $\omega = \omega_1$. The variable β is an arbitrary gain factor which is to be chosen to minimize the probability of error. If we define the random variables

$$k_\phi = \frac{1}{T}\int_0^T \cos\phi(t)\,dt$$

and

$$\lambda_\phi = \frac{1}{T}\int_0^T \sin\phi(t)\,dt$$

then the output of the ω_0 section when $\sin\omega_0 t$ was transmitted is

$$Q_0 = u_1^2 + u_1^2 + u_2^2 + \beta(u_3^2 + u_4^2)$$

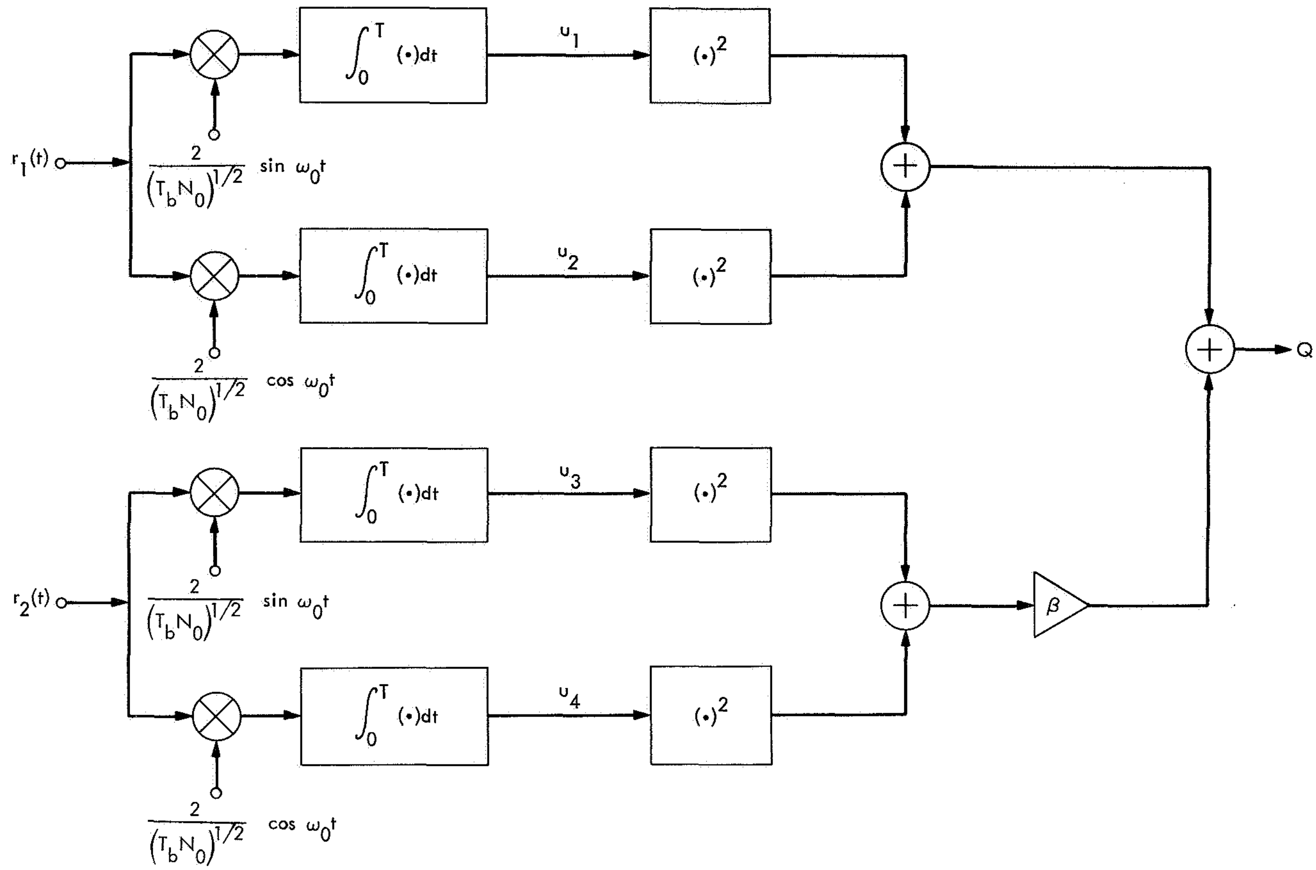

Fig. 5. Diagram of one section of the doubly incoherent receiver

where

$$u_1 = \left(\frac{2E}{N_0}\right)^{1/2} k_\phi \cos\theta_1 + n_1, \qquad u_2 = \left(\frac{2E}{N_0}\right)^{1/2} k_\phi \sin\theta_1 + n_2$$

