XX. Communications Systems Research

TELECOMMUNICATIONS DIVISION

A. Sequential Decoding: Short Constraint Length Convolutional Codes, *J. A. Heller*

1. Introduction

Much of the past interest in convolutional codes has been in the use of these codes with sequential coding. There the problem of finding "good" codes is not pressing, because arbitrarily small error probabilities can (albeit at the expense of some erasure probability) be had by using almost any code of sufficiently long constraint length. The complexity of a sequential decoder is not very sensitive to constraint length.

Recently, Viterbi (Ref. 1) described a decoder for convolutional codes which was subsequently shown to be maximum likelihood (Ref. 2). The number of operations required by the Viterbi algorithm per decoded bit is a constant for any given constraint length K and increases exponentially with K. This contrasts with brute force maximum likelihood decoding where the decoder complexity goes up exponentially with block length which is typically many times K.

The Viterbi algorithm, while not practical for large K, may perhaps be useful in the range $K = 4$ to 10. Since it is limited to small K, finding good short constraint length codes becomes important. The Viterbi algorithm has several advantages over sequential decoding for small K. First, the time to decode is fixed; hence buffering of the received data is unnecessary. Second, decoding progresses at the received data rate; there is no need for the decoder to be 10 to 20 times faster than the incoming data rate as is necessary in sequential decoding to combat buffer overflow. Third, resynchronization by inserting a known single-constraint-length-long sequence into the data stream periodically is unnecessary. The most important result is that for systems giving a bit error probability of about 5×10^{-3}, such as prevails in the *Mariner* Mars 1969 high rate telemetry system, convolutional codes with short constraint length decoded by the Viterbi algorithm yield a gain of about 1 dB over the use of biorthogonal codes. Also, with the Viterbi algorithm there are no erasures. And the bandwidth expansion is not high. Hence, the use of short constraint convolutional codes is very attractive in future telemetry systems.

2. Properties of Convolutional Codes

Figure 1 shows a typical binary (nonsystematic) convolutional coder. The code rate is $R_N = 1/v$. In Figure 1 $R_N = \frac{1}{3}$ bits/code symbol. The code itself is determined by the connections between the shift register stages and the mod-2 adders. It is usually specified by v connection

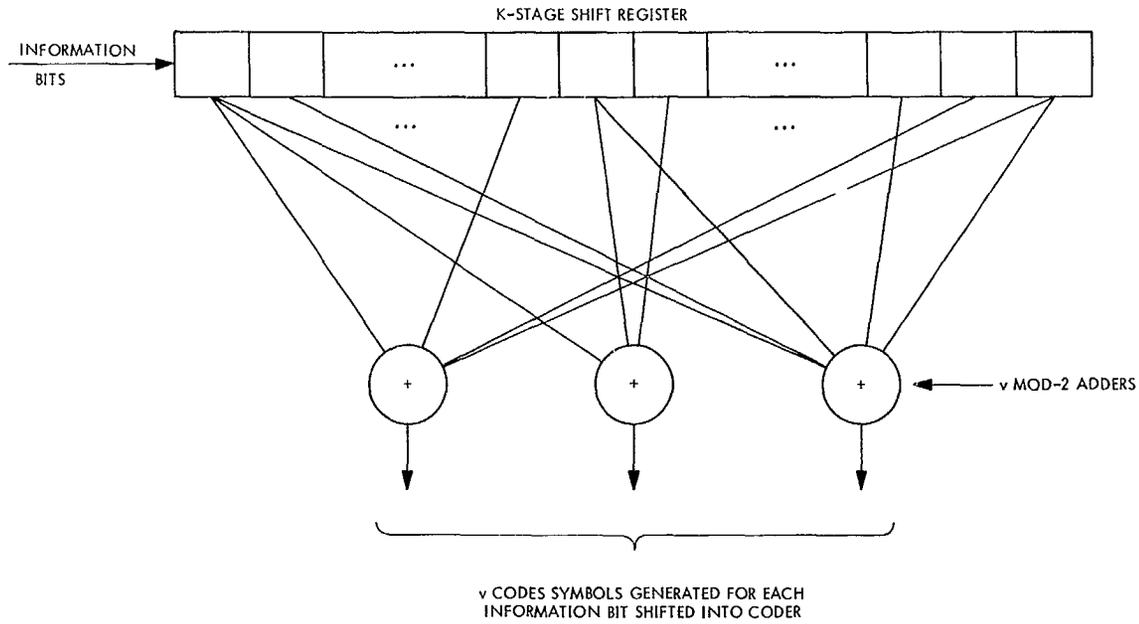


Fig. 1. Binary convolutional coder

vectors, one for each adder, $g_1, g_2, \cdots, g_v$. The jth component of g_i is a 1 if stage j of the shift register is connected to the ith adder, otherwise it is 0. A more compact method of specifying the code is a Kv component vector f which contains the components of the connection vectors interleaved, that is,

$$f_1 = g_{11},\ f_2 = g_{21},\ \cdots,\ f_v = g_{v1}$$

$$f_{v+1} = g_{12},\ \cdots,\ f_{Kv} = g_{vK}$$

Convolutional codes are special cases of group codes. The generator matrix, G, of a convolutional code is shown in Fig. 2. The first row of the matrix is the vector f followed by zeros. Each succeeding row is the previous row shifted v places to the right with the vacated elements filled with zeros. The number of rows L in the matrix is the length of the input information stream. It follows that there are $(K + L - 1)\,v$ columns in G; this is the block length of the code. The $(K + L - 1)\,v$ element code word y is related to the L element input sequence x by

$$y = xG \tag{1}$$

The code is thus the set of all 2^L linear combinations of the rows of G.

The problem of finding the best code, in the probability of error sense, for a given K is difficult if not impossible.

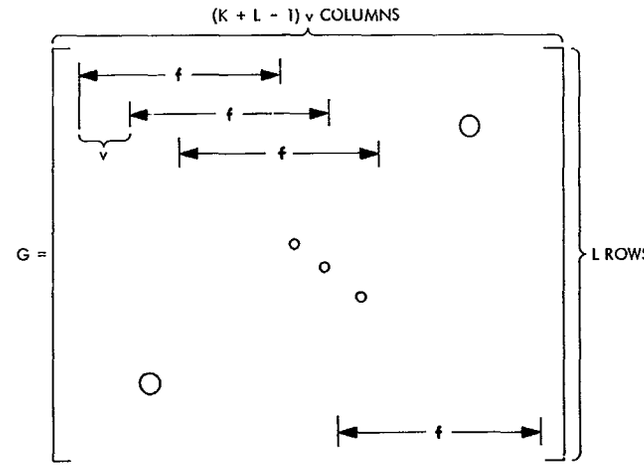


Fig. 2. Convolutional code generator matrix

Other criteria such as maximizing the minimum code word distance do not guarantee that error probability is minimized (this has been shown by simulation). However, low error probability seems at least to be well, if not perfectly, correlated with large minimum distances; that is, a large minimum distance implies that all code word distances are large.

An upper bound on minimum code-word distance for systematic codes has been obtained previously (SPS 37-50, Vol. III, pp. 248–251), and will be stated here for non-systematic codes. Since a linear convolutional code is a

group code, the set of distances from the kth code word to all other code words is independent of k. Thus without loss of generality assume the all-zero code word (corresponding to the all-zero information sequence) is the correct code word. It is a well-known property of group codes that the total number of ones summed over all code words is less than or equal to the total number of zeros (Ref. 3), with equality when no column of G is all zero. The minimum distance must be less than or equal to the average distance, and distance is simply the number of ones in a code word. There are 2^L code words of which $2^L - 1$ are nonzero. The average number of ones in the $2^L - 1$ nonzero code words is thus

$$\frac{2^L}{2^L - 1}\left[\frac{(K + L - 1)\,v}{2}\right]$$

Hence d_m, the minimum distance, is upper bounded by this quantity.

$$d_m \leq \frac{2^L}{2^L - 1}\left[\frac{(K + L - 1)\,v}{2}\right] \tag{2}$$

This bound is true for input sequences of any length L. It is also true for nonzero sequences of length $h < L$. Therefore, a tighter bound can be obtained by minimizing on h

$$d_m \leq \min_h \frac{2^h}{2^h - 1}\left[\frac{(K + h - 1)\,v}{2}\right] \tag{3}$$

It has been shown that the minimum occurs at a value of h that varies as log K (SPS 37-50, Vol. III, pp. 248–251). As a limiting case, for large K, Eq. (3) reduces to

$$d_m \lesssim \frac{Kv}{2} \tag{4}$$

which says that for large K the minimum distance is less than about half a constraint length of code symbols. For $K \lesssim 10$, the optimum h, h_0, will be from 2 to 4. When $h = 1$, we have the bound $d_m \leq Kv$ which can be achieved with a code with 2 code words (G has one row, namely f) with all elements of f equal to 1.

Equation (3) can be considered as a sequence of upper bounds for nonzero input sequences of length $h = 1$ to L (we may take $L = \infty$ since there is no need to segment the data into blocks in the Viterbi algorithm). This suggests the following *ad hoc* scheme which has been useful in generating small K codes with a minimum distance equal to or near the upper bound. Let $d(h)$ be the *actual* minimum distance for a convolutional code of block length h.

Initially choose all Kv elements of the connection vectors, g_i, equal to 1. Then as previously stated, $d(1) = Kv$. Looking at the generator matrix with two rows; however, it is clear that $d(2) = 2v$, which is the weight of the code word corresponding to the input sequence 11. Now change generator elements from 1 to 0 in such a way that each zero is surrounded by ones. Each time this is done, $d(1)$ will drop by 1, and $d(2)$ will increase by 2.

Continue this until min $\{d(1), d(2)\}$ is maximized. This minimum will then be the integer part of the right side of Eq. (3) for $h = 2$. If $K \lesssim 4$ this will be sufficient to achieve the upper bound of Eq. (3). For K up to about 6 the bound can be achieved simply by trying several permutations of the zeros in the connection vectors. For $K > 6$, h_0 gets larger, and $d(j)$ for $j > 2$ must be considered. The manipulations soon become too unwieldy for hand computation.

As an example of this procedure consider generating a $K = 4$, rate ⅓ ($v = 3$) code. Equation (3) yields $d_m \leq 10$ at $h_0 = 2$. Start out with

$$g_1 = g_2 = g_3 = 1111$$

This yields $d(1) = 12$ and $d(2) = 6$. Now make the following change

$$g_1 = g_2 = 1111, \qquad g_3 = 1011$$

Now $d(1) = 11$, $d(2) = 8$. Finally let

$$g_1 = 1111, \qquad g_2 = 1101, \qquad g_3 = 1011$$

Then we have $d(1) = d(2) = 10$. A simple check ensures that $d(i) \geq 10$, $i > 2$. This code is the $K = 4$ code used in the computer simulation described in the next two sections.

Using the Viterbi algorithm, convolutional codes require 2^{K-1} likelihood comparisons to decode one bit. K-bit binary block codes when encoded and decoded optimally require $(1/K)(2^K - 1)$ comparisons per bit. In the first analysis it is thus meaningful to compare the performance of a K-bit block code with a constraint length K convolutional code.

For the white gaussian noise channel a biorthogonal code is very nearly optimum. As an example, for a 4-bit

biorthogonal code, the minimum distance (in terms of signal energy-to-noise ratio) is $2E_b N_0$, where E_b/N_0 is the energy per bit to noise ratio. There are 14 signals at the minimum distance.

The $K = 4$ convolutional code previously described has $d_m = 10$. Since there are 3 code symbols per bit, the minimum distance in signal energy is $(10/3) E_b/N_0 = 3.33E_b/N_0$. Furthermore there are only 4 code words at this distance at any given decoding step. All other code words are at a greater distance.

3. The Decoding Algorithm

The Viterbi algorithm is well documented (Refs. 1 and 2), and the details will not be repeated here. Suffice it to say that operating on an infinite stream of input data the algorithm never makes final bit decisions. A step in the decoding process consists of generating the likelihoods for 2^K possible bit streams and narrowing this down to 2^{K-1} using 2^{K-1} pairwise comparisons. Each comparison is in many ways similar to a "look forward" step in sequential decoding. The comparisons can be done in parallel. In practice the decoder makes bit decisions after a delay of several constraint lengths. The algorithm seems in the first analysis to be quite amendable to special-purpose machine implementation. The major operations in a step are "add," "compare," "shift," and "register exchange." The principal memory required is 2^{K-1} shift registers of length about 3 to 5 constraint lengths.

As far as synchronization is concerned, the decoder needs only branch synchronization, i.e., it must know which code symbols corresponds to the first adder output in the coder. Since the decoder retains bit stream outputs corresponding to all coder states, the decoder will with high probability synchronize within several constraint lengths when started in an arbitrary place in the received data stream.

4. Experimental Results

The Viterbi decoder was simulated on a SDS 930 computer. Received data was generated in 1000-bit blocks, and the decoder was started unsynchronized at the beginning of each of these blocks. The codes used were rate ½, and the channel simulated was a binary input 8-level quantized output white gaussian noise channel. The quantization levels were chosen as in Ref. 4. After allowing 50 bits for resynchronization, the counting of decoding errors was started. Each of the 2^{K-1} decoder output bit streams was 48 bits long. After a decoding step, a decision was made on the oldest bit in the bit stream, by choosing the one corresponding to the most likely stream.

Three codes with different constraint lengths were tested for several values of E_b/N_0. At least 200,000 bits were decoded for each point. The bit error probabilities for the $K = 4$, 6, and 8 codes are shown in Figs. 3, 4, and 5, respectively. The code generators used are also shown in the figures. The $K = 4$, $d_m = 10$ code has already been described. For $K = 6$, Eq. (3) yields an upper bound of 13 on d_m. This d_m is actually achieved with the code shown. This code was also generated using the *ad hoc* scheme and some trial and error. Later, a search program was written to generate all $K = 6$, $v = 3$ codes with $d_m = 13$. Of several tested, the one shown in Fig. 4 proved the best.

At $K = 8$, the bound yields $d_m \leqq 17$. No $d_m = 17$ code was found by hand calculation, and a machine search has not yet been tried. The code used has $d_m = 16$ and is by no means necessarily the best $K = 8$ code. It was, in fact, the only one tested.

Also shown in Figs. 3, 4, and 5 are the corresponding K-bit biorthogonal code bit error probabilities (Ref. 5).

5. Conclusions

The error probability curves of Figs. 3, 4, and 5 clearly indicate that constraint length K convolutional codes perform considerably better than K-bit biorthogonal codes. For instance at $K = 6$ and $P_E = 10^{-3}$ about 1 dB is to be saved. There is also the fact that bandwidth and energy per code symbol required by convolutional codes is independent of K while for biorthogonal codes bandwidth goes as $2^{K-1}/K$ and energy per symbol as $K2^{-(K-1)}$. Included in Fig. 3 is the bit error probability curve for a $K = 7$ biorthogonal code. Observe that a $K = 4$ convolutional code is superior to a $K = 7$ biorthogonal code.

References

1. Viterbi, A. J., "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Trans. Inf. Theory*, IT-11, January 1965.
2. *Final Report on a Coding System Design for Advanced Solar Missions*, NASA Contract NAS2-3637. Codex Corp., Watertown, Mass., December 1967.
3. Peterson, W. W., *Error-Correcting Codes*. M.I.T. Press and John Wiley & Sons, Inc., New York, 1961.
4. Jacobs, I. M., "Sequential Decoding for Efficient Communication From Deep Space," *IEEE Trans. Commun. Tech.*, COM-15, August 1967.
5. Golomb, S. W., et al., *Digital Communications With Space Applications*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

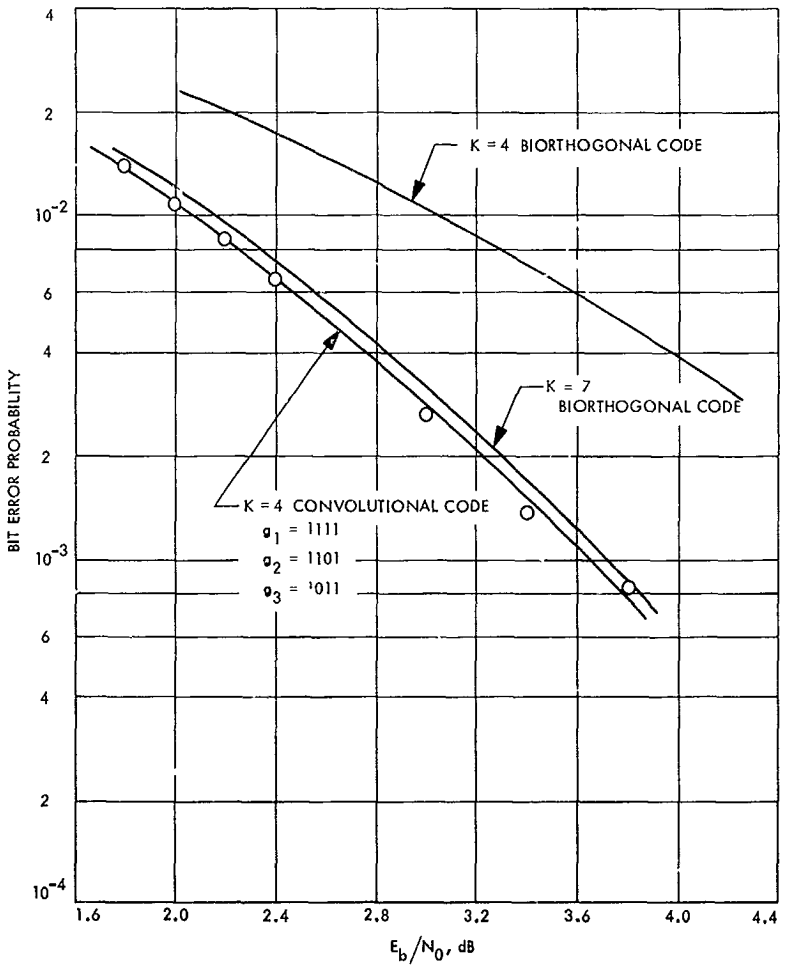


Fig. 3. Bit error probabilities for $K = 4$ codes and a comparison with a $K = 7$ biorthogonal code

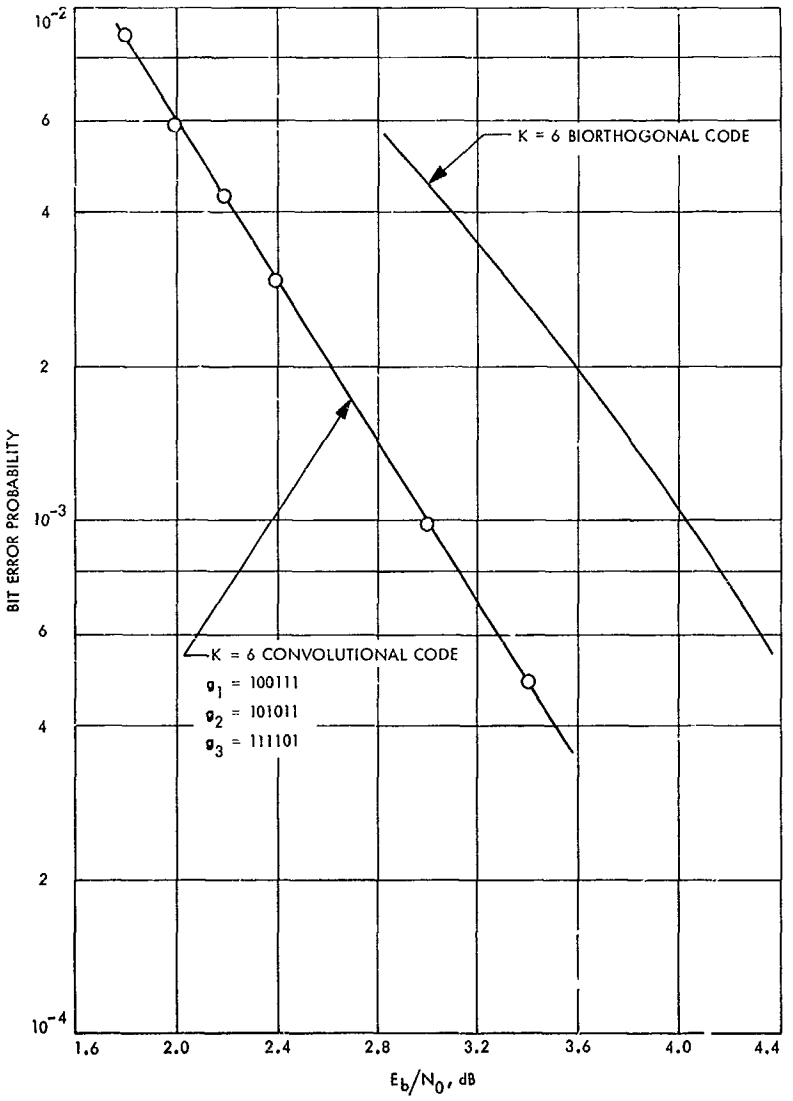


Fig. 4. Bit error probabilities for $K = 6$ codes

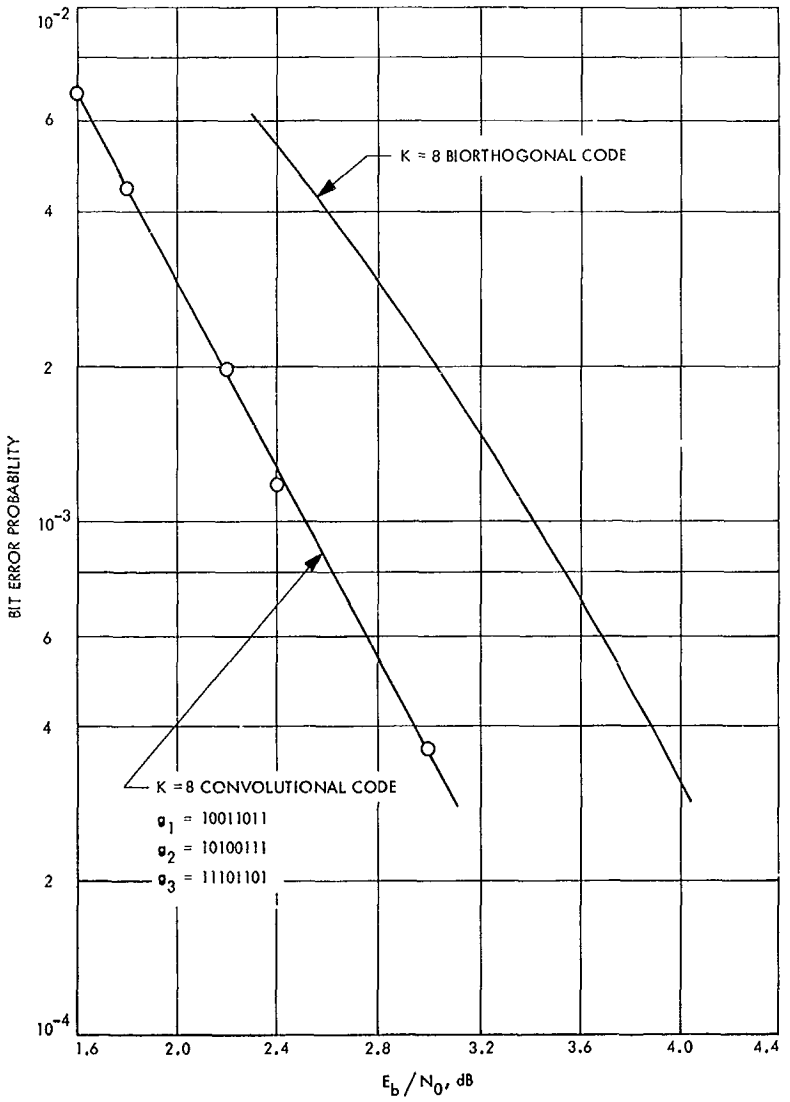


Fig. 5. Bit error probabilities for $K = 8$ codes

B. Coding and Synchronization Studies: Dynamics of Second-Order Phase-Locked Loops,

W. C. *Lindsey*[1]

1. Introduction

This article discusses some recent results obtained relative to the statistical dynamics of the phase-error process in second-order phase-locked loops.[2] Briefly, the problem can be described as follows: The phase-error process $\tilde{\phi}(t)$ in a phase-locked loop system undergoes diffusion much like a particle in Brownian motion (Ref. 1); hence, the variance of the phase-error becomes infinite in the steady state. Previous work (Refs. 2, 3, and 4) on determining the probability distribution of the phase-error in the steady state of a first-order loop was accomplished by reducing the phase-error modulo 2π to a $\phi(t)$ process. Such a reduction ignores how many cycles have slipped with the passage of time so that diffusion of the phase-error is taking place. For finding telemetry error probabilities, that is all that is needed; for estimating tracking accuracy, however, the $\tilde{\phi}(t)$ process itself must be studied. To completely describe the $\tilde{\phi}(t)$ process, one must account for that component of the variance of the phase-error which causes diffusion, i.e., cycle slipping.