$$u_3 = \left(\frac{2E}{N_0}\right)^{1/2} \lambda_\phi \cos\theta_2 + n_3, \qquad u_4 = \left(\frac{2E}{N_0}\right)^{1/2} \lambda_\phi \sin\theta_2 + n_4$$

and the output of the ω_1 section when $\sin\omega_0 t$ was transmitted is

$$Q_1 = v_1^2 + v_2^2 + \beta\,(v_3^2 + v_4^2)$$

where

$$v_1 = m_1, \qquad v_2 = m_2, \qquad v_3 = m_3, \qquad v_4 = m_4$$

The noises n_i and m_i, $i = 1$ to 4, are mutually independent gaussian random variables of unit variance. Similar variables are defined in a symmetrical way when $\sin\omega_1 t$ was transmitted.

The estimate of which value of ω was sent is taken to correspond to the section with the largest output. Thus, the probability of error is given by

$$P_E^D = \min_{0 \leq \beta \leq 1} P_r\,(Q_0 < Q_1 \,|\, \omega = \omega_0)$$

assuming that prob $(\omega = \omega_0) =$ prob $(\omega = \omega_1)$.

By first conditioning on Q_0, we readily find

$$P_r(Q_0 < Q_1 | \omega = \omega_0, k_\phi, \lambda_\phi, Q_0) = \frac{1}{1-\beta} \exp\left(-\frac{1}{2} Q_0^2\right) - \frac{1}{1-\beta} \exp\left(-\frac{1}{2} Q_0^2\right)$$

But we have that

$$P_E^D = \min_{0 \leq \beta \leq 1} E\{P_r(Q_0 < Q_1 | \omega = \omega_0, k_\phi, \lambda_\phi, Q_0)\}$$

where the expectation is taken over the variables n_1, n_2, n_3, and n_4 and the functionals k_ϕ and λ_ϕ. A straightforward integration yields

$$P_E^D = \min_{0 \leq \beta \leq 1} \frac{1}{2} E \frac{\exp\left[-\left(\frac{\beta}{1+\beta}\lambda_\phi^2 + \frac{1}{2}k_\phi^2\right)\frac{PT_b}{N_0}\right] - \beta^2 \exp\left[-\left(\frac{1}{2}\lambda_\phi^2 + \frac{1}{1+\beta}k_\phi^2\right)\frac{PT_b}{N_0}\right]}{1-\beta^2}$$

where the expectation is now only over k_ϕ and λ_ϕ. This is the expression used in *Subsection 3.*

---

9. Conclusion

For the schemes discussed, it can be seen that the optimum value of the index of modulation k for low rates is almost always given by the constraint $\rho_L = 6$.

Also, for certain rates the doubly incoherent receiver gives considerably better error performance than the incoherent receiver. Just how much better for a given rate, however, remains an open question which can perhaps best be answered by simulation.

References

1. Viterbi, A. J., *Optimum Detection and Signal Selection for Partially Coherent Binary Communication*, Wescon 13.1. Western Electronic Manufacturers' Association, Los Angeles, Calif., 1964.
2. Lindsey, W. C., "Performance of Phase-Coherent Receivers Preceded by Bandpass Limiters," *IEEE Trans. on Commun. Technol.*, April 1968.
3. Wozencraft and Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, Inc., New York, 1965.

B. Analysis of a Serial Orthogonal Decoder,

R. R. Green

1. Introduction

This article presents a more straightforward mathematical analysis of the decoder discussed in SPS 37-39, Vol. IV, pp. 247–252. As before, the problem is to perform the matrix vector product $y = H_n x$, where x is a real vector with 2^n components. H_n is the code matrix, or dictionary, defined inductively by $H_n = H_{n-1} \otimes H_1$, with

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and $\otimes$ denotes the Kronecker product.

2. Notation

Subscripts are used to denote the size of matrices in the following way: A_m implies that A_m is a 2^m by 2^m square matrix. I_m denotes the 2^m by 2^m identity matrix.

The Kronecker product of two matrices, say A and B, is defined by $A \otimes B = (a_{ij} B)$. This product is associative, i.e.,

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

and, if the dimensions are correct for the necessary ordinary matrix products to be defined, we have (Ref. 1)

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

From the foregoing, we have the following useful relations:

$$I_m \otimes I_n = I_{m+n}$$

$$(I_m \otimes A_n)(I_m \otimes B_n) = I_m \otimes A_n B_n$$

$$(A_n \otimes I_m)(B_n \otimes I_m) = A_n B_n \otimes I_m$$

$$(A_n \otimes I_m)(I_n \otimes B_m) = A_n \otimes B_m = (I_n \otimes B_m)(A_n \otimes I_m)$$

3. Motivation

The difficulty in evaluating $H_n x$ directly, on a term-by-term basis, is the size of H_n. Since every element in H_n is either 1 or -1, direct evaluation would involve $2^n(2^n - 1)$ additions or subtractions. This difficulty can be relieved by factoring H_n into the matrix product of n different 2^n by 2^n matrices, which will be denoted $M_n^{(1)}, M_n^{(2)}, \cdots, M_n^{(n)}$. Each matrix $M_n^{(i)}$ has only two non-zero elements per row, thus only $n2^n$ additions or subtractions are involved.

Furthermore, the structure of each matrix $M_n^{(i)}$ is such that it can be easily implemented with special-purpose digital equipment. Thus, we can construct a set of decoder stages, the first realizing $M_n^{(1)}$, the second $M_n^{(2)}$, etc. If these stages are then connected together serially, the input to stage 1 being x, the output of stage 1 being the input to stage 2, etc., the output of stage n will be y. Due to this serial structure of the decoder, n additions or subtractions are being done simultaneously. Thus, the digital equipment need only be fast enough to perform 2^n additions or subtractions per code word time, or one addition or subtraction per symbol time.

It is an interesting and somewhat surprising result that the stages of the decoder may be connected in an arbitrary order and the output of the last stage will still be the desired vector y.

4. Analysis

The following analysis is a special case of a more general result involving a code matrix which is the Kronecker product of n arbitrary matrices. Since the general case provides no particular additional insight into the decoder under consideration, the results have been particularized to this special case. It should be noted, however, that in the general case the factor matrices have the same form, the same commutivity result holds, and a somewhat more general product theorem can be proved.

Define

$$M_n^{(i)} = I_{n-i} \otimes H_1 \otimes I_{i-1}, \qquad \text{for } 1 \leqq i \leqq n$$

Theorem 1

$$M_n^{(i)} M_n^{(j)} = M_n^{(j)} M_n^{(i)}$$

Proof. Assume $i > j$ (if $i = j$, the result is trivial) then

$$\begin{aligned} M_n^{(i)} M_n^{(j)} &= (I_{n-i} \otimes H_1 \otimes I_{i-1})(I_{n-j} \otimes H_1 \otimes I_{j-1}) \\ &= [(I_{n-i} \otimes H_1 \otimes I_{i-j-1}) \otimes I_j]\,[I_{n-j} \otimes (H_1 \otimes I_{j-1})] \\ &= I_{n-i} \otimes H_1 \otimes I_{i-j-1} \otimes H_1 \otimes I_{j-1} \\ &= [I_{n-i+1} \otimes (I_{i-j-1} \otimes H_1 \otimes I_{j-1})][(I_{n-i} \otimes H_1) \otimes I_{i-1}] \\ &= (I_{n-j} \otimes H_1 \otimes I_{j-1})(I_{n-i} \otimes H_1 \otimes I_{i-1}) \\ &= M_n^{(j)} M_n^{(i)} \end{aligned}$$

Thus, Theorem 1 shows that the order of any two successive stages may be interchanged, and thus any possible permutation of the stages may be realized, without changing the final output. Also, the commutivity shown implies that we need not keep track of order when discussing matrix products of the $M_n^{(i)}$.

Theorem 2

$$\prod_{i=1}^{m} M_n^{(i)} = I_{n-m} \otimes H_m, \qquad 1 \leqq m \leqq n$$

Proof. For $m = 1$, we have

$$\prod_{i=1}^{m} M_n^{(i)} = M_n^{(1)} = I_{n-1} \otimes H_1$$

Assume the result is true for m, then prove for $m+1$:

$$\prod_{i=1}^{m+1} M_n^{(i)} = M_n^{(m+1)} \prod_{i=1}^{m} M_n^{(i)} = (I_{n-m-1} \otimes H_1 \otimes I_m)(I_{n-m} \otimes H_m) = I_{n-m-1} \otimes H_1 \otimes H_m = I_{n-m-1} \otimes H_{m+1}$$

Thus, by induction, the result is true for any m between 1 and n.