[1]Consultant, Electrical Engineering Dept., University of Southern California.

[2]Lindsey, W. C., "Nonlinear Analysis and Synthesis of Generalized Tracking Systems," *Proc. IEEE* (in press). Also USC EE317, University of Southern California, Los Angeles, Calif., Dec. 1968.

This is perhaps best described by evaluating the diffusion coefficient, i.e., the rate with which the variance of the phase-error is approaching infinity, and in some way combine this component of the variance with the variance of the phase-error reduced modulo 2π.

For a second-order loop, this article discusses an approximate solution for the steady-state distribution of the phase-error reduced modulo 2π and establishes the regions of validity of this solution by comparing it with experimental results. Also, formulas are presented for (1) the diffusion coefficient of the phase-error process, (2) the expected values of the time intervals between phase jumps, (3) the expected number of phase jumps per unit time, (4) the expected number of phase jumps "to the right" and "to the left," (5) the expected value of the phase-error rate in the steady state, and (6) the mean-squared value of the phase-error rate in the steady state. In the limit, as the system damping approaches infinity, the results are then valid for the first-order loop.

2. Loop Model and Phase-Error Density Reduced Modulo 2π

For a phase-locked loop system with loop filter,

$$F(p) = \frac{1 + \tau_2 p}{1 + \tau_1 p} \tag{1}$$

it has been recently shown (Footnote 2) that the probability distribution of the phase-error reduced modulo 2π is given to a good approximation by

$$p(\phi) = \frac{\exp[\beta\phi + \alpha\cos\phi]}{4\pi^2 \exp[-\pi\beta]\,|I_{j\beta}(\alpha)|^2} \times \int_{\phi}^{\phi+2\pi} \exp[-\beta x - \alpha\cos x]\,dx \tag{2}$$

where $I_v(x)$ is the imaginary Bessel function of order v and argument x and ϕ belongs to any interval of width 2π centered about any lock point $2n\pi$; n is any integer. The parameters α and β, which characterize Eq. (2), are related to the various system parameters through

$$\left.\begin{aligned} F_1 &= \tau_2/\tau_1 \\ \beta &= \Omega_0 - AK(1 - F_1)\overline{\sin\phi} \\ \alpha &= \left(\frac{r+1}{r}\right)\rho - \frac{1}{r\sigma_G^2} \end{aligned}\right\} \tag{3}$$

where $G = \overline{\sin\phi} - \sin\phi$, the overbar denotes statistical average, Ω_0 represents the amount of loop detuning or frequency offset, ρ is the signal-to-noise ratio in the loop bandwidth, and

$$w_L \approx \frac{r+1}{2\tau_2}, \qquad \text{if } r\tau_1 >> \tau_2$$

The parameter

$$r = AK\tau_2^2/\tau_1 = (\zeta/4)^{1/2}$$

where ζ is the "loop damping" and AK represents open loop gain. Moments of $\sin\phi$ (Footnote 2) are given by

$$\left.\begin{aligned} \overline{\sin\phi} &= \frac{\beta}{\alpha} - \frac{\sinh\pi\beta}{\pi\alpha}\,|I_{j\beta}(\alpha)|^{-2} \\ \overline{\sin^2\phi} &= \frac{1}{2\alpha}\,Re\left[\frac{2I'_{j\beta}(\alpha)}{I_{j\beta}(\alpha)}\right] - \frac{\beta}{\alpha}\left[2\,\overline{\sin\phi} - \frac{\beta}{\alpha}\right] \\ \overline{\cos\phi} &= Re\left[\frac{I_{j\beta+1}(\alpha) + I_{j\beta-1}(\alpha)}{2I_{j\beta}(\alpha)}\right] \end{aligned}\right\} \tag{4}$$

with

$$\sigma^2_{\sin\phi} = \overline{\sin^2\phi} - [\overline{\sin\phi}]^2 \tag{5}$$

and $Re[\cdot]$ denotes the real part of the quantity in the brackets. It is clear from Eqs. (2) and (4) that $p(\phi)$ will be symmetric when the loop is designed such that $\beta = 0$. It is easily shown that in the limit, as r approaches infinity, Eq. (2) reduces to a result given by V. I. Tikhonov (Refs. 2 and 3) for a first-order loop.

The expected value of the phase-error can be found from Eq. (2) and the well-known Bessel function expansions of $\exp(\pm x\cos\phi)$. Without going into details (Footnote 2), we have

$$\left.\begin{aligned} \overline{\phi} &= \int_{-\pi}^{\pi} \phi p(\phi)\,d\phi \\ \overline{\phi} &= \frac{2\sinh\pi\beta}{\pi|I_{j\beta}(\alpha)|^2} \sum_{m=1}^{\infty} \frac{m I_m(\alpha)}{m^2 + \beta^2} \\ &\quad \times \left[\frac{I_0(\alpha)}{\beta} + \frac{I_m(\alpha)}{4m} + \sum_{\substack{k=1 \\ k\neq m}}^{\infty} \frac{2m(-1)^k I_k(\alpha)}{m^2 - k^2}\right] \end{aligned}\right\} \tag{6}$$

It is clear from Eq. (6) that with $\beta = 0$, $\bar{\phi} = 0$. Furthermore, $\overline{\phi^2}$ is given (Footnote 2) by

$$\overline{\phi^2} = \int_{\pi}^{\pi} \phi^2 p(\phi)\, \phi$$

$$= \frac{\sinh \pi\beta}{\pi |I_{j\beta}(\alpha)|^2} \left\{ \frac{I_0(\alpha)}{\beta} \left[\frac{\pi^2 I_0(\alpha)}{3} + 4 \sum_{k=1}^{\infty} \frac{(-1)^k I_k(\alpha)}{k^2} \right] + 2\beta I_0(\alpha) \sum_{k=1}^{\infty} \frac{I_k(\alpha)}{k^2(\beta^2 + k^2)} \right.$$

$$\left. + 2\beta \sum_{k=1}^{\infty} \frac{(-1)^k I_k(\alpha)}{\beta^2 + k^2} \left[\left(\frac{\pi^2}{3} + \frac{1}{2k^2} \right) I_k(\alpha) + 4 \sum_{\substack{m=1 \\ m \neq k}}^{\infty} \frac{(-1)^m (k^2 + m^2) I_m(\alpha)}{(k^2 - m^2)^2} \right] \right\} \tag{7}$$

The variance $\sigma_{\tilde{\phi}}^2 = \overline{\tilde{\phi}^2} - (\bar{\tilde{\phi}})^2$ is minimized when the loop is designed such that $\beta = 0$ and α is maximized (Footnote 2). For this case, from Eqs. (6) and (7), we have

$$\sigma_{\tilde{\phi}}^2 = \frac{\pi^2}{3} + \frac{4}{I_0(\alpha)} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} I_k(\alpha) \tag{8}$$

Finally, the expected value of the phase-error rate $\dot{\phi}$ and the mean-squared value (Footnote 2) are given, respectively, by

$$\left.\begin{aligned} \bar{\dot{\phi}} &= \Omega_0 - AK\overline{\sin\phi} \\ \overline{\dot{\phi}^2} &= K_{00}^2 \left[\alpha^2 \overline{(\sin\phi)^2} + \beta^2 - 2\alpha\beta\, \overline{\sin\phi} \right] \\ \sigma_{\dot{\phi}}^2 &= K_{00}^2 \alpha^2 \sigma_{\sin\phi} + 2K_{00}(1 - K_{00}\beta^2) \end{aligned}\right\} \tag{9}$$

where

$$K_{00} = \frac{\rho(r+1)^2}{2w_L r^2}$$

The above equation, which relates to $\sigma_{\dot{\phi}}$ to $\sigma_{\sin\phi}$, may be interpreted as an "uncertainty hyperbola."

Figure 6 illustrates a plot of the variance of the phase-error $\sigma_{\tilde{\phi}}^2$ reduced modulo 2π for various values of r with $\beta = 0$. For purposes of checking the validity of the approximation which lead to Eq. (2), we have plotted various values of variance of the phase-error obtained by direct measurement in the laboratory (Ref. 5). From Fig. 6 it is clear that, for most practical purposes, Eq. (2) characterizes the distribution $p(\phi)$ for all $\rho > 0$ dB when $r = 2$, $\Omega_0 = 0$. The larger the value of r, the better the approximation.

3. Statistical Dynamics of the Phase-Error Process $\tilde{\phi}(t)$

One method of accounting for the fact that the loop actually skips cycles is to evaluate the rate at which the actual process at work in the loop undergoes diffusion. This parameter, the so-called diffusion coefficient D (Footnote 2 and Ref. 1), is $(2\pi)^2$ times the total average number $\bar{S}$ of phase jumps per unit of time has been shown (Footnote 2) to be given by

$$D = \frac{4w_L r^2}{\rho(r+1)^2} \frac{\cosh \pi\beta}{|I_{j\beta}(\alpha)|^2} = (2\pi)^2 \bar{S} \tag{10}$$

In *Section 4*, we shall see that the variance of the phase error $\tilde{\phi}(t)$ at time t starting from zero error at time $t = 0$ is

$$V = \overline{\tilde{\phi}^2} + Dt \tag{11}$$

since the cross term $E[4\pi\phi(t)\, k(t)]$ is essentially zero. This last fact follows from the fact that $E[\phi(t)] = 0$ while $\tilde{\phi}(t)$ is essentially independent of $k(t)$. Here, $k(t)$ is the unique integer with

$$\tilde{\phi}(t) = \phi(t) + 2\pi k(t)$$

Figure 7 illustrates a plot of the normalized diffusion coefficient for various values of ρ and r with $\beta = 0$. From this figure, it is seen that diffusion will not appreciably effect the measurement in a finite time if $\rho > 3$. Via Eq. (11), Eqs. (6) and (10) can be used to account for the effects in which cycle-slipping produces errors in Doppler measurements.

The expected time interval between successive cycle slipping events is given (Footnote 2) by

$$\Delta T = \frac{\pi^2 \rho(r+1)^2}{r^2 w_L} \frac{|I_{j\beta}(\alpha)|^2}{\cosh \pi\beta}$$

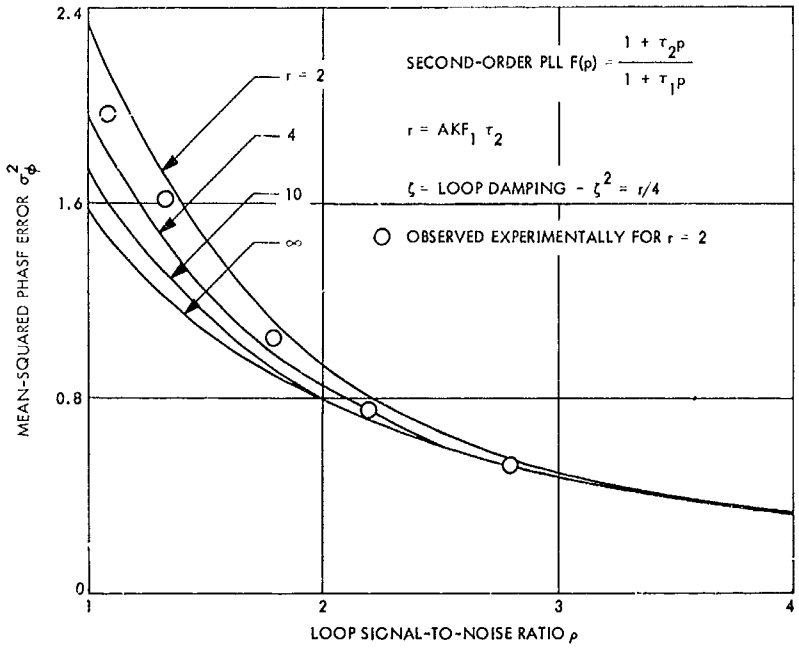


Fig. 6. Variance of the phase-error versus loop signal-to-noise ratio ρ for various values of r

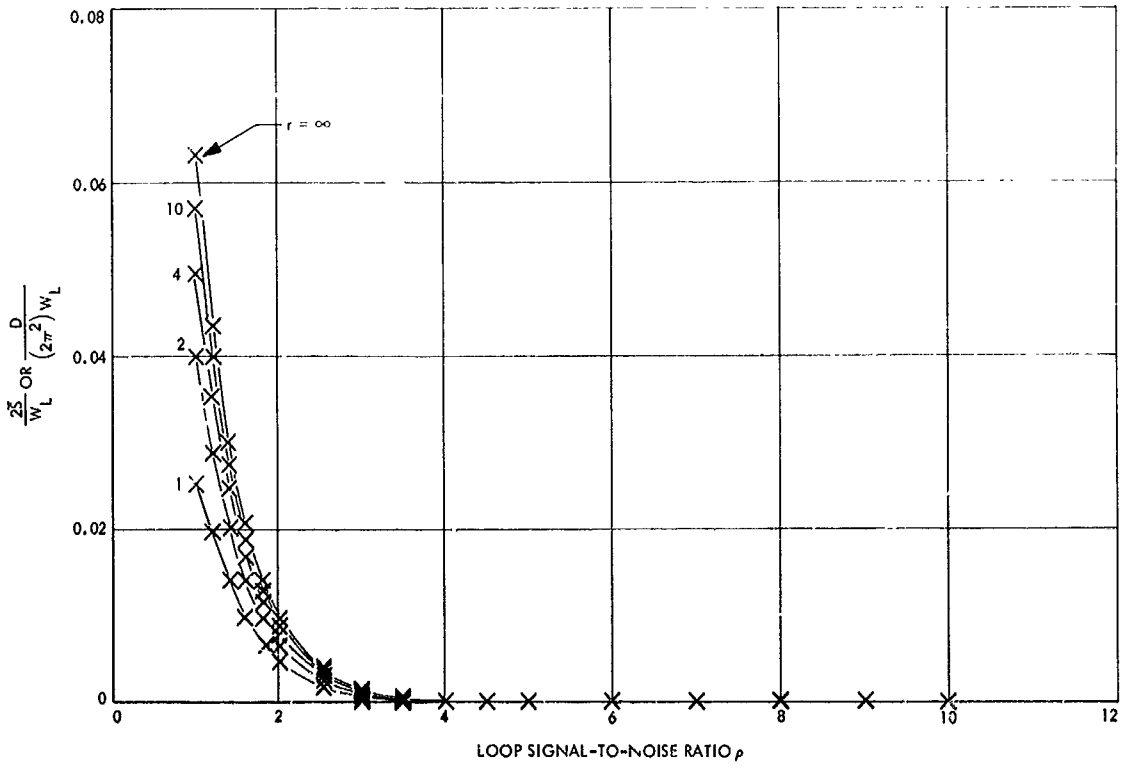


Fig. 7. Diffusion coefficient

where it can be and has been assumed that the successive instants in time when phase synchronization is lost are statistically independent.

In the case where $\Omega_0 \neq 0$, the case of greatest practical interest, the average number of phase jumps "to-the-right," say I, is of interest. Formulas for I_+ and I are given (Footnote 2) by

$$I_\pm = \frac{\exp(\pm\beta\pi) I}{2\sinh\pi\beta} \tag{12}$$

where I is the (net) average number of phase jumps per unit time; i.e.,

$$I = I_+ - I_- = \left[\frac{\rho(r+1)^2}{\omega_L (r\pi)^2}\right]^{-1} \frac{\sinh\pi\beta}{|I_{j\beta}(\alpha)|^2} \tag{13}$$

Both I_+ and I_- are zero for the unstressed loop with $\Omega_0 = 0$. The quantity I is related to the total expected number of phase jumps per unit time $\bar{S}$ through the equation

$$I = \frac{\sinh\pi\beta}{\cosh\pi\beta}\bar{S} = (\tanh\pi\beta)\bar{S} \tag{14}$$

4. Statistical Dynamics of the Phase-Error Process $\tilde{\phi}(t)$

The actual loop phase error $\tilde{\phi}(t)$ is related to the reduced modulo 2π process $\phi(t)$ through

$$\tilde{\phi}(t) = 2\pi k + \phi(t) \tag{15}$$

where

$$k = \left[\frac{\tilde{\phi}(t) - \phi(t)}{2\pi}\right] \tag{16}$$

is the largest integer which does not exceed the bracketed ratio. Thus, k is a discrete random variable which takes on integer values at random points in time. The mean-squared value of the $\tilde{\phi}(t)$ process is given by

$$\sigma^2_{\tilde{\phi}}(t) = \sigma^2_{\phi}(t) + (2\pi)^2\sigma^2_k + 2\pi[\overline{k\phi} - \bar{k}\bar{\phi}] \tag{17}$$

The event that k cycles are slipped in t seconds will be denoted by $\mathcal{J}$. If we assume that $\mathcal{J}$ is a Poisson-type process, then the quantity $\bar{S}$, representing the total average number of phase jumps per unit time, can be used to produce a probabilistic model for the phase jumping process that causes diffusion of the phase error process $\tilde{\phi}(t)$. Namely,

$$P(\mathcal{J}) = P(k) = \frac{(\bar{S}t)^k \exp[\bar{S}t]}{k!} \tag{18}$$

The mean of k is

$$\bar{k} = \bar{S}t \tag{19}$$

and the variance is also given by

$$\sigma^2_k = \overline{k^2} = \sigma^2_k = \bar{S}t \tag{20}$$

Experimental justification that supports the Poisson assumption is given in Ref. 5 and SPS 37-43, Vol. III, pp. 76–80. Thus, the mean-squared value of $\tilde{\phi}(t)$ becomes

$$\sigma^2_{\tilde{\phi}}(t) = \sigma^2_{\phi}(t) + (2\pi)^2\bar{S}t + 2\pi[\overline{k\phi} - \bar{k}\bar{\phi}] \tag{21}$$

and in the steady state, $\tilde{\phi}(t)$ has infinite variance. Now, $D = (2\pi)^2\bar{S}$ represents the diffusion coefficient of the $\tilde{\phi}(t)$ process, i.e., the rate at which $\tilde{\phi}(t)$ is undergoing diffusion. Furthermore, if the random variables k and ϕ are independent, as is reasonable, we can write

$$\lim_{t\to\infty} \Delta\sigma^2_{\tilde{\phi}} = \sigma^2_{\tilde{\phi}}(t+T) - \sigma^2_{\tilde{\phi}}(t) = DT \tag{22}$$

Finally, the probability of loosing phase-lock in t seconds (i.e., the probability of slipping one or more cycles) is given by

$$P[\mathcal{J} \geq 1] = 1 - \exp[-\bar{S}t] \tag{23}$$

This result should prove useful in the design of phase-coherent doppler tracking systems.

References

1. Wax, N., *Selected Papers on Noise and Stochastic Processes*, Dover Publications, Inc., N. Y., 1954.
2. Tikhonov, V. I., "The Effects of Noise on a Phase-Lock Oscillation Operation," *Automat. Rem. Cont.*, Vol. 22, No. 9, 1959.
3. Tikhonov, V. I., "Phase-Lock Automatic Frequency Control Application in the Presence of Noise," *Automat. Rem. Cont.*, Vol. 23, No. 3, 1960.
4. Viterbi, A. J., "Phase-Locked Loop Dynamics in the Presence of Noise by Fokker-Planck Techniques," *Proc. IEEE*, Vol. 51, No. 12, pp. 1737–1753, Dec. 1963.
5. Charles, F. J., and Lindsey, W. C., "Some Analytical and Experimental Phase-Locked Loop Results for Low Signal-to-Noise Ratios," *Proc. IEEE*, Vol. 55, No. 9, pp. 1152–1166, Sept. 1966.

C. Coding and Synchronization Studies: On Solution to the Second-Order Phase-Locked Loop, *J. K. Holmes*

1. Introduction

Understanding of phase-locked loop (PLL) operation began in 1955 with the now classical paper by R. M. Jaffe and E. Rechtin (Ref. 1). Starting with their basic linear analysis, there have been a wealth of papers concerning extensions, refinements, applications, and new approaches in the analysis of PLLs, as well as other automatic phase control devices.

This present article is concerned with computational simplification and an elaboration of the analysis initiated in a previous article (SPS 37-49, Vol. III, pp. 297–300), in which an expression for the variance and probability density function of the phase error was obtained for a second-order PLL. The solution was simplified with the introduction of a mean-square fit approach to approximate a certain unknown conditional mean. This idea, along with an extension and generalization of a technique due to A. J. Viterbi (Ref. 2), and some aspects of R. C. Tausworthe's "linear-spectral theory" (Ref. 3), allowed the development of a good approximation to the stationary phase-error variance reduced modulo 2π.

Recently, W. C. Lindsey,[3] using the mean-square approach introduced in SPS 37-49, Vol. III, has developed a generalization to Nth order loops and has considered many other aspects associated with the nonstationary case. An additional list of references on the subject of PLLs can be found in Refs. 2, 3, and 4.

[3]Lindsey, W. C., "Nonlinear Analysis and Synthesis of Generalized Tracking Systems," *Proc. IEEE* (in press).

2. Derivation

The equation of operation, for zero initial detuning and no modulation, is given in operator form by (Ref. 3)

$$\phi(t) + AK\frac{F(s)}{s}\sin\phi(t) = -K\frac{F(s)}{s}n(t) \tag{1}$$

where s is the Heaviside operator d/dt, $\phi(t)$ is the phase-error $\theta(t) - \hat{\theta}(t)$, and $K = k_1 k_m k_{vco}$. The PLL model under consideration is shown in Fig. 8. For the practical integrator, we have the following loop filter:

$$F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s} \qquad \tau_1, \tau_2 > 0 \tag{2}$$

where τ_1 and τ_2 are real nonnegative parameters. Substituting Eq. (2) into Eq. (1) yields

$$s(1 + \tau_1 s)\phi(t) + AK(1 + \tau_2 s)\sin\phi(t) = -K(1 + \tau_2 s)n(t) \tag{3}$$

To avoid obtaining an equation involving the derivative of white noise, let

$$\phi(t) = \tau_2 \dot{u}(t) + u(t)$$

Then Eq. (3) is equivalent to the third-degree equation

$$\left(1 + \tau_2\frac{d}{dt}\right)(\tau_1\dddot{u} + \ddot{u} + AK\sin[\tau_2\dot{u} + u] + Kn(t)) = 0 \tag{4a}$$

However, it is sufficient, as far as the solution for $\phi(t)$ is concerned, to solve

$$\mathcal{L}[u] \triangleq \tau_1\dddot{u} + \ddot{u} + AK\sin[\tau_2\dot{u} + u] = -Kn(t) \tag{4b}$$

as long as $\phi(0) = \dot{\phi}(0) = 0$ and $\mathcal{L}[u(0)] = -Kn(0)$.

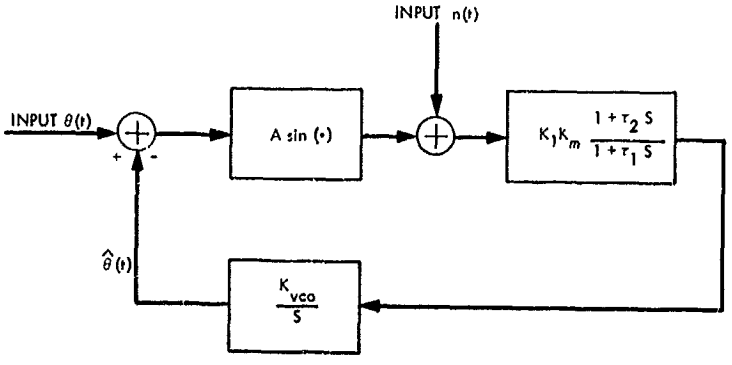


Fig. 8. Mathematically equivalent model for the second-order phase-locked loop

We can now form a Markov vector process in $\mathbf{y}(t)$ where the components are $y_0(t) = u(t)$ and $y_1(t) = \dot{u}(t)$. We have, using Eq. (4a),

$$\left.\begin{aligned} \frac{dy_0}{dt} &= y_1 \\ \frac{dy_1}{dt} &= -\frac{AK}{\tau_1}\sin(\tau_2 y_1 + y_0) - \frac{K}{\tau_1}n(t) - \frac{1}{\tau_1}y_1(t) \end{aligned}\right\} \tag{5}$$

The Fokker–Planck equation to this system of equations can be shown to be (SPS 37-49, Vol. III)

$$\frac{\partial p}{\partial t} = -y_1\frac{\partial p}{\partial y_0} + \frac{\partial}{\partial y_1}\left\{\left[\frac{AK}{\tau_1}\sin(\tau_2 y_1 + y_0)\right]p\right\} + \frac{1}{\tau_1}\frac{\partial}{\partial y_1}(y_1 p) + \frac{K^2 N_0}{4\tau_1^2}\frac{\partial^2}{\partial y_1^2}(p) \tag{6}$$

where we consider the class of solutions that have continuous first-order derivatives in t and y_0 and a continuous second-order derivative in y_1. The notation $p(y_1, y_0, t)$ denotes the joint probability density function of $\mathbf{y}_t = [y_0(t), y_1(t)]$ at time t. In order to obtain an equation in $p(\phi, t)$, we make the following change of variables. Let $z = y_0$ and $\phi = \tau_2 y_1 + y_0$, then we obtain

$$\frac{\partial p}{\partial t}(\phi, z, t) = -\frac{1}{\tau_2}(\phi - z)\left(\frac{\partial p}{\partial z} + \frac{\partial p}{\partial \phi}\right) + \tau_2\frac{\partial}{\partial \phi}\left\{\left[\frac{AK}{\tau_1}\sin\phi + \frac{1}{\tau_1\tau_2}(\phi - z)\right]p\right\} + \frac{K^2 N_0 \tau_2^2}{4\tau_1^2}\frac{\partial^2 p}{\partial \phi^2} \tag{7}$$

This equation appears impossible to solve in terms of known functions. Hence, we are led to obtain an approximation for the density of ϕ. In fact, let us find only the stationary density function with $\partial p/\partial t = 0$ in Eq. (7). First, we integrate out the variable z noting that $p(\phi, z)\big|_{z=\pm\infty} = 0$, which follows from the fact that

$$F(\phi, -\infty, t) = F(\phi, \infty, t) = 0$$

and from the continuity of $p(\phi, z, t)$. We then have

$$\left(\frac{1}{\tau_1} - \frac{1}{\tau_2}\right)\frac{\partial}{\partial \phi}\left(\phi p(\phi) - \int_{-\infty}^{\infty} z p(\phi, z)\, dz\right) + \frac{\tau_2 AK}{\tau_1}\left(\frac{\partial}{\partial \phi}\right)[\sin\phi\, p(\phi)] = 0 \tag{8}$$

Furthermore, since $z = \phi - \tau_2 y_1$, we see that

$$\int_{-\infty}^{\infty} z p(\phi, z)\, dz = p(\phi)\,[\phi - \tau_2 E(y_1 | \phi)]$$

so that Eq. (8) becomes

$$\left(\frac{\tau_2}{\tau_1} - 1\right)\frac{\partial}{\partial \phi}[E(y_1|\phi)\, p(\phi)] + \frac{\tau_2 AK}{\tau_1}[\sin\phi\, p(\phi)] + \frac{K^2 N_0}{4}\left(\frac{\tau_2}{\tau_1}\right)^2\frac{\partial^2 P(\phi)}{\partial \phi^2} = 0 \tag{9}$$

This equation cannot be solved exactly either since $E(y_1|\phi)$ requires the joint distribution function of y_1 and ϕ, which is not available. However, a generalization of the method of Viterbi's (Ref. 2) can be used to obtain a more workable form of this conditional expectation term. Multiply both sides of the differential equation for $y_1(t)$ (Eq. 5) by $\exp(t/\tau_1)$ and integrate from t^+ to a. We obtain

$$y_1(t^+) = y_1(a)\exp\left(\frac{a - t^+}{\tau_1}\right) + \frac{AK}{\tau_1}\int_{t^+}^{a}\exp\left(\frac{t' + t^+}{\tau_1}\right) \times \sin\phi(t')\, dt' + \frac{K}{\tau_1}\int_{t^+}^{a} n(t')\exp\left(\frac{t' - t^+}{\tau_1}\right)dt'$$

Forming $E[(\cdot)|\phi_t]$, and noting

$$E[n(t')|\phi_t, t' > t] = 0$$

we have, letting $t' = t^+ + \tau$,

$$E[y_1(t^+)|\phi_t] = \frac{AK}{\tau_1}\int_0^{a-t^+}\exp\left(\frac{\tau}{\tau_1}\right)E \times [\sin\phi(t^+ + \tau)|\phi(t)]\, d\tau + E\left[y_1(a)\exp\left(\frac{a - t^+}{\tau_1}\right)\middle|\,\phi(t)\right]$$

It can be shown that

$$\lim_{a\to\infty} E\left[y_1(a)\exp\left(\frac{a - t^+}{\tau_1}\right)\middle|\,\phi(t)\right] = 0, \qquad \text{if } \tau_1 > 0 \tag{10}$$

Hence, we have, assuming continuity in the $\phi(t)$ process,

$$E(y_1|\phi) = \frac{AK}{\tau_1}\int_0^{\infty}\exp\left(\frac{\tau}{\tau_1}\right)E[\sin\phi(t + \tau)|\phi(t)]\, d\tau \tag{11}$$

Hence, the evaluation of $E(y_1|\phi)$ has been reduced to evaluating $E[\sin\phi(t + \tau)|\phi(t)]$.

In order to proceed, it is necessary to approximate the conditional expectation term in Eq. (11). First, note that

$$\lim_{\tau \to 0} E\left[\sin \phi(t+\tau) \mid \phi(t)\right] = \sin \phi(t)$$

Secondly, results of simulation studies indicate the following relationship for $\tau > 0$:

$$E\left[\sin \phi(t+\tau) \mid \phi(t)\right] \cong f(\tau) \sin \phi(t)$$

Consequently, we are led to try an approximation of the form

$$E\left[\sin \phi(t+\tau) \mid \phi(t)\right] = C(\tau) \sin \phi(t)$$

A reasonable method of estimating $C(\tau)$ is to minimize the mean-square error between $C(\tau) \sin \phi(t)$ and $E\left[\sin \phi(t+\tau) \mid \phi(t)\right]$. To simplify the notation for the following minimization, let $\phi_2 = \phi(t+\tau)$ and $\phi_1 = \phi(t)$, and denote E_{ϕ_1} as the ensemble average over ϕ_1 and E_{ϕ_2} as the ensemble average over ϕ_2, given ϕ_1. With this notation, we then seek $C(\tau)$ such that

$$E_{\phi_1}\left\{\left[E_{\phi_2}(\sin \phi_2 \mid \phi_1) - C(\tau) \sin \phi_1\right]^2\right\}$$

is a minimum. Performing the minimization, it is found that the optimum $C(\tau)$, say $C_0(\tau)$, is given by

$$C_0(\tau) = \frac{E(\sin \phi_1 \sin \phi_2)}{E(\sin^2 \phi_1)} = \frac{R_{\sin \phi}(\tau)}{R_{\sin \phi}(0)} \tag{12}$$

Hence, with the above mean-square fit to the term $E(\sin \phi_2 \mid \phi_1)$, Eq. (11) becomes

$$E(y_1 \mid \phi) = \frac{AK \sin \phi}{\tau_1 R_{\sin \phi}(0)} \int_0^\infty \exp\left(\frac{\tau}{\tau_1}\right) R_{\sin \phi}(\tau)\, d\tau \tag{13}$$

It might appear that this expression for the conditional mean is divergent; however, as discussed below, the effect of the exponential in the integrand of Eq. (14) is negligible and, in fact, the integral is finite. Since the moments in Eq. (12) cannot be determined exactly, an approximation will be made similar to that used in linear spectral theory (Ref. 3). In this theory, it is assumed that a function $\gamma(s)\,\gamma(-s)$ exists such that

$$\gamma(s)\,\gamma(-s) = \frac{S_{\sin \phi}(s)}{S_\phi(s)}$$

where $S_\phi(s)$ is the spectral density of ϕ, etc. Furthermore, if $F(s)$ is very narrow band, $\gamma(s) = \gamma$ (a constant) and we have

$$\gamma^2 \int_0^\infty R_\phi(\tau)\, d\tau = \int_0^\infty R_{\sin \phi}(\tau)\, d\tau$$

A determination of γ^2 is not necessary since it will be seen later to cancel out. A tedious calculation shows that $R_\phi(\tau)$ is dominated by the exponential term

$$\exp\left(\frac{-r|\tau|}{2\tau_2}\right)$$

where $r = AK\tau_2^2/\tau_1$, and for "normal" design we select the parameters so that

$$\frac{r}{2\tau_2} >> \frac{1}{\tau_1}$$

Consequently, the exponential term in Eq. (13) is negligible in its effect on the integral for normal design. We neglect this term, $\exp(\tau/\tau_1)$, in the ensuing development. At this point then, we have

$$E(y_1 \mid \phi) = \frac{AK\gamma^2 \sin \phi}{\tau_1 R_{\sin \phi}(0)} \int_0^\infty R_\phi(\tau)\, d\tau \tag{14}$$

The final approximation to be made is to assume $\phi(t)$ is a gaussian random process in order to estimate $R_{\sin \phi}(0)$ for all signal-to-noise ratios (SNRs) of interest. (By using the gaussian assumption, we reduce the calculation of the phase-error density and its variance to a simple calculation for all normal parameter values. The correctness of this assumption is borne out by the comparison depicted in Fig. 9.) Then, by Price's Theorem (Ref. 5), we have

$$R_{\sin \phi}(0) = \exp(-\sigma_\phi) \sinh(\sigma_\phi^2) \tag{15}$$

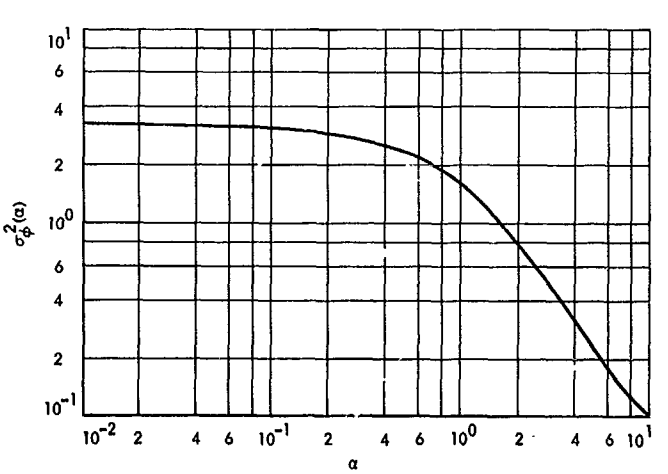


Fig. 9. Loop phase-error variance σ_ϕ^2 versus SNR

Also,

$$\int_0^\infty R_\phi(\tau)\,d\tau = \frac{1}{2}\int_{-\infty}^{\infty} R_\phi(\tau)\,d\tau = \frac{1}{2}S_\phi(0) \tag{16}$$

Now using the spectral equation (Ref. 3), we obtain

$$S_\phi(0) = \frac{N_0}{2A^2\gamma^2}$$

Hence, our approximation for $E(y_1|\phi)$ is finally given by

$$E(y_1|\phi) = \frac{N_0K}{4A\tau_1}\frac{\exp(\sigma_\phi^2)}{\sinh\sigma_\phi^2}\sin\phi \tag{17}$$

and we see that, as promised, our result does not depend on γ^2.

Since our goal is to obtain the density function of ϕ, modulo 2π, we define $P(\phi)$ as the sum of all $p(\phi + 2\pi n)$ and normalize it to integrate to one. Therefore, let

$$P(\phi) = \sum_{k=-\infty}^{\infty} p(\phi + 2\pi k)$$

Then, from Eq. (9), we have for $P(\phi)$

$$\left(\frac{\tau_2}{\tau_1} - 1\right)\frac{d}{d\phi}[E(y_1|\phi)P(\phi)] + \frac{\tau_2 AK}{\tau_1}\frac{d}{d\phi}[\sin\phi\,P(\phi)] + \frac{K^2N_0}{4}\left(\frac{\tau_2}{\tau_1}\right)^2\frac{d^2}{d\phi^2}P(\phi) = 0 \tag{18}$$

From Eqs. (17) and (18) it can be shown that

$$P(\phi) = \frac{\exp\left\{\frac{4A}{KN_0\tau_2}[\tau_1 - (\tau_1 - \tau_2)\beta]\cos\phi\right\}}{2\pi I_0\left\{\frac{4A}{KN_0\tau_2}[\tau_1 - (\tau_1 - \tau_2)\beta]\right\}} \tag{19}$$

where

$$\beta = \frac{N_0\exp(\sigma_\phi^2)}{4A^2\tau_2\sinh\sigma_\phi^2} \tag{20}$$

Subsection 3 demonstrates the method for obtaining σ_ϕ^2 versus $N_0/4A^2$ and, hence, through Eq. (20) the value of β.

In the high SNR case (SNR $\gtrsim$ 10 dB), β is given quite accurately by $\beta = 1/(1 + r)$ and the stationary phase-error probability density for high SNRs becomes

$$P(\phi) = (2\pi)^{-1}\left\{I_0\left[\left(\frac{4A\tau_1}{KN_0\tau_2}\right)\left(\frac{AK\tau_2^2 + \tau_2}{AK\tau_2^2 + \tau_1}\right)\right]\right\}^{-1} \times \exp\left[\left(\frac{4A\tau_1}{KN_0\tau_2}\right)\left(\frac{AK\tau_2^2 + \tau_2}{AK\tau_2^2 + \tau_1}\right)\cos\phi\right] \tag{21}$$

or, letting $r \cong AK\tau_2^2/\tau_1$, and noting $r >> \tau_2/\tau_1$, we have, to a good approximation for high SNRs, the following expression:

$$P(\phi) = (2\pi)^{-1}\left\{I_0\left[\frac{4A^2\tau_2}{N_0(r+1)}\right]\right\}^{-1}\exp\left[\frac{4A^2\tau_2}{N_0(r+1)}\right] \tag{22}$$

As a partial check on our answer, note from Eq. (21) that if $\tau_1 = \tau_2$, then $F(s) = 1$ (i.e., a first-order loop), and $P(\phi)$ is given by

$$P(\phi) = (2\pi)^{-1}\left[I_0\left(\frac{4A}{KN_0}\right)\right]^{-1}\exp\left(\frac{4A}{KN_0}\cos\phi\right) \tag{23}$$

the exact result for the first-order loop (Ref. 2). Further, if we let K and $\tau_1 \to \infty$ such that $K/\tau_1 \to K'$, $F(s)$ becomes $F(s) = (1 + \tau_2 s)/s$ (i.e., the perfect second-order loop), and we obtain in this case

$$P(\phi) = (2\pi)^{-1}\left\{I_0\left[\left(\frac{4A^2}{N_0}\right)\left(\frac{1}{AK\tau_2 + 1/\tau_2}\right)\right]\right\}^{-1} \times \exp\left[\left(\frac{4A^2}{N_0}\right)\left(\frac{1}{AK\tau_2 + 1/\tau_2}\cos\phi\right)\right] \tag{24}$$

which agrees with Viterbi's results (Ref. 2) derived for the high SNR case.

3. Comparison of Results

It is of interest to compare the results of this mean-square fit to the conditional mean (MSFCM) method with that of some experimental data reported by F. G. Charles and W. C. Lindsey (Ref. 4), as well as with some previous results. The PLL used in Ref. 4 had the following relevant parameters:

$$\tau_1 = 45.2$$

$$\tau_2 = 0.125$$

$$AK = 5{,}800$$

$$r = 2.0$$

The density function $P(\phi)$ in Ref. 4, was obtained experimentally and the variance was computed from this.

In order to obtain a numerical value for the variance from the method presented in Ref. 4, a knowledge of $\sigma_\phi^2(\alpha)$ is required where

$$\sigma_\phi^2(\alpha) = \int_{-\pi}^{\pi} \phi^2 \frac{\exp(\alpha \cos\phi)}{2\pi I_0(\alpha)} d\phi \tag{25}$$

This method deviates from that given in SPS 37-49, Vol. III, pp. 297–300, in that a simplification is introduced here that makes the evaluation of the variance a simple matter not requiring the use of a computer.

A graph of $\sigma_\phi^2(\alpha)$ versus α is given in Fig. 10. From Eqs. (19) and (20), assuming $\tau_1 >> \tau_2$, we have

$$\alpha = \left(\frac{4A^2\tau_1}{N_0\tau_2}\right)\left[\left(1 - \frac{N_0}{4A^2} \cdot \frac{2}{1 - e^{-2\sigma_\phi^2}}\right)\right] \tag{26}$$

Hence, once σ_ϕ^2 and α are specified from Fig. 10, $N_0/4A^2$ can be computed from Eq. (26). The SNR A^2/N_0W_L is then given by

$$SNR = \left(\frac{4A^2}{N_0}\right)\left(\frac{\tau_2}{r+1}\right) \tag{27}$$

and, hence, one has a simple method of obtaining the relationship between σ_ϕ^2 and SNR. Using this method, the variance was computed and plotted in Fig. 9. As can be seen from the figure, both the spectral method and the MSFCM method provide relatively close fits. In fact, the MSFCM method yields negligible errors for SNR $\geqq$ 2 dB, which encompasses the useful range of interest. For SNR $\geqq$ 10 dB, the linear spectral method, the MSFCM method, and the linear method, as well as the others, all converge to essentially the same curve.

References

1. Jaffe, R. M., and Rechtin, E., "Design and Performance of Phase-Locked Circuits Capable of Near-Optimum Performance over a Wide Range of Input Signals and Noise Levels," *Trans. PIGIT*, Mar. 1955.
2. Viterbi, A. J., *Principles of Coherent Communication*. McGraw–Hill Book Co., New York, 1966.
3. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers, Vol. I*, Technical Report 32-819. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1966.
4. Charles, F. G., and Lindsey, W. C., "Some Analytical and Experimental Phase Locked Loop Results for Low Signal-to-Noise Ratios," *Proc. IEEE*, Vol. 55, No. 9, pp. 1152–1166, Sept. 1966.
5. Price, R., "A Useful Theorem for Nonlinear Devices Having Gaussian Inputs," *Trans. PIGIT*, IT-4, June 1958.
6. Middleton, D., *Introduction to Statistical Communication Theory*, Sections 1.3 and 1.4-3. McGraw–Hill Book Co., New York, 1960.

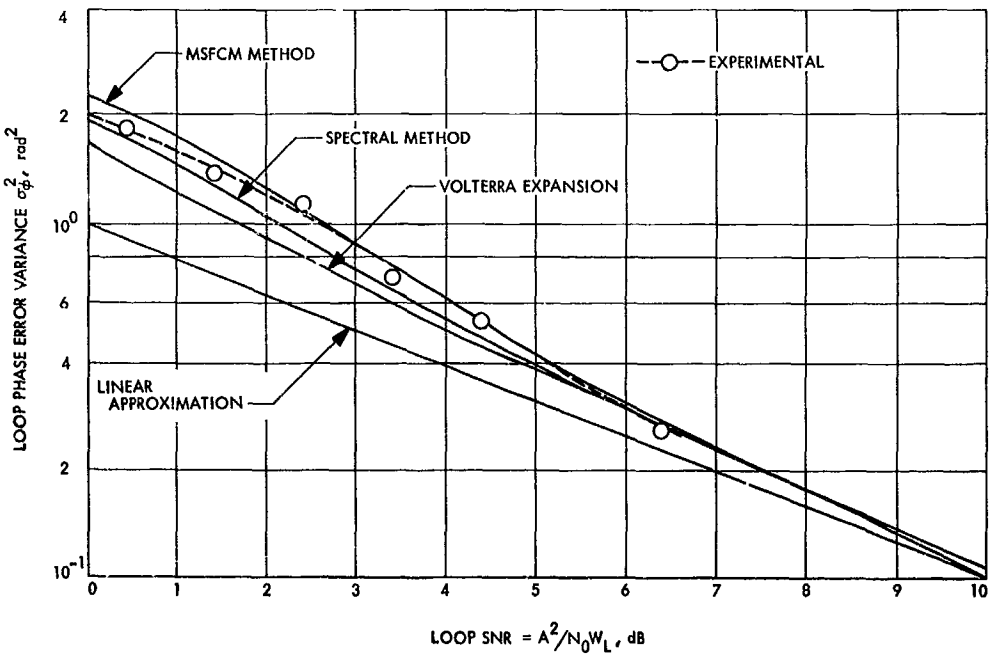


Fig. 10. $\sigma_\phi^2(\alpha)$ versus α

D. Coding and Synchronization Studies: Power Allocation Length Into Rapidly Varying Phase Error, *W. C. Lindsey[1] and J. F. Hayes*

1. Introduction

In a recent paper (Ref. 1) design trends were established for the allocation of power between the carrier and the information-bearing portion of a transmitted signal. Both one-way and two-way phase coherent communication systems were considered. This work was predicated on data rates high enough that any error in estimating the phase of the radio-frequency carrier could be taken to be constant over the duration of one signaling interval. In the present work the same problem is considered; however, the assumption that the phase error is constant over a symbol time is removed and the phase error is allowed to change appreciably over the signaling interval. In both works it is assumed that reference phase is derived exclusively from the carrier. Alternate approaches that envision deriving the reference phase from the data portion of the transmitted signal lead to quite different results (Ref. 2).