In particular, we see from Theorem 2, letting $m = n$, that

$$\prod_{i=1}^{n} M_n^{(i)} = I_{n-n} \otimes H_n = I_0 \otimes H_n = H_n$$

Also, as in the previous article on this decoder, it can be shown that

$$M_n^{(i)} = P_n^i R_n (P_n^{i-1})^T$$

where P_n and R_n are defined inductively for $n \geqq 1$ by

$$P_{n+1} = (I_1 \otimes P_n)(P_2 \otimes I_{n-1})$$

and

$$R_{n+1} = (P_2 \otimes I_{n-1})(I_1 \otimes R_n)$$

with $P_1 = I_1$ and $P_2 = (P_{ij})$. Here

$$P_{11} = P_{23} = P_{32} = P_{44} = 1$$

and otherwise $P_{ij} = 0$; $R_1 = H_1$. Thus, we see that connecting the n decoder stages $M_n^{(1)}$ through $M_n^{(n)}$ in any order whatever performs the operation $H_n x$. Furthermore, if the stages are connected in numerical order, $M_n^{(1)}$ first, $M_n^{(2)}$ second, etc., the output at any intermediate stage, say the jth stage, provides a decoder for H_j. Thus, the algorithm has multiple-mission capability.

Reference

1. Bellman, Richard, *Introduction to Matrix Analysis.* McGraw-Hill Book Co., Inc., New York, 1960.

C. Optimal Codes and a Strong Converse for Transmission Over Very Noisy Memoryless Channels, *A. J. Viterbi*[1]

1. Introduction

Wyner (Ref. 1) has obtained the following lower bounds on the asymptotic performance of the optimal codes for the additive white gaussian channel, where T is the message duration, R is the rate in nats/s, and C is the channel capacity:

$$P_E > \exp\{-T[E(R) + o(T)]\}$$

where

$$E(R) = \frac{C}{2} - R, \qquad 0 \leqq R \leqq \frac{C}{4}$$

$$= [(C)^{1/2} - (R)^{1/2}]^2, \qquad \frac{C}{4} \leqq R < C \qquad (1)$$

and

$$1 - P_E < \exp\{-T[E^*(R) + o(T)]\}$$

where

$$E^*(R) = [(R)^{1/2} - (C)^{1/2}]^2, \qquad R > C \qquad (2)$$

The second bound on the probability of correct decision for rates above capacity is generally referred to as a "strong converse."

It is well known (Ref. 2) that equal-energy orthogonal signals are asymptotically optimum because they achieve the error probability

$$P_E < \exp[-T E(R)] \qquad (3)$$

[1]Consultant, University of California at Los Angeles.

where $E(R)$ is given by Eq. (1). We begin by showing, through an application of extreme value theory (Ref. 3), that for rates above capacity orthogonal signals yield

$$1 - P_E > \exp\left\{-T\left[E^*(R) + O\left(\ln\frac{T}{T}\right)\right]\right\}, \qquad R > C \tag{4}$$

which proves their asymptotic optimality[2] above as well as below capacity.

We extend these results by showing the essential equivalence of all memoryless input-discrete very noisy channels to the white gaussian channel and thus extend the strong converse to this wider class of channels.

2. The Additive White Gaussian Channel

Balakrishnan (Ref. 4) has shown that for any equal-energy, *a priori* equiprobable set of M signals used on the white gaussian channel, the probability of correct decision using the optimum (maximum likelihood) decision rule is

$$1 - P_E = M^{-1} e^{-\lambda} E\{\exp[(2\lambda)^{1/2} \max_{1 \leq m \leq M} z_m]\} \tag{5}$$

where $\lambda = CT$, $C = S/N_0$, the ratio of received signal power to one-sided noise spectral density, while $\{z_m\}$ is a set of M zero-mean, unit-variance, gaussian random variables with a covariance matrix whose elements are the normalized integral inner products among signals.