Rapidly varying phase errors are known to affect the operation of command and telemetry systems in deep-space communication systems. Consequently, a theory for predicting their deleterious effects on system performance is needed when carrying out a particular design or in the testing of a fabricated system to see that system performance meets specification. In this article we draw heavily upon the notation introduced in Ref. 1. We also assume that the subcarrier and bit-sync jitter produce negligible effects on the detection process. Procedures for handling these effects, if appreciable, have been reported in Ref. 3.

2. System Model

We consider binary phase-shift-keyed signaling for which the transmitted signals are of the form

$$\rho(t) = (2P)^{1/2} \sin\left[\omega t + (\cos^{-1} m)\, x_k(t)\right], \qquad k = 1, 2 \tag{1}$$

where $x_k = \pm 1$, P is the total transmitted power, and m is the modulation factor which serves to apportion power between the carrier and the side-bands. The signal $\rho(t)$ is disturbed in the channel by additive white gaussian noise with a single-sided power density spectrum N_0 W/hz. The signal also suffers a random phase shift $\theta(t)$ in the channel. The received signal is demodulated at the receiver by means of a reference carrier preceded by a phase-locked loop (PLL). This demodulation produces (Ref. 1)

$$y(t) = x_k(t)\left[(1 - m^2)P\right]^{1/2} \cos\phi(t) + n'(t) \tag{2}$$

where

$$\phi(t) \triangleq \theta(t) - \hat{\theta}(t)$$

is the error between the phase shift in the channel $\theta(t)$ and the estimate $\hat{\theta}(t)$ of this phase shift given by the PLL, and $n'(t)$ is white gaussian noise with single-sided power density spectrum N_0 W/hz. The decision on the transmitted signal is made by correlating $y(t)$ with the possible received signals. Therefore we form

$$q = \int_0^T y(t)\left[x_1(t) - x_2(t)\right] dt \tag{3}$$

If $q \gtrless 0$, we decide that x_1 was transmitted; otherwise we take the transmitted signal to be $x_2(t)$.

3. System Error Probability

For a given realization of the random process $\phi(t)$, $0 \leq t \leq T$, q is normally distributed with mean

$$\mu = \pm 2\left[(1 - m^2)P\right]^{1/2} \int_0^T \cos\phi(t)\, dt$$

and variance $\sigma^2 = 2TN_0$. The conditional probability of error has been shown (Ref. 1) to be given by

$$P_E\left[\phi(t), 0 \leq t \leq T\right] = \int_{\left[R(1-m^2)Y\right]^{1/2}}^{\infty} \frac{\exp(-x^2/2)}{(2\pi)^{1/2}}\, dx \tag{4}$$

where

$$R \triangleq 2\left(\frac{PT}{N_0}\right)$$

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

The average probability of error is found by averaging over the random variable Y.

However, the exact computation of the probability density $f_z(y)$ appears to be a formidable problem. In order to circumvent this difficulty we assume that the phase error is small enough so that the linear PLL theory can be used. For most practical situations this assumption

[1]Consultant, Electrical Engineering Dept., University of Southern California.

is valid. The consequences of this assumption are two-fold: First, from the linear phase-locked loop theory (Ref. 4) we know that $\phi(t)$ is a gaussian process; second we can make the approximation (valid for the linear region of operation)

$$Y = 1 - \frac{1}{2T}\int_0^t \phi^2(t)\,dt \tag{5}$$

Thus the computation of $f_Y(y)$ reduces to computing the probability density of the integral of the square of a gaussian process.

In the case of a first-order PLL, a useful approximation to this density is known (Ref. 5). The PLL tracks a carrier of power m^2P immersed in white gaussian noise with power density spectrum N_0 W/hz. The correlation func- of $\phi(t)$ is

$$R_\phi(\tau) = \frac{B_L N_0}{m^2 P}\exp\left[-4B_L|\tau|\right] \tag{6}$$

where B_L is the equivalent noise bandwidth of the loop (Ref. 1). The probability density $f_z(z)$ of the random variable

$$z = \frac{m^2 P}{N_0 B_L T}\int_0^T \phi^2(t)\,dt \tag{7}$$

is well approximated by (Ref. 6)

$$f_z(z) = \frac{1}{(\pi\delta)^{1/2}}\exp(2/\delta)\,z^{-3/2}\exp\left[-\frac{1}{\delta}\left(z+\frac{1}{z}\right)\right], \quad 0 \leq z \leq \infty \tag{8}$$

where $\delta = 1/TB_L$ is the ratio of the system data rate $1/T$ to the bandwidth of the carrier tracking loop B_L. The density given by Eq. (8) is extremely accurate for $\delta \leq 5$, the case of interest here. We further assume that the data is modulated onto a square-wave subcarrier whose fundamental frequency is sufficiently large so that the power in the modulation does not ente. into the bandwidth of the loop. If, in fact, this is not the case, any appreciable sideband power which exists around the carrier would act as a self-noise in the loop and tend to introduce more jitter on the reference phase. Such designs are usually not interesting in practice, even though the analysis is relatively straightforward.

Averaging Eq. (4) over the random variable Z we find

$$P_E = \int_0^\infty f_z(z)\int_{A(z)}^\infty \frac{1}{(2\pi)^{1/2}}\exp(-y^2/2)\,dy\,dz \tag{9}$$

where

$$A(z) = [R(1-m^2)]^{1/2}\left(1 - \sigma_\phi^2\frac{z}{2}\right)$$

$$\sigma_\phi^2 = \frac{N_0 B_L}{m^2 P}$$

In the limit as δ approaches infinity Eq. (14) becomes

$$P_E = \frac{1}{(2\pi)^{1/2}}\int_{[R(1-m^2)]^{1/2}}^\infty \exp(-y^2/2)\,dy \tag{10}$$

which checks with the case for perfect phase synchronization.

4. Numerical Results

The double integral in Eq. (9) has been evaluated over a range of values for the parameters m, δ, and R. In Fig. 11 we plot P_E as a function of $m^2 = P_c/P$, with R as a parameter, for several different values of δ. These curves exhibit the same behavior (Ref. 1) as those plotted for the case of slowly varying phase error. For each value of R and δ there is an optimum modulation factor, e.g., m_0, for which the probability of error is minimum. Also, m_0 decreases as δ increases. In Fig. 12 we plot m_0^2 as a function of R with δ as a parameter. Finally, in Fig. 13 we show the minimum probability of error P_{E_0} corresponding to m_0 as a function of R with δ as a parameter. These results may be compared with those obtained in Ref. 1 for the case where $\phi(t)$ is constant for $0 \leq t \leq T$. They may be further used to design phase-coherent systems with $\delta < 1$.

5. Conclusions

As we have seen here and in Refs. 1 and 3, the phase-error introduced into the data detection process by a radio-frequency carrier tracking loop produces deleterious effects on system performance. As opposed to the problem considered in Ref. 1, this paper has treated the case where this error may vary over the duration of the received symbol. Useful design trends and results were established for the situation where the error $\phi(t)$ may be considered gaussian. For most practical applications this assumption is valid.

References

1. Lindsey, W. C., "Optimal Design of One-Way and Two-Way Coherent Communication Links," *IEEE Trans. Comm. Tech.*, Vol. COM-14, No. 4, pp. 418–431, August 1966.
2. Van Trees, H., "Optimum Power Division in Coherent Communication Systems," *IEEE Trans. Space Elect. Telem.*, Vol. SET, pp. 1–9, March, 1964.

3. Lindsey, W. C., "Determination of Modulation Indexes and Design of Two-Channel Coherent Communication Systems," *IEEE Trans. Comm. Tech.*, Vol. COM-15, No. 2, pp. 229–237, April, 1967.

4. Viterbi, A. J., *Principles of Coherent Communication*, McGraw-Hill Book Co., Inc., New York, 1966.

5. Grenander, U., Pollak, H. O., and Slepian, D., "The Distribution of Quadratic Forms in Normal Variates: A Small Sample Theory With Applications to Spectral Analysis," *J. Soc. Ind. Appl. Math.*, pp. 374–401, December 1959.

6. Slepian, D., "Fluxuations of Random Noise Power," *Bell System Technical Journal*, pp. 163–184, January 1958.

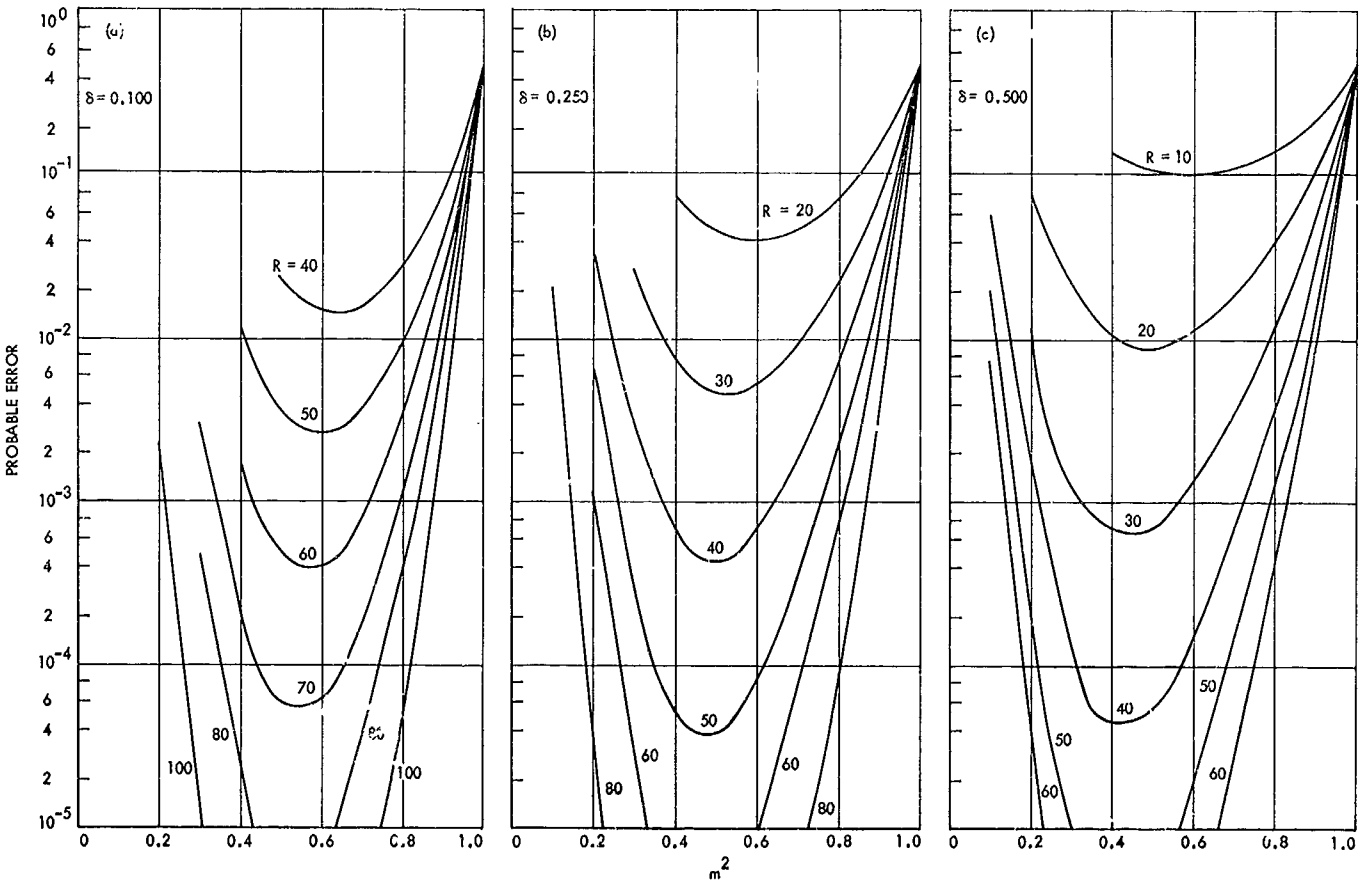


Fig. 11. Error probability P_E versus modulation factor m for various values of R: (a) $\delta = 0.1$, (b) $\delta = 0.5$, (c) $\delta = 0.25$

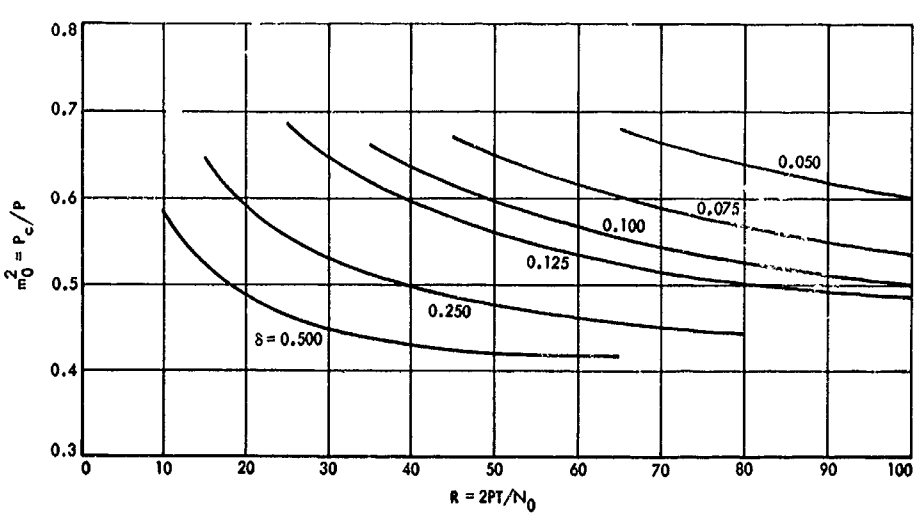


Fig. 12. Optimum modulation factor m_0 versus R for various values of δ

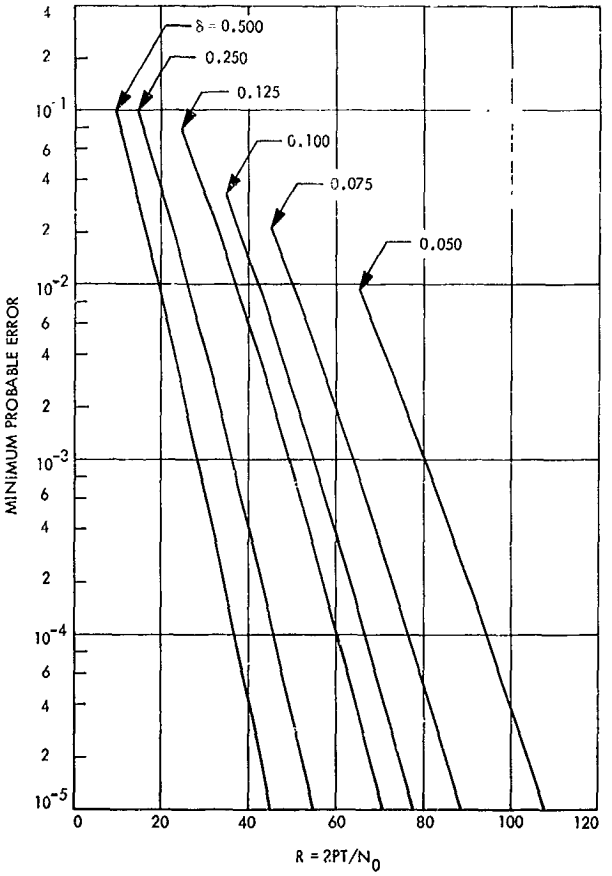


Fig. 13. Minimum error probability P_{E_0} versus R for various values of δ

E. Propagation Studies: Absolute Time by Pulsar,

G. A. Morris, Jr.

1. Introduction

Rapidly pulsating radio stars were reported early this year by the radio-astronomy group at Cambridge (Ref. 1). Further investigation confirms that the periods of these sources are quite regular, and several sources have been identified. The regular pulsations of these sources suggest they might be useful as an inexpensive means of determining absolute time among the Deep Space Network stations.

In order to be useful in determining absolute time these sources must have excellent long- and short-term stability. To be useful from an operational standpoint it must be possible to predict the arrival time of a pulse and also to make an accurate measurement of the arrival time in a short period of observation.

To obtain information about the usefulness of these sources an observational program was started in early August at S-band. The object of this program is to examine all pulsars as they are discovered to determine their usefulness as an absolute time standard. It is preferable to use sources with a large stable flux density in order to increase the signal-to-noise ratio and therefore reduce the time required to make an estimate of pulse arrival time. The pulse shape should be stable and, if possible, should have a structure that can be used to estimate arrival time readily. As of November 1, four sources have been found which satisfy the above criteria. Time versus average flux plots of these sources are shown in Fig. 14. Although individual pulses are generally too weak to be observed, averages of many pulses are easily detectable. Regular weekly observations of these sources will be made for a period of at least 1 yr. The difference between predicted and measured arrival times will be minimized in the least-squares sense, using differential correction techniques to confirm the source's celestial position and pulse period. Right ascension, declination, distance, period, and galactic rotation are among the parameters to be estimated by the least-squares fit. This information is required before accurate predictions of pulse arrival time can be made. Further, this information is valuable in determining the inherent stability of period and celestial position for possible optical identification.

2. System Description

A block diagram of the system is shown in Fig. 15. The only special equipment required is the pulsar timer, which supplies one interrupt to the computer per pulsar period and 5000 sampling pulses to the analog-to-digital converter every period. The 1-pulse/s signal from the cesium clock is used to start the timer on an exact second so that arrival time measurements may be made. The synthesizer used to determine the period is driven from the cesium clock 1-MHz output so that stability of period is obtained. Data is obtained by sampling the signal 5000 times per period and integrating the data for 500 periods. Data is normally taken for 1 h. Then the synthesizer is changed to a new frequency to correct for variations in the apparent period of the pulsar caused by the orbital and rotational motions of the earth. Data is recorded on magnetic tape, after being integrated over 500 pulses, for later analysis in the least-squares fit.

3. Variance–Covariance Analysis

A variance–covariance analysis was performed to determine how accurately several parameters could be measured for the source CP0950, assuming arrival time

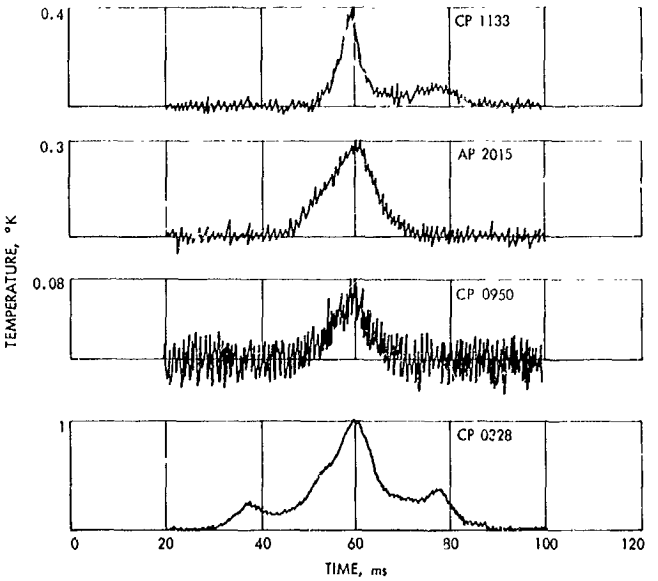


Fig. 14. Time versus flux of S-band pulsars

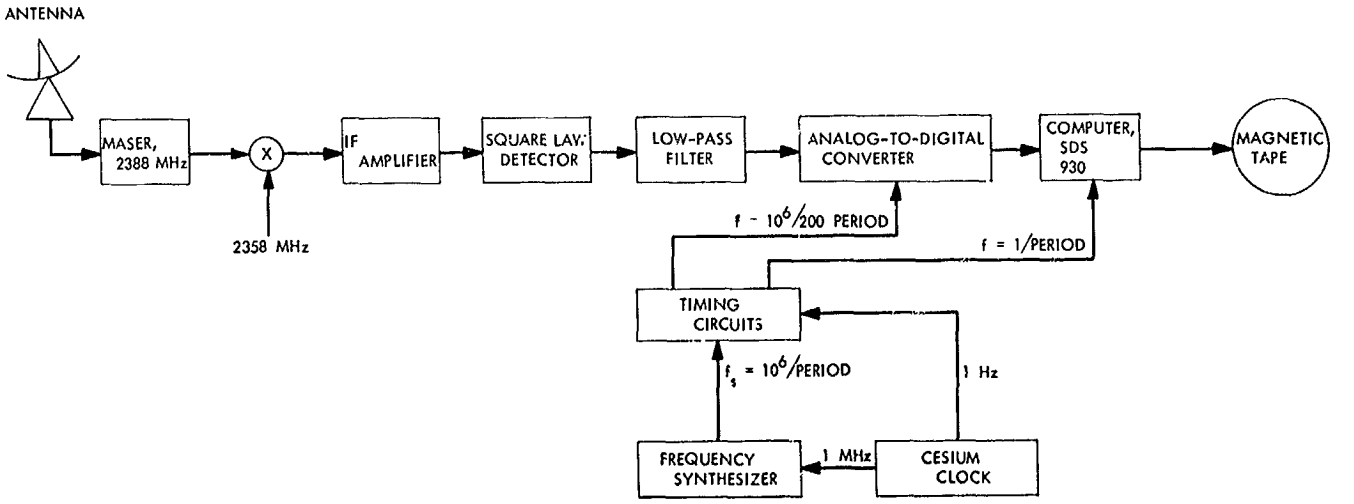


Fig. 15. Block diagram of pulsar system

measurements accurate to 100 μs. Present data indicates this is a reasonable estimate of accuracy. The standard deviations for position and period are shown in Table 1.

Table 1. Variance–covariance analysis based on arrival time measurement accurate to 100 μs for CP0950

Parameter	Standard deviation
Period	3.07×10^{-13} sec
Right ascension	0.00293 arc sec
Declination	0.00763 arc sec

A simplified earth orbit was used, and the source was assumed to have no proper motion, no radial acceleration, and a constant period. The position obtained by this technique, after 1 yr, should be of great benefit in making an optical identification. The standard deviation of the period ranks with the cesium standard as to stability.

Reference

1. Hewish, A., et al., "Observation of a Rapidly Pulsating Radio Source," *Nature*, Vol. 217, pp. 709–713, 1968.

F. Communications Systems Development: Effects of Phase-Locked Loop Dynamics on Phase-Coherent Communications,

I. F. Blake[5] and W. C. Lindsey[6]

1. Introduction

In a recent paper (Ref. 1) one of the authors analyzed the performance of a phase-coherent receiver preceded by a bandpass limiter. Two expressions for the error probability were given for the cases: (1) where the phase error varies rapidly over the symbol interval, and (2) where the phase error remains essentially constant over the symbol interval. The parameter which essentially characterizes how fast the phase error varies relative to the symbol time T is the ratio of the system data rate $\mathcal{R}$ to the bandwidth W_{LO} of the loop at the design point; i.e., $\delta = \mathcal{R}/W_{LO}$. In practice, the question naturally arises as for what region of δ will case (2) analysis be valid. It is also of interest to understand how the error rate changes as the parameter δ varies over all regions of the phase-error variations. Consequently, the purpose of this article is to present an analysis which deals with the general case of the phase-error variation for all values of δ. The results are particularly useful in the design of medium-rate phase-coherent communication systems used for space communication, e.g., *Mariner*- and *Pioneer*-type telecommunications systems.

2. System Model

If the received signal is tracked by means of a phase-locked loop, the output of the phase detector can be written as

$$y(t) = x_k(t)\cos\phi(t) + n(t) \tag{1}$$

where $\phi(t)$ is the time-varying phase error in the loop, and $n(t)$ is white gaussian noise of two-sided spectral density $N_0/2$ watts/sec (Ref. 1). The signals

$$x_k(t), \qquad k = 1, 2$$

are constant over the bit interval of duration T sec:

$$x_1(t) = +A$$

$$x_2(t) = -A$$

The correlator output, on which the decision as to whether $x_1(t)$ or $x_2(t)$ was sent is based, can be shown to be given by (Ref. 1)

$$q_k = \int_0^T [x_k(t)\cos\phi(t) + n(t)]\,[x_2(t) - x_1(t)]\,dt \tag{2}$$

If q_k is negative, we decide that $x_1(t)$ was sent, and if positive, $x_2(t)$. It is assumed that the signal-to-noise ratio in the tracking loop is such that the phase error process $\phi(t)$ is gaussian with correlation function

$$R_\phi(\tau) = \frac{1}{\rho_L}\exp[-\gamma|\tau|] = \sigma_L^2\exp[-\gamma|\tau|] \tag{3}$$

This assumption is justified in the case of a first-order loop with no bandpass limiter preceding it. For the case under consideration here, it is an approximating assumption essential for the analysis. The agreement of known experimental and theoretical results tend to support its use.

The parameter γ in Eq. (3) is defined later in terms of loop and filter parameters in Eq. (11). When the loop is preceded by a band-pass limiter, the quantity σ_L^2 in Eq. (3) is well approximated in the region of interest in practice by (Ref. 1)

$$\frac{1}{\sigma_L^2} = \rho_L$$

$$= \frac{3z}{\Gamma(1 + 2/\mu)}$$

$$\Gamma = \frac{1 + 0.345\,zy}{0.862 + 0.690\,zy}$$

$$z = \frac{P_c}{N_0 b_{LO}}$$

$$y = \frac{b_{LO}}{b_H}$$

$$\mu = \frac{(\Upsilon_0)^{1/2}\exp\left(-\frac{\Upsilon_0 y}{2}\right)\left[I_0\left(\frac{\Upsilon_0 y}{2}\right) + I_1\left(\frac{\Upsilon_0 y}{2}\right)\right]}{z^{1/2}\exp\left(-\frac{zy}{2}\right)\left[I_0\left(\frac{zy}{2}\right) + I_1\left(\frac{zy}{2}\right)\right]}$$

$$\Upsilon_0 = \frac{P_c o}{N_0 B_{LO}} \tag{4}$$

[5]Resident Research Associate, NASA–NRC.

[6]Consultant, Electrical Engineering Dept., University of Southern California.

where P_c is the carrier power, b_L and b_{IF} are loop and IF amplifier (in the double heterodyne receiver) bandwidths ($w_{IF} = 2b_{IF}$, $w_L = 2b_L$) (Ref. 1). The subscript 0 refers to the values of the loop parameters at the design point, which is defined as $z_0 = \gamma_0 =$ constant. The parameter μ is the ratio of the limiter suppression factor at the design point to its value at any other point and

$$I_k(x), \qquad k = 1, 2$$

are modified Bessel functions of order k. The parameter zy is the signal-to-noise ratio at the output of the receiver IF amplifier.

3. Error Probability Calculations

If the parameter z is sufficiently large, the cosine term in Eq. (1) can be replaced by the approximation

$$\cos\phi(t) \cong 1 - \frac{\phi^2(t)}{2} \tag{5}$$

It turns out that in practice this approximation is sufficiently accurate in the region of interest. The decision variable becomes

$$q_k = \beta_k T - \frac{1}{2}\beta_k \int_0^T \phi^2(t)\,dt - 2A\int_0^T n(t)\,dt \tag{6}$$

where

$$\begin{aligned} \beta_k &= x_k(t)\,[x_2(t) - x_1(t)] \\ &= -2A^2, \qquad k = 1 \\ &= +2A^2, \qquad k = 2 \end{aligned} \tag{7}$$

Define the new decision variable

$$\begin{aligned} z_k &= -\frac{(q_k - \beta_k T)}{\frac{1}{2}\beta_k T} \\ &= \frac{1}{T}\int_0^T \phi^2(t)\,dt + \frac{2A}{\frac{1}{2}\beta_k T}\int_0^T n(t)\,dt \end{aligned} \tag{8}$$

where now it is decided that $x_1(t)$ was sent if $z_k > 2$, and $x_2(t)$, otherwise. The second term on the right-hand side of Eq. (8) is a gaussian random variable with mean zero and variance

$$\sigma^2 = \frac{2N_0}{A^2T} \triangleq \frac{2}{R} \tag{9}$$

The probability density of the first integral in Eq. (8) is very difficult to determine. However, a useful approximation to it has been found (Ref. 2):

$$\begin{aligned} g(x) = \rho_L\left(\frac{\beta_L}{2\pi}\right)^{1/2} e^{\beta_L}(\rho_L x)^{-3/2} \\ \times \exp\left[-\frac{\beta_L}{2}\left(\rho_L x + \frac{1}{\rho_L x}\right)\right], \qquad x \geqq 0 \end{aligned} \tag{10}$$

where

$$\begin{aligned} \beta_L &= \frac{\gamma T}{2} \\ &= \frac{1}{\delta}\left(\frac{1 + r_{0/\mu}}{1 + r_0}\right) \end{aligned} \tag{11}$$

where γ is the parameter appearing in the exponent of the correlation function in Eq. (3), r_0 is a loop parameter, μ is as defined in Eq. (4), and

$$\delta = \frac{\mathscr{R}}{W_{LO}} = \frac{\mathscr{R}}{2b_{LO}} \tag{12}$$

The density in Eq. (10) is a unimodal function which is zero at the origin and a maximum at

$$x = \frac{1}{2}\left[\frac{9}{\beta_L^2\rho_L^4} + \frac{4}{\rho_L^2}\right]^{1/2} - \frac{3}{2\beta_L\rho_L^2} \tag{13}$$

The approximation to the true density becomes very close for $\beta_L > 5$. As $\beta_L \to \infty$, the maximum tends to $x = 1/\rho_L$ and the function tends to a delta function at this point. As $\beta_L \to 0$, $g(x)$ tends to a delta function at the origin. Similarly, for fixed β_L as $\rho_L \to \infty$, the function tends to a delta function at the origin. Notice that in order to make the approximation to the cosine valid, ρ_L should be at least 3 or larger.

If it is assumed that the two integrals in Eq. (8) are independent random variables (an approximation which can be partially justified on physical grounds for the case of interest here), then the probability density of the random variable z_k can be written as

$$p(z_k) = \int_0^\infty \frac{1}{(2\pi\sigma)^{1/2}} \exp\left[-\frac{(z_k - x)^2}{2\sigma^2}\right] g(x)\,dx \tag{14}$$

Assuming $x_1(t)$ was sent, the probability of error

$$P_E = \text{prob}\,[z_k > 2]$$
$$= \int_2^\infty \int_0^\infty \frac{1}{(2\pi\sigma)^{1/2}} \epsilon \exp\left[-\frac{(z-x)^2}{2\sigma^2}\right] g(x)\,dx\,dz \tag{15}$$

Interchanging the integration yields

$$P_E = \int_0^\infty \text{erfc}\left(\frac{2-x}{\sigma}\right) g(x)\,dx \tag{16}$$

where $\text{erfc}(\cdot)$ is the complementary error function

$$\text{erfc}(y) = \int_y^\infty \frac{1}{(2\pi)^{1/2}} \exp(-\mu^2/2)\,du \tag{17}$$

Notice that as $\rho_L \to \infty$, $P_E \to \text{erfc}(2R)^{1/2}$ which coincides with the result obtained in Eqs. (24) and (38) of Ref. 1 under this condition.

From the results of Ref. 2 it appears that the density function of Eq. (10) becomes a poor approximation as β_L decreases much below unity. Since

$$\beta_L = \frac{1}{\delta}\left(\frac{1 + r_0/\mu}{1 + r_0}\right) \tag{18}$$

this corresponds to δ being greater than

$$\left(\frac{(1 + r_0/\mu)}{1 + r_0}\right)$$

The parameter μ is a function of $x(= z/2)$ and other parameters. To find the region of validity of the curves, consider first Fig. 16, where $R = 4.8$. For $x = 4$, $\mu \cong 0.5$ and the curve is valid for $\delta < 2$. For $x = 300$, $\mu \cong 0.078$ and the curve is valid for $\delta < 9$. Drawing a straight line between these points gives a good approximation to the region of validity. The same limits hold for Fig. 17 also.

4. Results and Conclusions

That component of the error probability, assuming phases reduced modulo 2π, has been evaluated based upon the approximation that

$$\cos\phi(t) = 1 - \frac{\phi^2(t)}{2}$$

This does not imply the assumption that $\phi(t)$ is small, i.e., linear PLL theory holds. If linear PLL theory were

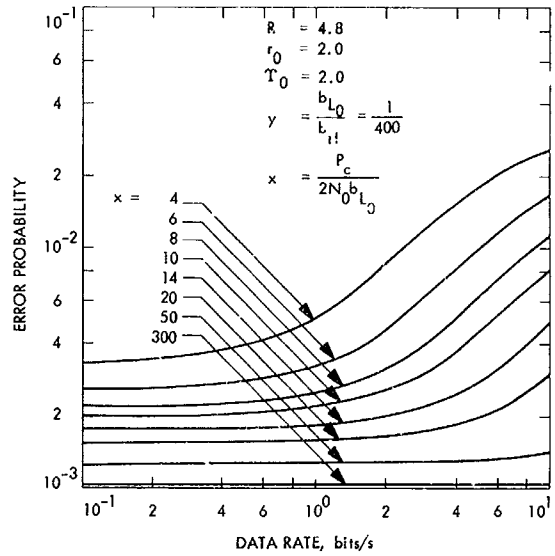


Fig. 16. Error probability versus data rate (phase error varies over bit interval), $R = 4.8$

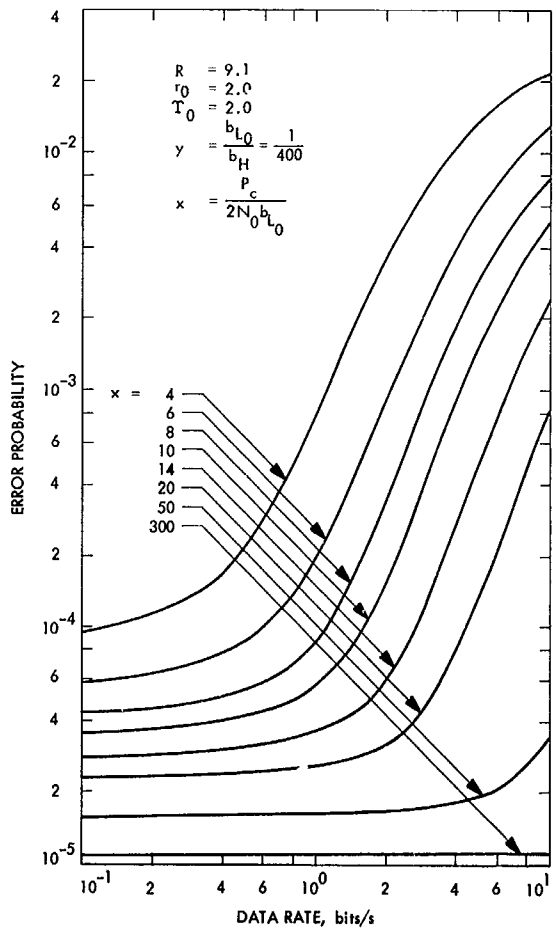


Fig. 17. Error probability versus data rate (phase error varies over bit interval), $R = 9.1$

invoked $\cos\phi(t)$ would be replaced by $\cos\phi(t) = 1$ and $\sin\phi(t) \simeq \phi(t)$. On the other hand, in order to describe the statistics of

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

$$\cong 1 - \frac{1}{2T}\int_0^T \phi^2(t)\,dt$$

it has been necessary assume that $\phi(t)$ is gaussian. In principle, however, this assumption is not at all restrictive since the variance of the actual ϕ process at work in the loop can be substituted into the probability distribution $f_Y(y)$. Thus the nonlinear effects of the loop are taken into account.

Figures 16 and 17 illustrate the performance of a binary phase-shift-keyed communication system when $\phi(t)$ varies over the symbol interval. The signal-to-noise ratio $R = ST_0/N_0$ has been set such that the error probability of the system would be 10^{-3} (Fig. 16) and 10^{-5} (Fig. 17) in a perfectly synchronized system. For values of $4 \leq \delta \leq 5$ the results check, for all practical purposes, with those given previously (Ref. 1) where it is assumed that $\cos\phi$ is essentially constant over the symbol interval. For $\delta < 4$ the results presented here begin to deviate appreciably from those where $\cos\phi$ is assumed constant; hence, the model introduced here will be useful in designing and testing of phase-coherent systems which operate with $\delta < 4$, the low-rate end of the region of δ.

References

1. Lindsey, W. C., "Performance of Phase Coherent Receivers Preceded by Bandpass Limiters," *IEEE Trans. Comm. Tech.*, Vol. COM-16, No. 2, pp. 245–251, April 1968.
2. Grenander, U., Pollak, H. O., and Slepian, D., "The Distribution of Quadratic Forms in Normal Variates: A Small Sample Theory With Applications to Spectral Analysis," *J. Soc. Ind. Appl. Math.*, Vol. 7, No. 4, pp. 374–401, Dec. 1959.

G. Communications Systems Development: Efficiency of Noisy Reference Detection,

R. C. Tausworthe

1. Introduction

Lindsey (Ref. 1) has published results which, for a given modulation index, relate the observed signal-to-noise ratios (SNR) to equivalent signal losses caused by the noisy demodulation process. Recent measurements of the performance of the 8⅓-bits/s *Mariner* Mars 1969 engineering telemetry instituted a reevaluation of these analyses. The following report is a tabulation of the method for determining the demodulation efficiency as a function of loop phase error. Performance is then related to these efficiencies through the modulation index. This method was chosen because of the flexibility it affords when many indices, bandwidths, etc., are being considered.

2. Efficiency Equations

The output of a coherent amplitude detector is a process of the form (Fig. 18)

$$z(t) = P^{1/2} m(t)\, g(\phi) + n(t) \tag{1}$$

in which $n(t)$ is wide-band noise normalized to have the same two-sided spectral density $N_0 = N_+/2$, as the input noise; $P = A^2$ is the rms detected sideband power; $g(\phi)$ is the detector phase characteristic, normalized so that $g(0) = 1$; and $m(t)$ is the detected modulation process, normalized so that $E(m^2(t)) = 1$. The modulation waveform we shall assume is one of M messages $\{m_k(t)\}$, for $0 \leq t < T$.

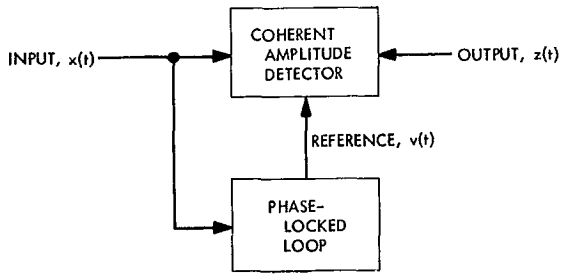


Fig. 18. Coherent detection by loop-derived reference

The process $z(t)$ is the input to a set of correlators, whose outputs at the end of a T-sec message are

$$w_k = \frac{A}{T}\int_0^T m(t)\, m_k(t)\, g(\phi(t))\,dt + N(T) \tag{2}$$

$N(T)$ then is a gaussian random variable with variance

$$\sigma_N^2 = \frac{N_0}{T} \tag{3}$$

We shall also assume that the phase-error process is derived from a phase-locked loop tracking the carrier or subcarrier. The loop bandwidth will be denoted w_L

and the phase-error density $p(\phi)$. The integral term of Eq. (2) can be written as a mean value plus a variational term

$$w_k = A(r_k\mu + \nu_k) + N(T) \tag{4}$$

in terms of the normalized cross-correlation r_k between the incoming message $m(t)$ and the kth stored one, $m_k(t)$, where $\mu = E[g(\phi)]$, the ν_k are random variables which depend on the value of k, and $\nu = (\nu_1, \cdots, \nu_m)$ possesses some distribution $p(\nu)$. The actual overall probability of error is the average conditional error probability

$$\bar{P}_E = \int Pr[\text{error}|\nu_1, \cdots, \nu_m]\, p(\nu_1, \cdots, \nu_m)\, d\nu_1 \cdots d\nu_m \tag{5}$$

The difficulty in specifying the characteristics of $p(\nu_1, \cdots, \nu_m)$ lead to approximations for Eq. (5).

First, if the loop response is considerably more rapid than the integration time T (i.e., $\delta = 2/w_L T << 1$) then the correlator output tends to the average

$$\nu_k \approx 0 \tag{6}$$

in which case the outputs appear all to have an equivalent constant factor $E[g(\phi)]$ multiplying the signal amplitude A. Performance is then the same as it would be if the signal power were reduced by the factor $E^2[g(\phi)]$. The error rate will fit the usual maximum likelihood theory, giving rise to a probability of error as a function of the matched filter SNR parameter ρ_{mf}:

$$\rho_{mf} = \frac{PT}{N_+} E^2[g(\phi)] = RE^2[g(\phi)] \tag{7}$$

in which $R = PT/N_+$ is the undegraded value of ρ_{mf}. In this case, it is easy to see that the detector efficiency η is merely

$$\eta_0 = E^2[g(\phi)] = \mu^2 \tag{8}$$

A second approximation can be made when the loop is very sluggish with respect to the message (i.e., $\delta = 2/w_L T >> 1$). Then, over the interval $(0, T)$, the phase error is nearly constant (but randomly distributed according to $p(\phi)$). In this case, the correlator outputs are very nearly

$$w_k = Ar_k g(\phi) + N(T) \tag{9}$$

so that the averaged error probability yields the overall error rate

$$\bar{P}_E(R^{1/2}) = \int P_E(R^{1/2} g(\phi)|\phi)\, p(\phi)\, d\phi \tag{10}$$

The degradation is then clearly

$$\eta_\infty = \frac{1}{R}\{\bar{P}_E^{-1}[P_E(R^{1/2})]\}^2 \tag{11}$$

Because of the convexity of the function $Pr[E|\nu]$, it follows that the actual efficiency is bounded by Eqs. (8) and (11):

$$\eta_\infty \leq \eta_\delta \leq \eta_0 \tag{12}$$

3. Error Probability

Since Eq. (11) requires it, let us consider the error probability function. For no coding and antipodal binary signals, the error rate is

$$P_E = \frac{1}{2}\operatorname{erfc}(R)^{1/2} \tag{13}$$

For orthogonal, equi-energy signals, the error rate is

$$\begin{aligned} P_E &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \exp(-y^2/2) \\ &\quad \times \left[\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{v+(2R)^{1/2}} \exp(-v^2/2)\, dv\right]^{M-1} dy \\ &\approx \frac{M-1}{2} \operatorname{erfc}\left(\frac{R}{2}\right)^{1/2} \end{aligned} \tag{14}$$

As a function of R, biorthogonal codes behave much the same (Ref. 2) as Eq. (14) indicates. Thus, for the three cases (no coding, orthogonal, biorthogonal) we have

$$P_E(R_1^{1/2}) \approx P_E(R_0^{1/2}) \frac{\operatorname{erfc}(\lambda R_1)^{1/2}}{\operatorname{erfc}(\lambda R_0)^{1/2}} \tag{15}$$

for any values R_0 and R_1 of R. The coefficient λ relates to the coding:

$$\lambda = \begin{cases} 1, & \text{uncoded} \\ \dfrac{1}{2}, & \text{coded (orthogonal/biorthogonal)} \end{cases} \tag{16}$$

We can set $R_0^{1/2} = R^{1/2}$ and $R_1^{1/2} = R^{1/2} g(\phi)$, above, to evaluate Eq. (10):

$$\bar{P}_E(R^{1/2}) = \frac{P_E(R^{1/2})}{\operatorname{erfc}(\lambda R)^{1/2}} \int_{-\pi}^{\pi} \operatorname{erfc}(\lambda R)^{1/2} g(\phi)) \, p(\phi) \, d\phi \tag{17}$$

But the same error rate would occur if the value of R were R_{eq} and no degradation were allowed:

$$\bar{P}_E(R^{1/2}) = P_E(R_{eq}^{1/2}) = \frac{P_E(R^{1/2})}{\operatorname{erfc}(\lambda R)^{1/2}} \operatorname{erfc}(\lambda R_{eq})^{1/2} \tag{18}$$

It is then clear that

$$R_{eq} = \frac{1}{\lambda} \left\{ \operatorname{erfc}^{-1} \left[\int_{-\pi}^{\pi} \operatorname{erfc}((\lambda R)^{1/2} g(\phi)) \, p(\phi) \, d\phi \right] \right\}^2 \tag{19}$$

and, correspondingly, that

$$\eta_\infty = \frac{1}{\lambda R} \left\{ \operatorname{erfc}^{-1} \left[\int_{-\pi}^{\pi} \operatorname{erfc}((\lambda R)^{1/2} g(\phi)) \, p(\phi) \, d\phi \right] \right\}^2 \tag{20}$$

Thus, it remains only to evaluate Eqs. (8) and (20) for given $g(\phi)$ and $p(\phi)$ to obtain limits on η_δ.

4. Carrier-Extraction Degradation

For the carrier-extraction process the detector characteristic is

$$g(\phi) = \cos\phi \tag{21}$$

and the phase-error distribution, based on the first-order loop theory, is approximately (Ref. 3)

$$p(\phi) = \frac{\exp(\rho \cos\phi)}{2\pi I_0(\rho)} \tag{22}$$

in terms of the loop equivalent SNR (Ref. 4)

$$\rho = \frac{P_c}{N_0 \omega_L \Gamma} \tag{23}$$

Based on this ρ, a certain loop phase error σ^2 is present in the loop:

$$\sigma^2 \approx \frac{\pi}{3} + \frac{4}{I_0(\rho)} \sum_{n=1}^{\infty} \frac{(-1)^n I_n(\rho)}{n^2}$$

$$\sim \frac{1}{\rho}, \qquad \text{as } \rho \to \infty \tag{24}$$

It also follows that the degradation for $\delta << 1$ is

$$\eta_0 = \left(\frac{I_1(\rho)}{I_0(\rho)} \right)^2 \tag{25}$$

The value for η_∞ has been obtained by numerical integration, and appears along with the η_0 of Eq. (25) in Fig. 19, cross-plotted as a function of the loop error. It may be noted that when σ^2 is small, the two bounds converge approximately to the gaussian-phase-error result

$$1 - \eta_0 \sim \frac{1}{\rho} \sim \sigma^2 \sim 1 - \exp(-\sigma^2) \tag{26}$$

But as degradation becomes an appreciable percent, the two separate and depend not only on σ^2, but on λR as well. Because of the increasing steepness of P_E with λR, the degradation for $\delta >> 1$ becomes more drastic as λR increases. The degradation for $\delta << 1$ is, however, independent of λR.