Let $R = \ln M/T$ and restrict to orthogonal signals; Eq. (5) becomes

$$1 - P_E = \exp[-T(C+R)] \times E\left\{\exp\left[(2CT)^{1/2} \max_{1 \leq m \leq e^{RT}} z_m\right]\right\} \tag{6}$$

where $\{z_m\}$ are independent normalized gaussian variables, since the covariance matrix for orthogonal signals is the identity matrix.

Equation (6) can be rewritten as

$$\begin{aligned} 1 - P_E &= \exp[-T(C+R)] \times \int_{-\infty}^{\infty} \exp[(2CT)^{1/2} x] \frac{d}{dx}[F(x)]^{e^{RT}} dx \\ &= \exp(-TC) \int_{-\infty}^{\infty} \exp[(2CT)^{1/2} x] [F(x)]^{e^{RT}-1} d[F(x)] \end{aligned} \tag{7}$$

where

$$F(x) \triangleq \int_{-\infty}^{x} e^{-y^2/2} \frac{dy}{(2\pi)^{1/2}}$$

is the (cumulative) gaussian distribution.

We proceed to evaluate Eq. (7) by applying a technique from extreme value theory due to Cramér (Ref. 3). Consider the transformation

$$1 - \xi e^{-RT} = F(x) \tag{8}$$

which has the inverse (Ref. 3)

$$\begin{aligned} x &= F^{-1}(1 - \xi e^{RT}) \\ &= (2RT)^{1/2} - \frac{\ln 4\pi RT}{2(2RT)^{1/2}} - \frac{\ln \xi}{(2RT)^{1/2}} + O\left(\frac{1}{RT}\right) \end{aligned} \tag{9}$$

Substituting Eqs. (8) and (9) into Eq. (7), we obtain

$$\begin{aligned} 1 - P_E &= \exp[-T(C+R)] \int_0^{e^{RT}} \exp[(2CT)^{1/2} F^{-1}(1 - \xi e^{RT})] (1 - \xi e^{-RT})^{e^{RT}-1} d\xi \\ &= \exp\left\{-T\left[C + R - 2(RC)^{1/2} + \frac{\ln 4\pi RT}{2\left(2\frac{R}{C}\right)^{1/2} T} - O(T^{-3/2})\right]\right\} \int_0^{e^{RT}} \xi^{-(C/R)^{1/2}} (1 - \xi e^{-RT})^{e^{RT}-1} d\xi \end{aligned} \tag{10}$$

[2]It has long been conjectured that regular simplex signals are globally optimum for all rates on the white gaussian channel, and this obviously implies the asymptotic optimality of orthogonal signals. However, only the local first- and second-order conditions of optimality of regular simplex signals have been shown (Ref. 4) and all attempts at proving global optimality at all rates have met with failure.

The last integral is bounded from below by

$$\left\{e\left[1-\left(\frac{C}{R}\right)^{1/2}\right]\right\}^{-1}$$

for $R > C$. Thus, it follows that for orthogonal signals on the white gaussian channel

$$1 - P_E > \exp\left(-T\left\{[(R)^{1/2} - (C)^{1/2}]^2 + O\left(\frac{\ln T}{T}\right)\right\}\right), \quad R > C \tag{11}$$

which proves Inequality (4).

3. Input-Discrete Very Noisy Memoryless Channels

The error probability expression for the white gaussian channel, Eq. (5), can be generalized to any memoryless finite-dimensional (or time-discrete) channel. For any set of M equally likely messages and a maximum likelihood decision rule, for any set of N-dimensional channel input sequences $\{\mathbf{x}^{(j)}; j = 1, 2, \cdots, M\}$, and for $\mathbf{y}$, an N-dimensional output sequence, we have

$$1 - P_E = M^{-1}\sum_{j=1}^{M}\int_{D_j} p(\mathbf{y}|\mathbf{x}^{(j)})\,dy \tag{12}$$

where

$$D_j = \{\mathbf{y} : p(\mathbf{y}|\mathbf{x}^{(j)}) = \max_m p(\mathbf{y}|\mathbf{x}^{(m)})\} \tag{13}$$