5. Subcarrier-Loop Degradation

Assuming that the subcarrier is a square wave, the detector characteristic becomes triangular:

$$g(\phi) = 1 - \frac{2}{\pi}|\phi|, \qquad \text{for } |\phi| \leq \frac{3\pi}{2} \tag{27}$$

The approximate loop error density (based on a first-order loop) is again related (SPS 37-31, Vol. IV, pp. 311–325) to the loop equivalent SNR by

$$p(\phi) = \begin{cases} \dfrac{1}{C} \exp\left[-\dfrac{2\rho}{\pi^2} \phi^2 \right], & \text{for } |\phi| \leq \pi/2 \\ \dfrac{1}{C} \exp\left[-\rho \left(1 - \dfrac{2(\phi - \pi)^2}{\pi^2} \right) \right], & \pi/2 \leq \dfrac{3\pi}{2} \end{cases}$$

$$C = \frac{\pi^{3/2}}{(2\rho)^{1/2}} \left\{ \operatorname{erf}\left(\frac{\rho}{2} \right)^{1/2} + e^{-\rho} h\left(\frac{\rho}{2} \right)^{1/2} \right\} \tag{28}$$

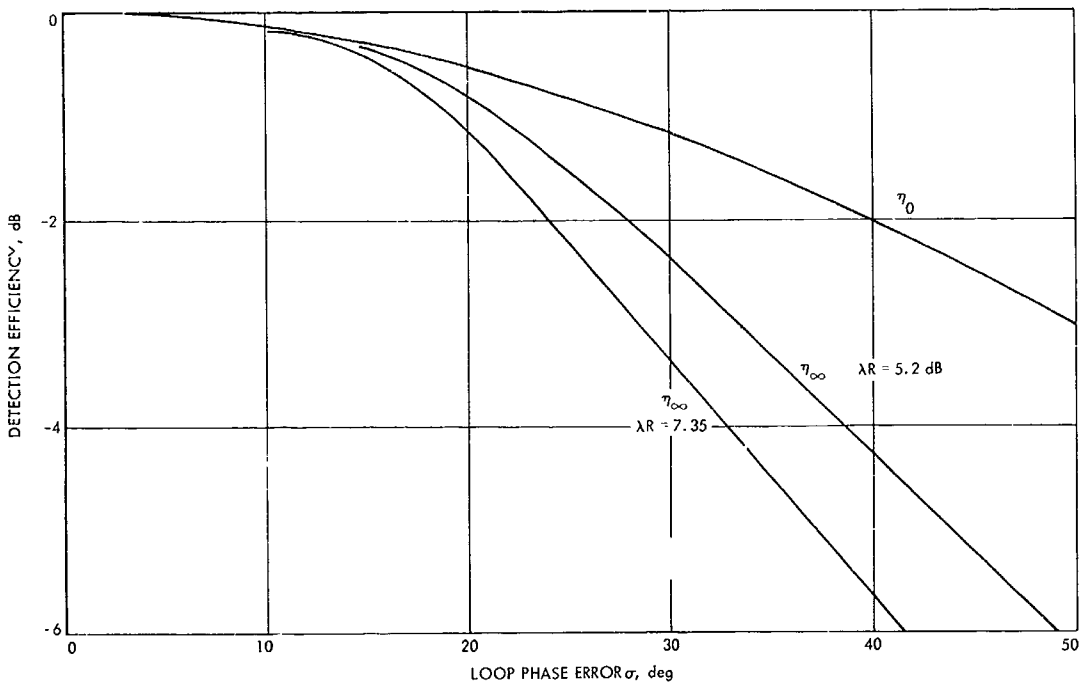


Fig. 19. Loop-derived reference degradation bounds for carrier tracking loop

where the imaginary error function

$$h(x) = \frac{2}{\pi^{1/2}} \int_0^x e^{t^2}\, dt$$

$$= \frac{2}{\pi^{1/2}}\, x \sum_{n=0}^{\infty} \frac{x^{2n}}{n!\,(2n+1)} \tag{29}$$

The phase-error variance is then

$$\sigma^2 = \left(\frac{\pi}{2}\right)^{5/2} \frac{\pi}{\rho^{3/2} C} \left\{ \operatorname{erf}\left[\left(\frac{\rho}{2}\right)^{1/2}\right] + (4\rho - 1)\, h\left[\left(\frac{\rho}{2}\right)^{1/2}\right] - 4\left(\frac{2\rho}{\pi}\right)^{1/2} e^{-\rho/2}\left(1 - e^{-\rho/2}\right) \right\} \tag{30}$$

and the low-rate efficiency can be straightforwardly obtained as

$$\eta_0 = \frac{\operatorname{erf}\left[\left(\frac{\rho}{2}\right)^{1/2}\right] - e^{-\rho}\, h\left[\left(\frac{\rho}{2}\right)^{1/2}\right]}{\operatorname{erf}\left[\left(\frac{\rho}{2}\right)^{1/2}\right] + e^{-\rho}\, h\left[\left(\frac{\rho}{2}\right)^{1/2}\right]} \times \left\{ 1 - \frac{(1 - e^{-\rho/2})^2}{\left(\frac{\pi\rho}{2}\right)^{1/2} \left\{ \operatorname{erfc}\left[\left(\frac{\rho}{2}\right)^{1/2}\right] - e^{-\rho}\, h\left[\left(\frac{\rho}{2}\right)^{1/2}\right] \right\}} \right\} \tag{31}$$

Again, the values for η_∞ have been obtained by numerical integration, and the two different behaviors plotted in Fig. 20 for comparison. As was evident in the previous case as well, the two degradations at low σ^2 behave like

$$\eta_0 = \left[1 - \left(\frac{2}{\pi}\right)^{3/2} \sigma\right]^2 \approx 1 - 2\left(\frac{2}{\pi}\right)^{3/2} \sigma \tag{32}$$

as would be predicted by a gaussian ϕ-process theory.

6. Interpolation Between η_0 and η_∞

For a given normalized code-word rate $\delta = 2/w_L T$, the actual efficiency η_δ lies between η_0 and η_∞. To compute η_δ exactly is an extremely difficult task, since the statistics of ϕ_k required by Eq. (4) are unknown. What we shall develop here is an interpolation formula for η_δ, rather than a direct evaluation of the efficiency. One very good approximation of the error probability in the vicinity of R_0 is obtained by a Taylor expansion of $\ln [P_E(R_1^{1/2})]$:

$$P_E(R_1^{1/2}) \approx P_E(R_0^{1/2})\, e^{\lambda(R_0 - R_1)} \tag{33}$$

for two comparative values R_0 and R_1 of $R = PT/N_+$, for both no-coding and orthogonal/biorthogonal coding, according to the value of λ. We take R_0 to be the value

$$R_0 = R\mu^2 = R\,E^2[g(\phi)] \tag{34}$$

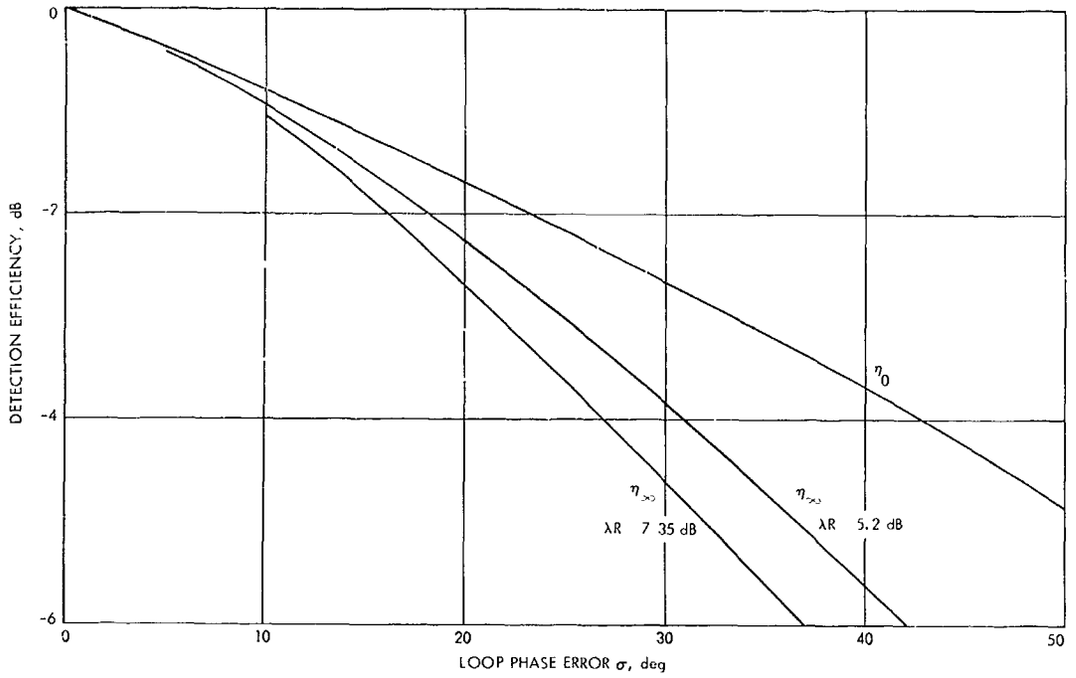


Fig. 20. Loop-derived reference degradation bounds for subcarrier-tone-tracking loop

corresponding to η_0, and take R_1 to be the apparent SNR corresponding to the correct word in Eq. (4):

$$R_1 = R(\mu + \nu)^2 \tag{35}$$

Then we can equate the observed error probability as having occurred with an equivalent value of $R = R_{eq}$, with no reference phase error:

$$\begin{aligned} P_E(R_{eq}^{1/2}) &= \int P_E(R_1^{1/2}|\nu)\, p(\nu)\, d\nu \\ &= P_E(R_0^{1/2}) \int \exp\left\{ -\lambda R_0 \right. \\ &\quad \left. \times \left[\left(\frac{\nu}{\mu}\right)^2 + 2\left(\frac{\nu}{\mu}\right) \right] \right\} p(\nu)\, d\nu \end{aligned} \tag{36}$$

The form of P_E displayed in Eq. (33) then provides

$$\lambda(R_0 - R_1) = \ln \int \exp\left\{ -\lambda R_0 \left[\left(\frac{\nu}{\mu}\right)^2 + 2\left(\frac{\nu}{\mu}\right) \right] \right\} p(\nu)\, d\nu$$

$$\begin{aligned} \lambda R(\eta_0 - \eta_\delta) &= \ln \int \left\{ 1 - \lambda R_0 \left[\left(\frac{\nu}{\mu}\right)^2 + 2\left(\frac{\nu}{\mu}\right) \right] \right. \\ &\quad \left. + \frac{\lambda^2 R_0}{2} \left[\left(\frac{\nu}{\mu}\right)^2 + 2\left(\frac{\nu}{\mu}\right) \right]^2 + \cdots \right\} p(\nu)\, d\nu \end{aligned}$$

$$\begin{aligned} &= \ln \left\{ 1 - \lambda R_0 \left(\frac{\sigma_\nu}{\mu}\right)^2 + 2\lambda^2 R_0^2 \left(\frac{\sigma_\nu}{\mu}\right)^2 + \cdots \right\} \\ &= \left(\frac{\sigma_\nu}{\mu}\right)^2 [2\lambda^2 R_0^2 - \lambda R_0] + \cdots \end{aligned} \tag{37}$$

When σ_ν^2 is small, the first term will dominate the behavior of Eq. (37). Hence, as a result, we see that

$$\frac{\eta_0 - \eta_\delta}{\eta_0 - \eta_\infty} \approx \left(\frac{\sigma_\nu}{\sigma_{\nu,\infty}}\right)^2 = a \tag{38}$$

in which $\sigma_{\nu,\infty}^2$ is the variance of ν as $\delta \to \infty$. Hence, the interpolation formula we seek is

$$\eta_\delta = (1 - a)\,\eta_0 + a\eta_\infty \tag{39}$$

and is valid whenever Eq. (37) is dominated by its first term.

The parameter a defined by Eq. (38) involves only the expectation of the square of

$$\nu = \frac{1}{T} \int_0^T \{ g(\phi(t)) - E[g(\phi)] \}\, dt \tag{40}$$

which is given straightforwardly by (Ref. 5)

$$\sigma_v^2 = \delta \int_0^{2/\delta} \left(1 - \frac{\delta x}{2}\right) [R_{g(\phi)}(x/w_L) - R_{g(\phi)}(\infty)]\, dx \tag{41}$$

The asymptotic values of σ_v^2 at very small and very large δ then verify our previous intuitive claim:

$$\sigma_v^2 \approx \begin{cases} \delta \int_0^{\infty} [R_{g(\phi)}(x/w_L) - R_{g(\phi)}(\infty)]\, dx, & \text{as } \delta \to 0 \\ \sigma_g^2, & \text{as } \delta \to \infty \end{cases} \tag{42}$$

It thus remains only to evaluate σ_v^2 at a particular value of δ. The ratio of the two variances σ_v^2 and σ_η^2 then give the parameter a. But because the loop is nonlinear, $R_{g(\phi)}(\tau)$ is not known, although there are several approximations available for calculation of $R_\phi(\tau)$. We can model $\phi(t)$ as a gaussian process having the same variance and bandwidth as the ϕ-process and thereby evaluate the autocorrelation of $g(\phi)$ in terms of that of $\phi(t)$ by Price's Theorem (Ref. 6). For example, if $g(\phi) = \cos\phi$, then

$$R_{\cos\phi}(\tau) - R_{\cos\phi}(\infty) = 2e^{-\sigma_\phi^2} \sinh^2\left(\frac{1}{2} R_\phi(\tau)\right) \approx \frac{1}{2} e^{-\sigma_\phi^2} R_\phi^2(\tau) \tag{43}$$

while, if $g(\phi)$ is the triangular function present in square-wave subcarrier extraction, then

$$R_{\mathrm{tri}(\phi)}(\tau) - R_{\mathrm{tri}(\phi)}(\infty) = \left(\frac{2}{\pi}\right)^3 \left[R_\phi(\tau) \sin^{-1}(R_\phi(\tau)) \sin^{-1}\left(\frac{R_\phi(\tau)}{\sigma_\phi^2}\right) + (\sigma_\phi^4 - R_\phi^2(\tau))^{1/2} - \sigma_\phi^2 \right] \tag{44}$$

Further, we can model the correlation function of ϕ by the simple first-order loop result

$$R_\phi(\tau) = \sigma_\phi^2 \exp(-2w_L|\tau|) \tag{45}$$

and thereby evaluate the parameter a.

At best, the evaluation of a requires numerical integration. of σ_ϕ^2 is small in the carrier loop case, however, the approximation in Eq. (43) can be used to give

$$a\,(\text{carrier loop}) = \delta \int_0^{2/\delta} \left(1 - \frac{\delta x}{2}\right) e^{-x}\, dx = \frac{\delta}{4}\left[1 - \frac{\delta}{8}(1 - e^{-8/\delta})\right] \tag{46}$$

and is independent of σ_ϕ^2. This a is plotted in Fig. 21. The numerically integrated, more exact value at $\sigma_\phi^2 = 1$ is almost indistinguishable from Eq. (46).

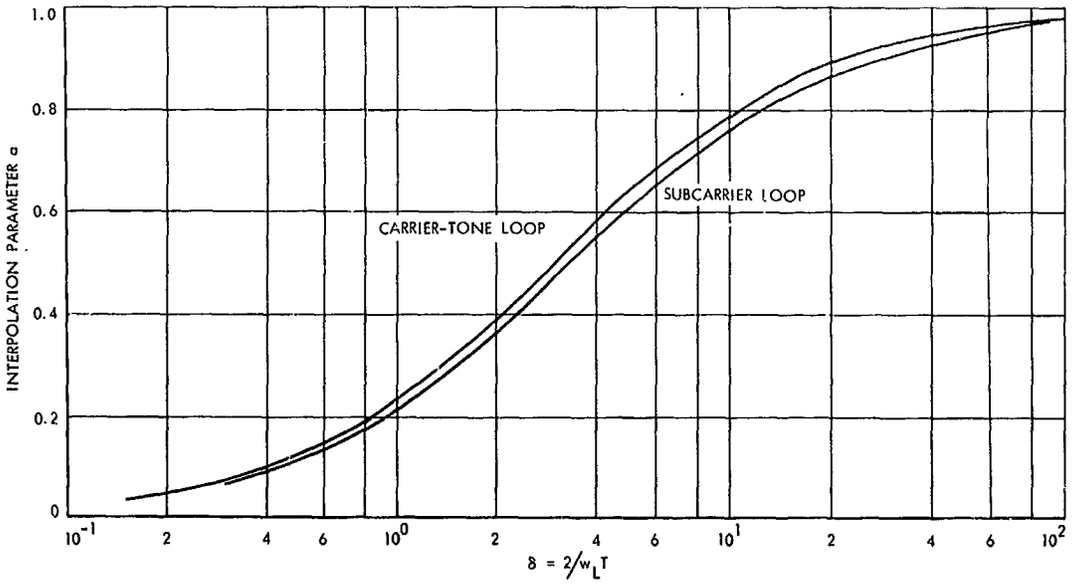


Fig. 21. Interpolation parameter a for sine wave (carrier tone) and square wave (subcarrier tone) loops as a function of the normalized data word rate

In the triangular subcarrier case, a reduces to the integral

$$a\,(\text{subcarrier}) = \frac{\delta}{\pi - 2}\int_0^{4/\delta}\left(1 - \frac{\delta x}{4}\right)\left[e^{-x}\sin^{-1}\left(e^{-x}\right) + (1 - e^{-2x})^{1/2} - 1\right]dx \qquad (47)$$

which is also independent of σ_ϕ^2. The variation of a with δ is depicted also in Fig. 21. The approximate expression

$$a\,(\text{subcarrier loop}) = \frac{0.09135\,\delta + \delta^2}{1 + 3.3718\,\delta + \delta^2} \qquad (48)$$

provides a simple formula for amazingly accurate results.

7. Conclusions

The efficiency of a coherent amplitude detector lies somewhere between limits set by two extreme theories, depending on the value of $\delta = 2/w_L T$. In the discussion we have considered (by assumption of the form $p(\phi)$ only) the effects of wide-band input noise. However, if there were other processes causing phase error, such as loop voltage-controlled oscillator noise, detection instabilities, etc., they can be considered as an equivalent phase-error term to be included in $p(\phi)$.

As long as the loop SNR, ρ, is greater than 10, $p(\phi)$ is very nearly gaussian, and a normal density can be substituted for $p(\phi)$, with the other instabilities reflected in the value of σ^2.

References

1. Lindsey, W. C., *Performance of Phase-Coherent Receivers Preceded by Bandpass Limiters*, Technical Report 32-1162. Jet Propulsion Laboratory, Pasadena, Calif., Sept. 1967.
2. Golomb, S. W., et al., *Digital Communications With Space Applications*, pp. 120–131. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
3. Viterbi, A. J., *Principles of Coherent Communications*, pp. 86–96. McGraw-Hill Book Co., Inc., New York, 1966.
4. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers, Vol. I*, Technical Report 32-819. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1966.
5. Davenport, W. B., and Root, W. L., *Random Signals and Noise*, Sec. 6-5. McGraw-Hill Book Co., Inc., New York, 1958.
6. Price, R., "A Useful Theorem for Nonlinear Device Having Gaussian Inputs," *IRE Trans. Info. Theory*, Vol. IT-4, pp. 69–72, June 1958.

H. Information Processing: Limiters in Phase-Locked Loops: A Correction to Previous Theory, *R. C. Tausworthe*

1. Introduction

In 1953 Davenport (Ref. 1) published a now-classic paper which showed that at very large values, the asymptotic output signal-to-noise ratio (SNR) of a limiter is twice its input SNR. Because of this, it was supposed that the same improvement ultimately should be evident in the performance of a phase-locked loop tracking the limiter output. In fact, the author (Ref. 2) used this result (erroneously, but subtly so) to derive a limiter performance factor Γ. Recently, however, G. D. Forney (Ref. 3) has presented a simple argument to show that the asymptotic factor of 2 is not realized in loop performance, although it is indeed present in output SNR. In this article, the author extends the asymptotic result to rederive the equivalent limiter performance factor.

2. Loop Theory and Noise Components

We shall assume (Fig. 22) that a loop has incident a sinusoid in wide-band noise, and we shall express this process in the form (Ref. 4)

$$y(t) = \alpha\, 2^{1/2}\sin(\omega_0 t + \theta) + n_Q(t)\, 2^{1/2}\sin(\omega_0 t + \hat{\theta}) + n_I(t)\, 2^{1/2}\cos(\omega_0 t + \hat{\theta}) \qquad (1)$$

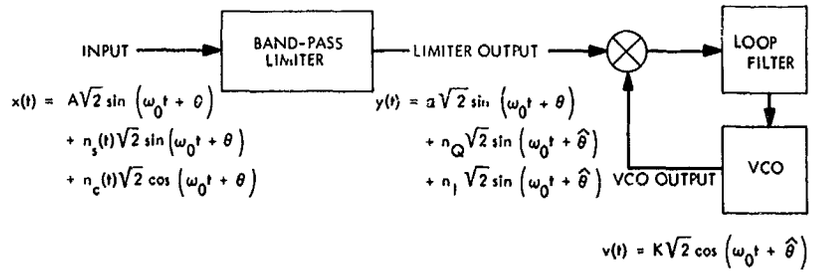


Fig. 22. The bandpass limiter phase-locked loop

where $n_I(t)$ and $n_Q(t)$ are independent wide-band in-phase and quadrature components [with respect to the voltage-controlled oscillator (VCO) output] of the input noise process, α is the rms signal amplitude, and $\hat{\theta}$ is the loop estimate of the input phase process θ. The VCO output we shall take to be of the form

$$v(t) = K 2^{1/2} \cos(\omega_0 t + \hat{\theta}) \tag{2}$$

so that the phase detector output is

$$z(t) = \langle y(t)\, v(t) \rangle = K[\alpha \sin\phi + n_I(t)] \tag{3}$$

with $\phi = \theta - \hat{\theta}$, the loop phase error. The terms enclosed in "$\langle\ \rangle$" refer to the "low-pass part of" the expression.

The important point in Eq. (3) is that the only noise having any effect on loop performance is the noise in phase with the VCO. The resulting linear theoretical phase error is then

$$\sigma_\phi^2 = \frac{S_{n_I n_I}(0)\, w_L}{\alpha^2} \tag{4}$$

where $S_{n_I n_I}(j\omega)$ is the spectral density of the process $n_I(t)$ and w_L is the loop bandwidth (Ref. 2).

In previous analyses, it has always been assumed that the input noise process has components of equal spectral density, each equal to that of the input process. However, *the noise components at the limiter output do not have equal spectra*; this is the point that was overlooked in extending Eq. (4) to the limiter-loop case. Hence, to correct the result published in Ref. 2, the value of $S_{n_I n_I}(0)$ merely has to be reevaluated.