Then, since

$$\bigcup_{j=1}^{M} D_j = Y_N$$

the N-dimensional output space, we can rewrite Eq. (12) as

$$\begin{aligned} 1 - P_E &= M^{-1}\int_{Y_N} \max_m p(\mathbf{y}|\mathbf{x}^{(m)})\,dy \\ &= M^{-1}\int_{Y_N} q(\mathbf{y}) \max_m \frac{p(\mathbf{y}|\mathbf{x}^{(m)})}{q(\mathbf{y})}\,dy \\ &= M^{-1} E_{\mathbf{y}}\left[\max_m \frac{p(\mathbf{y}|\mathbf{x}^{(m)})}{q(\mathbf{y})}\right] \end{aligned} \tag{14}$$

where $q(\mathbf{y})$ is an arbitrary probability measure on the output space and $E_{\mathbf{y}}$ is the expectation with respect to this measure. Substitution of the appropriate likelihood functions for the white gaussian channel and for the $q(\mathbf{y})$ corresponding to the likelihood function for a zero-signal hypothesis reduces Eq. (14) to Eq. (5) (cf Helstrom, Ref. 5).

For memoryless time-discrete channels,

$$p(\mathbf{y}|\mathbf{x}^{(m)}) = \prod_{n=1}^{N} p(y_n|x_n^{(m)})$$

and specializing to the independent output measure,

$$q(\mathbf{y}) = \prod_{n=1}^{N} q(y_n)$$

Eq. (14) becomes

$$1 - P_E = M^{-1} E_{\mathbf{y}}\{\exp[\max_m z_m(\mathbf{y})]\} \tag{15}$$

where

$$z_m(\mathbf{y}) = \sum_{n=1}^{N} \ln\left[\frac{p(y_n|x_n^{(m)})}{q(y_n)}\right] \tag{16}$$

Such channel is said to be *very noisy* if

$$p(y_n|x_m^{(n)}) = q(y_n)[1 + \epsilon(x_n^{(m)}, y_n)] \tag{17}$$

where $\epsilon(x, y) \to 0$ uniformly in x and y, and $q(y_n)$ is an arbitrary probability density or distribution. It follows that for any n and m

$$0 = \int_Y p(y|x)\,dy - 1 = \int_Y q(y)\,\epsilon(x, y)\,dy \tag{18}$$

We now restrict attention to a discrete input alphabet of K symbols, so that each $x_n^{(m)}$ is taken from the set $\{x_1, x_2, \cdots, x_K\}$. For this class of channels, we need only consider the class of *fixed composition* codes, which are characterized by the property that each code word is some permutation of the same sequence of N symbols, since Shannon, Gallager, and Berlekamp (Ref. 6) have shown that the asymptotic performance of the best code in the restricted subclass is the same as for the best code in the unrestricted class. Thus, given that the relative frequency of the symbol x_k in each code word of the fixed composition code is

$$\rho_k \triangleq \frac{\text{number of occurrences of } x_k}{N}, \qquad k = 1, 2, \cdots, K$$

we have that the means of the random variables $z_m(\mathbf{y})$ relative to the output measure

$$\prod_{n=1}^{N} q(y_n)$$

are all equal to

$$\left.\begin{aligned} E_{\mathbf{y}}[z_m(\mathbf{y})] &= N\sum_{k=1}^{K}\rho_k\int_Y q(y)\left[\ln\frac{p(y|x_k)}{q(y)}\right]dy \\ &\approx -N\sum_{k=1}^{K}\rho_k\int_Y q(y)\left[\frac{\epsilon^2(x_k,y)}{2}\right]dy \end{aligned}\right\} \tag{19}$$

where we have used Eq. (18) and also Condition (17) to neglect all terms above quadratic in ϵ.

Similarly, since the channel is memoryless,

$$\mathrm{var}_{\mathbf{y}}[z_m(\mathbf{y})] = N\sum_{l=1}^{K}\rho_k\,\mathrm{var}_y\left[\ln\frac{p(y|x_k)}{q(y)}\right] \tag{20}$$

But again neglecting terms above quadratic in ϵ,

$$\mathrm{var}_y\left[\ln\frac{p(y|x_k)}{q(y)}\right] \approx \int_Y q(y)\,\epsilon^2(x_k,y)\,dy \tag{21}$$