3. Limiter Output Noise

Let the input to an ideal bandpass limiter be

$$x(t) = A 2^{1/2} \sin(\omega_0 t + \theta) + n_s(A) 2^{1/2} \sin(\omega_0 t + \theta) + n_c(t) 2^{1/2} \cos(\omega_0 t + \theta) \tag{5}$$

where A is the rms signal amplitude, and $n_s(t)$ and $n_c(t)$ are independent, zero mean, quadrature gaussian noise processes with respect to the input signal, having identical variances and spectra

$$\left.\begin{aligned} \sigma_{n_c}^2 = \sigma_{n_s}^2 = \sigma_n^2/2 \\ N_{0c} = N_{0s} = N_0 \end{aligned}\right\} \tag{6}$$

It follows directly that the fundamental-zone limiter output is of the form

$$\begin{aligned} y(t) &= \frac{2^{3/2} L}{\pi} \left\{ \left(\frac{[A + n_s(t)]}{A(t)} \right) 2^{1/2} \sin(\omega_0 t + \theta) + \left(\frac{n_c(t)}{A(t)} \right) 2^{1/2} \cos(\omega_0 t + \theta) \right\} \\ &= \frac{2^{3/2} L}{\pi} \{ [\alpha + n_1(t)]\, 2^{1/2} \sin(\omega_0 t + \theta) + n_2(t)\, 2^{1/2} \cos(\omega_0 t + \theta) \} \end{aligned} \tag{7}$$

where L is the limiter clipping level, and $A(t)$ is the process

$$A(t) = [(A + n_s(t))^2 + n_c^2(t)]^{1/2} \tag{8}$$

For convenience, we shall normalize $L = \pi\, 2^{3/2}$ to give unity output power in the fundamental harmonic zone. Previous calculations (Ref. 1 and SPS 37-44, Vol. IV, pp. 303–307) have shown that the mean signal level is

$$\begin{aligned} \alpha^2 &= \frac{\pi}{4} \rho_{II} \exp(-\rho_{II}) [I_0(\rho_{II}/2) + I_1(\rho_{II}/2)]^2 \\ &\approx \frac{0.7854\rho_{II} + 0.4768\rho_{II}^2}{1 + 1.0240\rho_{II} + 0.4768\rho_{II}^2} \end{aligned} \tag{9}$$

in terms of the input SNR, $\rho_{II} = A^2/N_0 w_{II}$, the input process bandwidth w_{II} and the modified Bessel functions of the first kind, $I_n(x)$.

The second term of Eq. (7) is noise whose variance is

$$\sigma_2^2 = E\left\{ \frac{n_c^2(t)}{A^2(t)} \right\} \tag{10}$$

This expectation involves the evaluation of a somewhat tedious, but straightforward, double integral. The result is the simple expression

$$\sigma_2^2 = \frac{1}{2\rho_{II}} [1 - \exp(-\rho_{II})] \tag{11}$$

The noise component $n_1(t)$ appearing in Eq. (7) is uncorrelated with $n_2(t)$; so its variance is

$$\sigma_1^2 = 1 - \alpha^2 - \sigma_2^2 \tag{12}$$

4. Limiter Noise in-Phase With the VCO

The reader will note that the expressions for the loop input $y(t)$ given in Eqs. (1) and (7) are slightly different.

When multiplied by the VCO output in Eq. (2), the relevant loop noise $n_I(t)$ is

$$n_I(t) = n_1(t)\sin\phi + n_2(t)\cos\phi \tag{13}$$

whose variance is then

$$\sigma_I^2 = \sigma_2^2 + (\sigma_1^2 - \sigma_2^2)\,E(\sin^2\phi) \tag{14}$$

5. Phase-Locked Loop Variance

The variance in Eq. (4) can now be evaluated by noting that, according to the definition of the in-phase noise bandwidth w_I,

$$S_{n_I n_I}(0) = \frac{1}{w_I}\int_{-\infty}^{+\infty} S_{n_I n_I}(jw)\,\frac{dw}{2\pi} = \frac{\sigma_I^2}{w_I} \tag{15}$$

and therefore

$$\sigma_\phi^2 = \frac{\sigma_I^2 w_L}{w_I \alpha^2} = \frac{N_0 w_L}{A^2}\left(\frac{w_{II}\rho_{II}}{w_{Ii}\rho_{Ii}}\right)\left(\frac{w_{Ii}\sigma_{Ii}^2}{w_I(1-\alpha^2)}\right) \tag{16}$$

The ratio $w_{Ii}\rho_{Ii}/\rho_{II}w_{II}$ is the Springett factor (SPS 37-36, Vol. IV, pp. 241–244), previously defined as $1/\Gamma$, which relates output and input spectral densities. The only unknown in Eq. (16) is the ratio w_I/w_{Ii}. Asymptotically, as $\rho_{II}\to\infty$, all the output noise comes from $n_2(t)$, so $w_I = w_{Ii}/2$ (the limiter output bandwidth w_{Ii} differs by a factor of 2 from that of the baseband process); and as $\rho_{II}\to 0$, both noises $n_1(t)$ and $n_2(t)$ have the same spectra, so again $w_I = w_{Ii}/2$.

Insertion of the value for σ_{Ii}^2 and the linear-theoretical assumption $\sin^2\phi \approx \phi^2$ yield

$$\sigma_\phi^2 = \frac{N_0 w_L}{A^2}\Gamma\,. \tag{17}$$

in which Γ is the true limiter performance factor

$$\Gamma = \frac{\left(\frac{w_{II}\rho_{II}}{w_{Ii}\rho_{Ii}}\right)\left[\frac{1-\exp(-\rho_{II})}{(1-\alpha^2)\rho_{II}}\right]\left(\frac{w_{Ii}}{w_I}\right)}{1+\left(\frac{2w_L}{w_{II}\rho_{II}}\right)\left(\frac{w_{II}\rho_{II}}{w_{Ii}\rho_{Ii}}\right)\left[\frac{1-\exp(-\rho_{II})}{(1-\alpha^2)2\rho_{II}}-1\right]\left(\frac{w_{Ii}}{w_I}\right)} \tag{18}$$

In limiting cases, Γ behaves as follows:

$$\Gamma \sim \frac{\frac{1}{2}(2)}{1+0} = 1, \qquad \rho_{II}\text{ very large} \tag{19}$$

as predicted by the Forney result, and

$$\Gamma \sim \frac{1.16}{1+0.335\,\frac{w_L}{w_{II}}} < 1.16, \qquad \rho_{II}\text{ very small} \tag{20}$$

The conclusion reached is that *the effect of a limiter on loop performance is less than 16% (0.66 dB) in effective input SNR!* Under the assumption that $w_{II} > 10\,w_L$, the dependence of Γ on w_L/w_{II} disappears for practical purposes (within 0.1 dB). A rational function approximation emulating the cross-over behavior of Eq. (18) near $\rho = 0.5$ (Fig. 23), is

$$\Gamma = \frac{1+\rho_{II}}{0.862+\rho_{II}} \tag{21}$$

With this approximation, the linear-theoretical limiter performance follows directly.

6. Nonlinear Loop Theory

The only alteration that must be made to make the derivation of Γ general enough for the linear-spectral loop theory (Ref. 2, Chapt. 9) is the estimation of $E(\sin^2\phi)$ in Eq. (14) by a more accurate value. But the only change which would result in Γ would be a factor

$$\frac{E[\sin^2\Phi]}{E[\Phi^2]} = \frac{(1-\exp(-2\sigma_\Phi^2))}{2\sigma_\Phi^2} \leq 1 \tag{22}$$

(under the assumptions of Ref. 2, Chapt. 9) inserted into the second denominator term of Eq. (18), which we have agreed, for $w_{II} > 10\,w_L$, is negligible anyway. Hence, the nonlinear theory is also essentially unaltered from the previously published results, except for the new Γ that must be used.

References

1. Davenport, W. B., "Signal-to-Noise Ratios in Band-Pass Limiters," *Journal of Applied Physics*, Vol. 24, No. 6, pp. 720–727, June 1953.
2. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers*, Technical Report 32-819. Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1966.
3. Forney, G. D., *Coding in Coherent Deep-Space Telemetry*. McDonnell Douglas Astronautics Corp., Santa Monica, Calif., March 19, 1967.
4. Davenport, W. B., and Root, W. L., *Random Signals and Noise*, pp. 158–165. McGraw-Hill Book Co., Inc., New York, 1958.

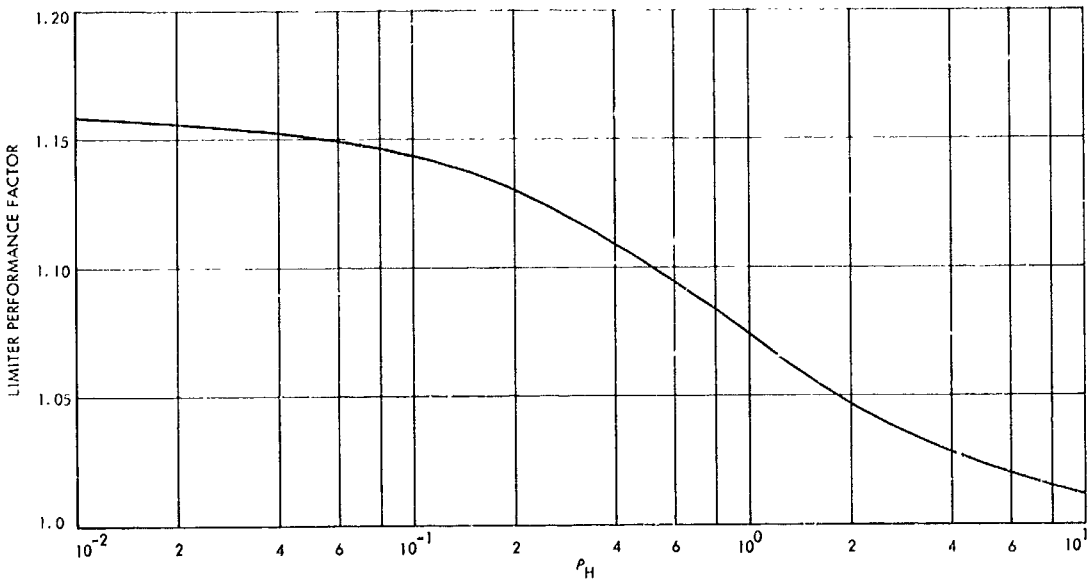


Fig. 23. Variation of the parameter Γ with limiter input SNR, $\rho_{II} = A^2/N_0 w_L$

XXI. Communications Elements Research

TELECOMMUNICATIONS DIVISION

A. System Studies for 90-GHz Space Communications Systems: Atmospheric Effects on Millimeter Wave Propagation, *T. Sato*

The 60-in. radio telescope used for radio astronomy experiments (SPS 37-49, Vol. III, pp. 329–338) is being converted into a radio sextant. An optical tracker automatically tracks the sun in both hour angle and declination by generating error signals for the antenna servo. The error-sensing elements attached to the telescope consist of two photoconducting cells per channel. When the sun is centered, both cells are illuminated equally and produce equal resistances; no error signal is then produced. If the sun's image shifts, the cells are illuminated unequally, thereby producing an error signal. The error signal to the digital processor generates pulses that are fed to the stepping motor in a direction to remove the angular error in pointing.

The 90-GHz nodding subdish radiometer mounted on the 60-in. radio telescope receives emissions from the sun. The slant path through the atmosphere changes with the position of the sun. The temperature of the antenna when pointed at the sun is given by

$$T_a(\phi) = T'_a 10^{-a_0/10 \sec\phi} + T_s(\phi)$$

where T'_a is the antenna temperature with no intervening atmosphere, a_0 is the zenith attenuation in decibels, ϕ is the zenith angle, and $T_s(\phi)$ is the sky temperature. Use of the nodding subdish radiometer balances out the $T_s(\phi)$ component. The expression for the received temperature is then given by

$$\log T_a(\phi) = -\frac{a_0}{10 \sec\phi} + \log T'_a$$

The absolute zenith attenuation can then be determined from measurements of the received power and the zenith angle without knowledge of the solar temperature or antenna gain. Certain weather conditions may invalidate the secant law giving an average value for the zenith attenuation.

B. Spacecraft Antenna Research: RF Breakdown Characteristics of S-Band Sterilizable High-Impact Antennas, *K. Woo*

The power-handling capabilities of the S-band sterilizable high-impact antennas (coaxial cavity,[1] cupped

[1]SPS 37-40, Vol. IV: discussion on pp. 201–206; photograph in Fig. 5, p. 202.

turnstile,[2] square cup,[3] and circular cup[4]) have been investigated at very low pressures. The antennas were each tested in the vacuum chamber of the JPL voltage breakdown facility with dry air, 100% carbon dioxide, and a mixture of 50% carbon dioxide and 50% argon.

The ionization breakdown power level at 2298 MHz of each antenna was obtained as a function of pressure near and at the point where the breakdown power is minimum, such as shown in Fig. 1 (obtained for the coaxial cavity radiator). The minimum ionization breakdown power levels at 2298 MHz for all the antennas were found within the pressure range of 1.5 to 2.5 torr. The multipacting breakdown power levels at 2298 MHz of all the antennas were tested around 5×10^{-5} torr. The multipacting breakdown and minimum ionization breakdown power levels for all the antennas are summarized in Table 1.

Each value given in Table 1 represents the power that each antenna actually received, i.e., the power fed into the antenna minus the power reflected back due to mismatch. The values to which ">" signs are attached represent the peak levels to which those antennas were tested due to power limitations imposed by either the output of the RF generator or the feeding hybrid (if used) or by other difficulties.

[2]SPS 37-42, Vol. IV: discussion on pp. 180, 181; photograph in Fig. 1, p. 181.

[3]SPS 37-49 and -53, Vol. III: discussion on pp. 345–347 (37-49) and pp. 164–166 (37-53); photographs in Figs. 32, p. 346 (37-49) and 5, p. 165 (37-53).

[4]SPS 37-47, Vol. III: discussion on pp. 240–242; photograph in Fig. 4, p. 240 and Fig. 5, p. 241.

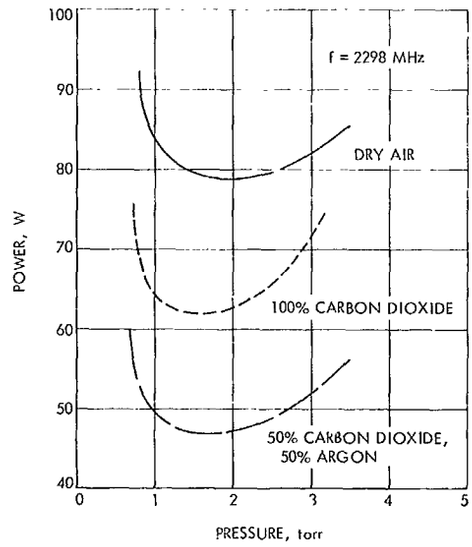


Fig. 1. Ionization breakdown characteristic of coaxial cavity radiator

The power levels for the square cup and circular cup radiators represent only momentary power; i.e., the power was shut off immediately after it was applied to those antennas. When a steady power was applied, those antennas experienced a thermal breakdown (hence, degradation of electrical performance) at relatively low power levels. The thermal breakdown was in the form of charring of the foam in the vicinity of the feed probe due to the accumulation of heat resulting from the poor thermal conductivity of the foam. Figure 2 shows the charred area around the feed of the square cup radiator after removal of the foam. Test results at 2.0 torr showed that a square cup radiator potted with Stafoam AA630 would sustain at least 3 h of continuous operation at the 50-W level, but would suffer thermal breakdown (voltage standing-wave

Table 1. Ionization and multipacting breakdown power levels and locations for the S-band sterilizable high-impact antennas[a]

Antenna	Ionization breakdown				Multipacting breakdown			
	Minimum breakdown power, W, for indicated atmosphere			Location of breakdown	Breakdown power, W, for indicated atmosphere			Location of breakdown
	A	B	C		A	B	C	
Coaxial cavity	79	62	47	Around the probes	> 130	> 130	> 130	None observed
Cupped turnstile	31	23	17	At balun split	> 130	> 130	> 130	None observed
Square cup[b]	> 560	> 560	> 560	None observed	> 560	> 560	> 560	None observed
Circular cup[b]	> 560	> 560	> 560	None observed	> 560	> 560	> 560	None observed

[a]Atmosphere A: dry air.
Atmosphere B: 100% carbon dioxide.
Atmosphere C: 50% carbon dioxide, 50% argon.
[b]Potted with Stafoam AA630 or Eccofoam PT.

Fig. 2. Square cup radiator after thermal breakdown

ratio changed from 1.8 to 4.6) after 5 min of operation at the 100-W level. The same antenna potted with Eccofoam PT would sustain at least 3 h of continuous operation at the 40-W level.

On the basis of the test results, it can be seen that, for high-power operation, the ionization breakdown is the problem area for the coaxial cavity and cupped turnstile antennas, and the thermal breakdown is the problem area for the square cup and circular cup antennas. To improve the power-handling capabilities of the existing antennas, the following modifications are recommended:

Antenna	Recommendation
Coaxial cavity	Widen cavity and enclose feed probes with high-thermal-conductivity dielectric
Cupped turnstile	Redesign balun
Square cup Circular cup	Improve thermal conductivity and reduce loss of potting dielectric

C. Radiation From a Horizontal Electric Dipole Antenna Located in a Cylindrically Stratified Plasma, *R. Woo*

1. Introduction

In SPS 37-47, Vol. III, pp. 247–257, the wake of a Mars entry capsule was represented by a cylindrical plasma column and the problem of a horizontal electric dipole located in the wake was formulated. In SPS 37-50, Vol. III, pp. 312–316, the representation of the wake was improved by surrounding the plasma column with a plasma shell of different electron density than that of the column. Here the wake will be represented by a cylindrically stratified plasma. The problem of a horizontal electric dipole located in the wake will be formulated using a scheme that readily lends itself to computer calculation.

2. Integral Formulation of Fields for Horizontal Dipole

The regions are numbered i, where $i=1, 2, 3, \cdots, N+1$. Region $N+1$ is free space.

The free-space region $i = N + 1$ is characterized by permittivity ϵ_0 and permeability μ_0. According to the magneto-ionic theory description for a homogeneous, isotropic, and lossless plasma, the plasma region i is represented by a medium whose permittivity ϵ_i is given by (Ref. 1)

$$\epsilon_i = \left[1 - \left(\frac{\omega_{pi}}{\omega}\right)^2\right]\epsilon_0 \tag{1}$$

where

$$\omega_{pi}^2 = \frac{n_i e^2}{m\epsilon_0} \tag{2}$$

with n_i representing the electron density, e the electron charge, m the electron mass, and ω_{pi} the electron plasma frequency. The propagation constant of region i is then given by

$$k_i^2 = \omega^2\mu_0\epsilon_i$$

Assuming a time dependence of $e^{j\omega t}$, the respective electric and magnetic field intensities, **E** and **H**, are related to the electric (π) and magnetic (π^*) Hertz potentials through the following equations:

$$\mathbf{E} = \nabla \times \nabla \times \boldsymbol{\pi} - j\omega\mu_0 \nabla \times \boldsymbol{\pi}^* \tag{3}$$

$$\mathbf{H} = \nabla \times \nabla \times \boldsymbol{\pi}^* + j\omega\epsilon \nabla \times \boldsymbol{\pi} \tag{4}$$

where ϵ is the intrinsic permittivity of the medium.

The electric dipole of moment Il is assumed to be located at (ρ_0, ϕ_0, z_0) and oriented in the x direction in plasma region m (Fig. 3). The x component of π, π_{xm}, will therefore represent the primary field and satisfy the following equation:

$$(\nabla^2 + k_m^2)\,\pi_{xm} = -\frac{Il}{j\omega\epsilon_m}\,\delta(\phi - \phi_0)\,\delta(z - z_0)\,\frac{\delta(\rho - \rho_0)}{\rho} \tag{5}$$

where δ is the Dirac delta function and the subscript m refers to the region m. Since both E_z and H_z will be present in the secondary fields, the respective z components of π and π^*, π_{zi} and π^*_{zi}, will be used for region i and will satisfy the following wave equations:

$$(\nabla^2 + k_i^2)\,\pi_{zi} = 0 \tag{6}$$

$$(\nabla^2 + k_i^2)\,\pi^*_{zi} = 0 \tag{7}$$

Integral representations for the Hertz potentials can be written using a Fourier integral in the z direction and a Fourier series in the ϕ direction (Ref. 2). For the primary field in region m,

$$\pi_{xm} = \frac{Il}{j\omega\epsilon_m}\,\frac{1}{8\pi j} \times \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} H_n^{(2)}(\lambda_m\rho)\,J_n(\lambda_m\rho_0)\,e^{-jn(\phi-\phi_0)}\,e^{-jh(z-z_0)}\,dh \tag{8}$$

where

$$\lambda_m^2 = k_m^2 - h^2$$

J_n is the nth-order Bessel function of the first kind and $H_n^{(2)}$ is the nth-order Hankel function of the second kind. Equation (8) is for $\rho > \rho_0$; when $\rho < \rho_0$, ρ and ρ_0 must be interchanged in the equation. For the secondary fields, the following results apply for region i:

$$\pi_{zi} = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} [a_n^{(i)} J_n(\lambda_i\rho) + c_n^{(i)} H_n^{(2)}(\lambda_i\rho)] \times e^{-jn\phi}\,e^{-jh(z-z_0)}\,dh \tag{9}$$

$$\pi^*_{zi} = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} [b_n^{(i)} J_n(\lambda_i\rho) + d_n^{(i)} H_n^{(2)}(\lambda_i\rho)] \times e^{-jn\phi}\,e^{-jh(z-z_0)}\,dh \tag{10}$$

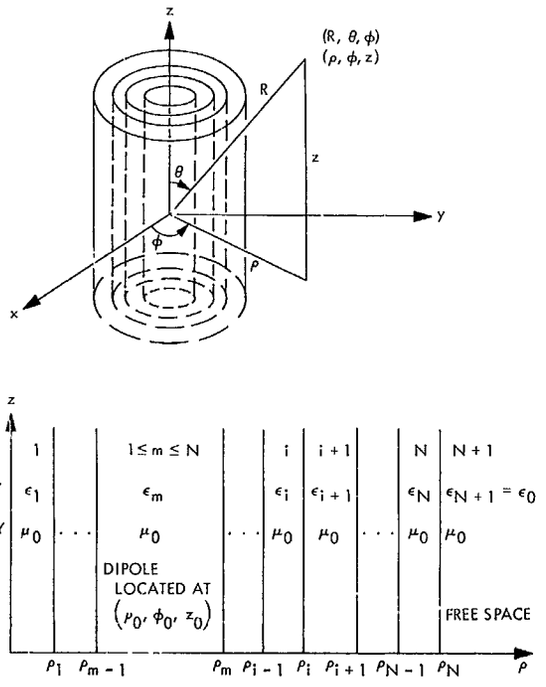


Fig. 3. Geometry of the problem

where

$$\lambda_i^2 = k_i^2 - h^2$$

The superscript i in the coefficients a_n, b_n, c_n, and d_n refers to the fields in region i. Since the fields are finite at $\rho = 0$, $c_n^{(1)} = d_n^{(1)} = 0$. Also, since outgoing waves are in the free-space region $i = N + 1$, $a_n^{(N+1)} = b_n^{(N+1)} = 0$.