Also, the capacity of a very noisy input-discrete memoryless channel is given by

$$\begin{aligned} C &= \max_{\{p_k\}}\sum_{k=1}^{K}p_k\int_Y p(y|x_k)\ln\frac{p(y|x_k)}{q(y)}dy \\ &\approx \max_{\{p_k\}}\sum_{k=1}^{K}p_k\int_Y q(y)\left[1+\epsilon(x_k,y)\right] \\ &\qquad\times\left[\epsilon(x_k,y)-\frac{\epsilon^2(x_k,y)}{2}\right]dy \\ &\approx \max_{\{p_k\}}\sum_{k=1}^{K}p_k\int_Y q(y)\frac{\epsilon^2(x_k,y)}{2}dy \end{aligned} \tag{22}$$

Thus, choosing the relative frequencies $\{\rho_k\}$ corresponding to the maximizing distribution for capacity, we have from Eqs. (19), (20), and (21)

$$E_{\mathbf{y}}[z_m(\mathbf{y})] \approx -NC \tag{23}$$

$$\mathrm{var}_{\mathbf{y}}[z_m(\mathbf{y})] = 2NC \tag{24}$$

Furthermore, since $z_m(\mathbf{y})$ is the sum of N independent random variables, by the central limit theorem it must be asymptotically gaussian. In fact, if we normalize by letting

$$v_m(\mathbf{y}) \triangleq \frac{z_m(\mathbf{y})+NC}{(2NC)^{1/2}} \tag{25}$$

it follows from Eqs. (23) and (24) that $v_m(\mathbf{y})$ is a zero-mean, unit-variance, random variable and by the Berry–Esseen theorem (cf Loève, Ref. 7) we have that $P_v(x)$, the distribution function of the normalized variable v_m, differs from the normalized gaussian distribution $F(x)$ by no more than

$$\begin{aligned} |P_v(x)-F(x)| &\leq \frac{\theta E(|z_m|^3)}{(\mathrm{var}\,z_m)^{3/2}} \\ &\approx \frac{\theta N\sum_{k=1}^{K}\rho_k\int_Y q(y)|\epsilon(x_k,y)|^3dy}{(2NC)^{3/2}} \\ &\approx 0 \end{aligned}$$

when we neglect all terms in ϵ of order higher than quadratic.

Thus, all the variables v_m are asymptotically gaussian with zero means and unit variances. Applying Eq. (25) to Eq. (15) and letting $R' = (\ln M)/N$ nats/symbol,

$$\begin{aligned} 1-P_E &= \exp[-N(R'+C)] \\ &\qquad\times E_{\mathbf{y}}\{\exp[(2NC)^{1/2}\max_{1\leq m\leq e^{NR'}}v_m(y)]\} \end{aligned} \tag{26}$$

This formula is identical to the form of Eq. (6) for orthogonal signals on white gaussian channels, except that the variables v_m are not necessarily independent. However, for rates below capacity it is well known (Ref. 6) that the error probability for the best code on memoryless very noisy input-discrete channels behaves asymptotically exactly as that for orthogonal signals in the white gaussian channel [i.e., Expressions (1) and (3) hold with T replaced by N and R replaced by R']. For this to be the case, the best code on memoryless very noisy input-discrete memoryless channels must asymptotically lead to independent $v_m(y)$ in Eq. (26), since any other covariance matrix would lead asymptotically to a greater P_E below capacity. Thus, Eq. (26) reduces to Eq. (6), and the asymptotic behavior above capacity given by Expressions (2) and (4) must hold also for the best code on this class of channels.

References

1. Wyner, A. D., "On the Probability of Error for Communication in White Gaussian Noise," *IEEE Trans. on Inform. Theory*, Vol. IT-13, pp. 86–90, Jan. 1967.
2. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*. John Wiley & Sons, Inc., New York, 1965.
3. Cramér, H., *Mathematical Methods of Statistics*, pp. 374–376. Princeton University Press, Princeton, N. J., 1946.
4. Balakrishnan, A. V., "A Contribution to the Sphere-Packing Problem of Communication Theory," *J. Math. Analysis and Appl.*, Vol. 3, No. 3, pp. 485–506, Dec. 1961.
5. Helstrom, C. W., Editor's Note, *IEEE Trans. Inform. Theory*, Vol. IT-14, No. 2, p. 311, Mar. 1968.
6. Shannon, C. E., Gallager, R. G., and Berlekamp, E. R., "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels, I," *Inform. Contr.*, Vol. 10, No. 1, pp. 81–83, Jan. 1967.
7. Loève, M., *Probability Theory*, p. 288. Van Nostrand, New York, 1965.