The cylindrical components of **E** and **H** for region i can be obtained using Eqs. (3) and (4).

$$E_{zi} = \left(\frac{\partial^2}{\partial z^2} + k_i^2\right)\pi_{zi} \tag{11}$$

$$E_{\phi i} = \frac{1}{\rho}\,\frac{\partial^2 \pi_{zi}}{\partial\phi\,\partial z} + j\omega\mu_0\,\frac{\partial \pi^*_{zi}}{\partial\rho} \tag{12}$$

$$E_{\rho i} = \frac{\partial^2 \pi_{zi}}{\partial\rho\,\partial z} - \frac{j\omega\mu_0}{\rho}\,\frac{\partial \pi^*_{zi}}{\partial\phi} \tag{13}$$

$$H_{zi} = \left(\frac{\partial^2}{\partial z^2} + k_i^2\right)\pi^*_{zi} \tag{14}$$

$$H_{\phi i} = \frac{1}{\rho}\,\frac{\partial^2 \pi^*_{zi}}{\partial\phi\,\partial z} - j\omega\epsilon_i\,\frac{\partial \pi_{zi}}{\partial\rho} \tag{15}$$

$$H_{\rho i} = \frac{\partial^2 \pi^*_{zi}}{\partial\rho\,\partial z} + \frac{j\omega\epsilon_i}{\rho}\,\frac{\partial \pi_{zi}}{\partial\phi} \tag{16}$$

The following are additional components in region m due to the primary field:

$$E_{zm} = \frac{\partial^2 \pi_{xm}}{\partial z\, \partial \rho} \cos\phi - \frac{1}{\rho} \frac{\partial^2 \pi_{xm}}{\partial z\, \partial \phi} \sin\phi \tag{17}$$

$$E_{\phi m} = \left(\frac{\partial^2 \pi_{xm}}{\partial z^2} + \frac{\partial^2 \pi_{xm}}{\partial \rho^2} \right) \sin\phi + \frac{\partial}{\partial \rho} \left(\frac{1}{\rho} \frac{\partial \pi_{xm}}{\partial \phi} \right) \cos\phi \tag{18}$$

$$E_{\rho m} = \left(\frac{1}{\rho^2} \frac{\partial \pi_{xm}}{\partial \phi} - \frac{1}{\rho} \frac{\partial^2 \pi_{xm}}{\partial \phi\, \partial \rho} \right) \sin\phi - \left(\frac{1}{\rho} \frac{\partial \pi_{xm}}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 \pi_{xm}}{\partial \phi^2} + \frac{\partial^2 \pi_{xm}}{\partial z^2} \right) \cos\phi \tag{19}$$

$$H_{zm} = -j\omega\epsilon_m \frac{\partial \pi_{xm}}{\partial \rho} \sin\phi - j\omega\epsilon_m \frac{1}{\rho} \left(\frac{\partial \pi_{xm}}{\partial \phi} \right) \cos\phi \tag{20}$$

$$H_{\phi m} = j\omega\epsilon_m \frac{\partial \pi_{xm}}{\partial z} \cos\phi \tag{21}$$

$$H_{\rho m} = j\omega\epsilon_m \frac{\partial \pi_{xm}}{\partial z} \sin\phi \tag{22}$$

The boundary conditions at ρ_i require that the tangential **E** and **H** fields be continuous. Imposing these boundary conditions at each boundary ρ_i yields the following results in matrix form:

$$\begin{bmatrix} 0 \\ 0 \\ c_n^{(N+1)} \\ d_n^{(N+1)} \end{bmatrix} = M_n^{(N)} M_n^{(N-1)} \cdots M_n^{(1)} \begin{bmatrix} a_n^{(1)} \\ b_n^{(1)} \\ 0 \\ 0 \end{bmatrix} - M_n^{(N)} M_n^{(N-1)} \cdots M_n^{(m+1)} \left\{ [B_n^{(m)}]^{-1} \begin{bmatrix} \gamma_n^{(m)} \\ \eta_n^{(m)} \\ \alpha_n^{(m)} \\ \beta_n^{(m)} \end{bmatrix} - M_n^{(m)} [B_n^{(m-1)}]^{-1} \begin{bmatrix} \gamma_n^{(m-1)} \\ \eta_n^{(m-1)} \\ \alpha_n^{(m-1)} \\ \beta_n^{(m-1)} \end{bmatrix} \right\} \tag{23}$$

where

$$M_n^{(i)} = [B_n^{(i)}]^{-1} A_n^{(i)}$$

$$A_n^{(i)} = \begin{bmatrix} \lambda_i^2 J_n(\lambda_i \rho_i) & 0 & \lambda_i^2 H_n^{(2)}(\lambda_i \rho_i) & 0 \\ 0 & \lambda_i^2 J_n(\lambda_i \rho_i) & 0 & \lambda_i^2 H_n^{(2)}(\lambda_i \rho_i) \\ \dfrac{-nh}{\rho_i} J_n(\lambda_i \rho_i) & j\omega\mu_0 \lambda_i J_n'(\lambda_i \rho_i) & \dfrac{-nh}{\rho_i} H_n^{(2)}(\lambda_i \rho_i) & j\omega\mu_0 \lambda_i H_n^{(2)\prime}(\lambda_i \rho_i) \\ -j\omega\epsilon_i \lambda_i J_n'(\lambda_i \rho_i) & \dfrac{-nh}{\rho_i} J_n(\lambda_i \rho_i) & -j\omega\epsilon_i \lambda_i H_n^{(2)\prime}(\lambda_i \rho_i) & \dfrac{-nh}{\rho_i} J_n(\lambda_i \rho_i) \end{bmatrix}$$

$B_n^{(i)}$ is similar to $A_n^{(i)}$ except that λ_i and ϵ_i are replaced by λ_{i+1} and ϵ_{i+1}, respectively; and $[B_n^{(0)}]^{-1} = 0$.

$$\gamma_n^{(m)} = -\frac{Il}{4\pi} \omega\mu_0 \frac{jh\lambda_m}{4k_m^2} \left[e^{j(n+1)\phi_0} J_{n+1}(\lambda_m \rho_0) - e^{j(n-1)\phi_0} J_{n-1}(\lambda_m \rho_0) \right] H_n^{(2)}(\lambda_m \rho_m)$$

$$\eta_n^{(m)} = -\frac{Il}{4\pi} \omega\mu_0 \frac{\lambda_m}{4\omega\mu_0} \left[e^{j(n+1)\phi_0} J_{n+1}(\lambda_m \rho_0) + e^{j(n-1)\phi_0} J_{n-1}(\lambda_m \rho_0) \right] H_n^{(2)}(\lambda_m \rho_m)$$

$$\alpha_n^{(m)} = \frac{Il}{4\pi} \omega\mu_0 \frac{j}{8k_m^2} \left\{ \left[(k_m^2 + h^2) e^{j(n+1)\phi_0} J_{n+1}(\lambda_m \rho_0) + \lambda_m^2 e^{j(n-1)\phi_0} J_{n-1}(\lambda_m \rho_0) \right] H_{n+1}^{(2)}(\lambda_m \rho_m) \right.$$
$$\left. - \left[(k_m^2 + h^2) e^{j(n+1)\phi_0} J_{n-1}(\lambda_m \rho_0) + \lambda_m^2 e^{j(n+1)\phi_0} J_{n+1}(\lambda_m \rho_0) \right] H_{n-1}^{(2)}(\lambda_m \rho_m) \right\}$$

$$\beta_n^{(m)} = \frac{Il}{4\pi} \omega\mu_0 \frac{h}{4\omega\mu_0} \left[e^{j(n+1)\phi_0} J_{n+1}(\lambda_m \rho_0) H_{n+1}^{(2)}(\lambda_m \rho_m) + e^{j(n-1)\phi_0} J_{n-1}(\lambda_m \rho_0) H_{n-1}^{(2)}(\lambda_m \rho_m) \right]$$

The quantities $\gamma_n^{(m-1)}$, $\eta_n^{(m-1)}$, $\alpha_n^{(m-1)}$ and $\beta_n^{(m-1)}$ are obtained from $\gamma_n^{(m)}$, $\eta_n^{(m)}$, $\alpha_n^{(m)}$ and $\beta_n^{(m)}$, respectively, by changing ρ_0 to ρ_{m-1} and ρ_m to ρ_0. Equality (23) consists of four equations and four unknown coefficients and can be readily solved. The remaining coefficients are obtained through the following recursion relationship:

$$\begin{bmatrix} a_n^{(i+1)} \\ b_n^{(i+1)} \\ c_n^{(i+1)} \\ d_n^{(i+1)} \end{bmatrix} = M_n^{(i)} \begin{bmatrix} a_n^{(i)} \\ b_n^{(i)} \\ c_n^{(i)} \\ d_n^{(i)} \end{bmatrix} - [B_n^{(m)}]^{-1} \begin{bmatrix} \gamma_n^{(m)} \\ \eta_n^{(m)} \\ \alpha_n^{(m)} \\ \beta_n^{(m)} \end{bmatrix} \delta(i-m) + [B_n^{(m-1)}]^{-1} \begin{bmatrix} \gamma_n^{(m-1)} \\ \eta_n^{(m-1)} \\ \alpha_n^{(m-1)} \\ \beta_n^{(m-1)} \end{bmatrix} \delta(i-m+1) \tag{24}$$

3. Radiation Patterns for Horizontal Dipole

The radiation fields are obtained by using the method of saddle-point integration (Ref. 3) to evaluate the integral expressions of the fields in free space. The resulting spherical components of the radiation fields are

$$\left.\begin{aligned} E_\theta &= -2k^2 \sin\theta\, e^{jkz_0\cos\theta} \frac{e^{-jkR}}{R} \\ &\quad \times \sum_{n=-\infty}^{\infty} c_n^{(N+1)}\, e^{-jn\phi + j(n+1)\pi/2} \\ H_\phi &= \left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E_\theta \end{aligned}\right\} \tag{25}$$

$$\left.\begin{aligned} E_\phi &= 2\omega\mu_0 k \sin\theta\, e^{jkz_0\cos\theta} \frac{e^{-jkR}}{R} \\ &\quad \times \sum_{n=-\infty}^{\infty} d_n^{(N+1)}\, e^{-jn\phi + j(n+1)\pi/2} \\ H_\theta &= -\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E_\phi \end{aligned}\right\} \tag{26}$$

where k is the free-space wave number, $c_n^{(N+1)}$ and $d_n^{(N+1)}$ are given in Eq. (23), and

$$\lambda_i^2 = k_i^2 - k^2\cos^2\theta$$
$$h = k\cos\theta$$

It should be noted that the corresponding modified Bessel functions must be used when $\lambda_i^2 < 0$.

References

1. Wait, J. R., *Electromagnetics and Plasmas.* Holt, Rinehart and Winston, New York, 1968.

2. Wait, J. R., *Electromagnetic Radiation from Cylindrical Structures.* Pergamon Press, New York, 1959.

3. Collin, R. E., *Field Theory of Guided Waves.* McGraw-Hill Book Co., Inc., New York, 1960.

XXII. Spacecraft Telemetry and Command

TELECOMMUNICATIONS DIVISION

A. Approximate Analysis of Channel Imbalance Effects in Non-Coherent FSK Receivers With Large BT Products, C. Carl

1. Introduction

Binary non-coherent frequency-shift-keyed (FSK) systems with large IF bandwidth-to-bit-rate ratios, or BT, are being considered for applications where simplicity and reliability of operation are worth the cost of decreased efficiency compared with, for example, phase-shift-keyed (PSK) modulation techniques. Such an application is the planetary capsule-spacecraft relay link. The typical receiver is shown in Fig. 1. Either frequency f_1 or f_2 is sent, corresponding to a data 1 or 0, square law detected, integrated over the bit interval T, and detected as a data 1 or 0 (bit synchronization at the data detector is assumed). For receiver design purposes, the effect on performance must be determined when the two channels (at f_1 and f_2) are unbalanced with respect to bandwidth, carrier attenuation, etc., so that the appropriate hardware tolerances may be specified. The effect of ac-coupling the data detector must also be considered.

2. Analysis

The analysis involves a straightforward expansion of Glenn's (Ref. 1) work on the balanced receiver. As in Ref. 1, the central-limit theorem is invoked as an approximation for $N = BT \geq 10$; i.e., the statistics at the output of the integrator approach gaussian for large values of N.[1]

The receiver model to be analyzed is shown in Fig. 2, where B_i, α_i, and A_i are the channel bandwidths, carrier attenuation, and channel attenuations, respectively; i.e.,

$$B_i = \frac{\int_{-\infty}^{\infty} |H_i(f)|^2 \, df}{|H_{i_{\max}}|^2}$$

$$\alpha_i = \frac{|H_i(f_i)|^2}{|H_{i_{\max}}|^2}$$

A "balancing amplifier" has been added to channel 2 to ascertain whether the imbalance may be minimized by adjustment of the gain. Let the integrator–detector be dc-coupled for the present. The error probability will be evaluated by calculating the conditional error probabilities, given that a *one* (hypothesis H_1) or a *zero* (H_2) is sent, and averaging them with respect to H_i.

[1]Subsequent work has shown this assumption to be pessimistic in performance by a few tenths of a decibel (to be published).

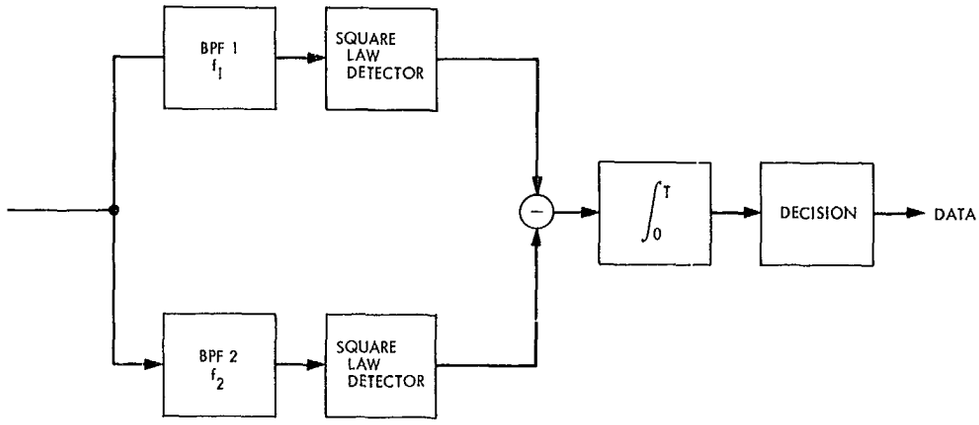


Fig. 1. Typical non-coherent FSK receiver

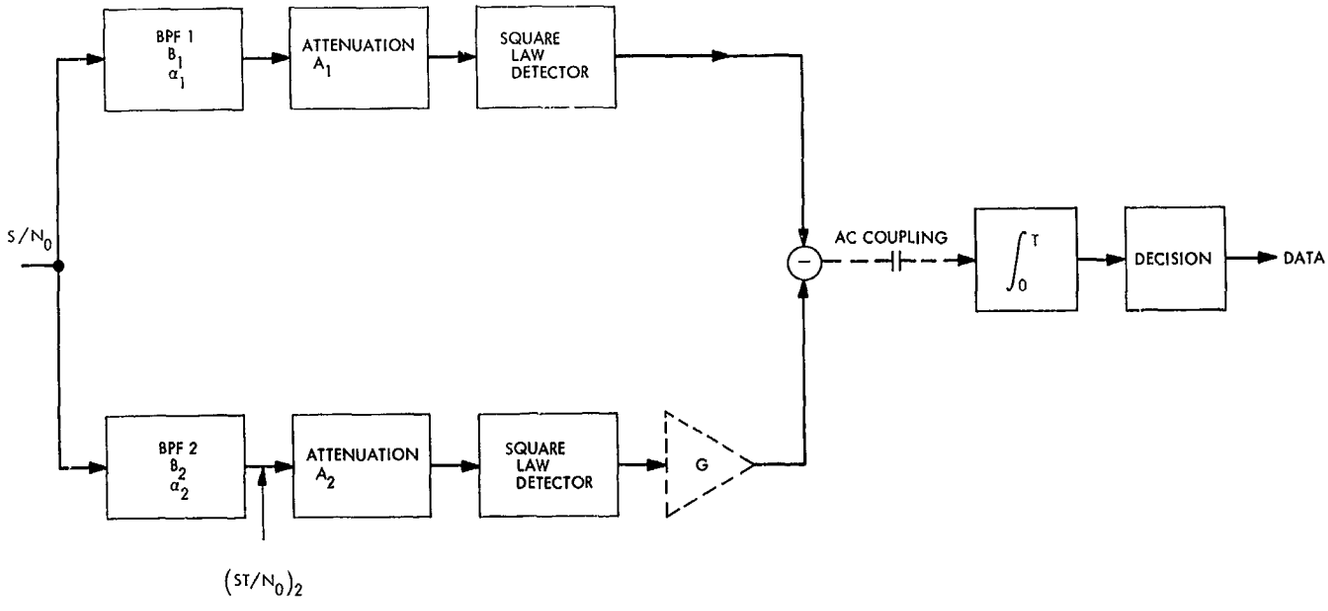


Fig. 2. FSK receiver model

Marcum (Ref. 2) has shown that the probability density of a sum of N samples (i.e., integral) of the output of a square law detector, when the input is signal-plus-noise and noise only, is

$$\left.\begin{aligned} p_{s+n}(y) &= \left(\frac{y}{Na}\right)^{N-1/2} e^{-(y+Na)}\, I_{N-1}\left[2(Nay)^{1/2}\right], \\ p_n(y) &= \frac{y^{N-1}e^{-y}}{(N-1)!}, \end{aligned}\right\} y > 0$$

and the required moments are

$$\begin{aligned} \mu_{s+n} &= N(1+a) \\ \sigma^2_{s+n} &= N(1+2a) \\ \mu_n &= N \\ \sigma^2_n &= N \end{aligned}$$

where y is normalized to $2BN_0$ and $a = S/(N_0B)$; S/N_0 is the signal-to-noise density ratio at the receiver input.

Removing the normalization (since B is not the same for both channels) and assuming $H = H_1$, the moments at the integrator output due to each channel separately, neglecting the subtractor, are

$$\mu_1 = 2B_1^2 TN_0A_1\left(1 + \frac{\alpha_1 S}{N_0B_1}\right)$$

$$\sigma_1^2 = 4B_1^3 TN_0^2A_1^2\left(1 + \frac{2\alpha_1 S}{N_0B_1}\right)$$

$$\mu_2 = 2B_2^2 TN_0A_2G$$

$$\sigma_2^2 = 4B_2^3 TA_2^2N_0^2G^2$$

where G is the amplifier gain. Combining the channels through the subtractor, the probability of error, with the gaussian assumption, is

$$P(e \mid H_1) = \frac{1}{2}\left\{1 - \text{erf}\left[\frac{1}{2}\frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}\right]^{1/2}\right\}$$

where

$$\text{erf}\,(x) = \frac{2}{\pi^{1/2}}\int_0^x e^{-t^2}\,dt$$

After some algebra,

$$P(e \mid H_1) = \frac{1}{2}\left(1 - \text{erf}\left\{\frac{1}{2}\frac{\left[N_1^2A_1 - N_2^2A_2G + A_1N_1\left(\frac{\alpha_1 ST}{N_0}\right)\right]^2}{N_1^3A_1^2 + N_2^3A_2^2G^2 + A_1^2N_1^2\left(\frac{2\alpha_1 ST}{N_0}\right)}\right\}^{1/2}\right) \tag{1a}$$

where $N_i = B_iT$. Similarly for $H = H_2$, the probability of error may be evaluated to obtain

$$P(e \mid H_2) = \frac{1}{2}\left(1 - \text{erf}\left\{\frac{1}{2}\frac{\left[N_2^2A_2G - N_1^2A_1 + N_2A_2G\left(\frac{\alpha_2 ST}{N_0}\right)\right]^2}{N_1^3A_1^2 + N_2^3A_2^2G^2 + G^2A_2^2N_2^2\left(\frac{2\alpha_2 ST}{N_0}\right)}\right\}^{1/2}\right) \tag{1b}$$

Since $P(H_1) = P(H_2) = \frac{1}{2}$, the average probability of error is

$$P(e) = \frac{1}{2}P(e \mid H_1) + \frac{1}{2}P(e \mid H_2) \tag{2}$$

3. Choice of Amplifier Gain

This amplification may be thought of as an operation that is compensating for *some* of the receiver imbalance by improving the detector efficiency (i.e., approaching a maximum likelihood decision rule). One might speculate on "optimum" values of G that minimize $P(e)$, which in general would be functions of ST/N_0 and therefore be difficult to mechanize in hardware. As a practical matter then, G will be selected by some simple method to determine what improvements in performance (if any) that may be achieved.

First, for $G = 1$ (the unbalanced receiver), the probability of error from Eq. (2) becomes

$$P(e) = \frac{1}{2}\left(2 - \text{erf}\left\{\frac{1}{2}\frac{\left[N_1^2R - N_2^2 + RN_2\Gamma\left(\frac{ST}{N_0}\right)_2\right]^2}{N_1^3R^2 + N_2^3 + 2\Gamma R^2N_1^2\left(\frac{ST}{N_0}\right)_2}\right\}^{1/2} - \text{erf}\left\{\frac{1}{2}\frac{\left[N_2^2 - N_1^2R + N_2\left(\frac{ST}{N_0}\right)_2\right]^2}{R^2N_1^3 + N_2^3 + 2N_2^2\left(\frac{ST}{N_0}\right)_2}\right\}^{1/2}\right) \tag{3}$$

where $(ST/N_0)_2$ is that ratio in channel 2 bandpass; i.e.,

$$\left(\frac{ST}{N_0}\right)_2 = \frac{\alpha_2 ST}{N_0}$$

and

$$\Gamma = \frac{\alpha_1}{\alpha_2}$$

$$R = \frac{A_1}{A_2}$$

As the second case, let G be chosen such that the signal-only output from each channel are equal or

$$\mu_1 \big|_{H_1} = \mu_2 \big|_{H_2}, \qquad N_0 \to 0$$

Hence,

$$2B_1^2 T N_0 A_1 \left(1 + \frac{\alpha_1 S}{N_0 B_1}\right) \underset{N_0 \to 0}{=} 2B_2^2 T N_0 A_2 G \left(1 + \frac{\alpha_2 S}{B_2 N_0}\right)$$

As $N_0 \to 0$,

$$G = \frac{B_1 \alpha_1 A_1}{B_2 \alpha_2 A_2} = \frac{N_1}{N_2} R\Gamma$$

Substituting this into Eq. (2),

$$P(e) = \frac{1}{4}\left(2 - \operatorname{erf}\left\{\frac{1}{2}\frac{\left[N_1 - N_2\Gamma + \Gamma\left(\frac{ST}{N_0}\right)_2\right]^2}{N_1 + N_2\Gamma^2 + 2\Gamma\left(\frac{ST}{N_0}\right)_2}\right\}^{1/2} - \operatorname{erf}\left\{\frac{1}{2}\frac{\left[N_2\Gamma - N_1 + \Gamma\left(\frac{ST}{N_0}\right)_2\right]^2}{N_1 + N_2\Gamma^2 + 2\Gamma^2\left(\frac{ST}{N_0}\right)_2}\right\}^{1/2}\right) \tag{4}$$

Finally, let G be chosen such that the variance of the two conditional detection densities be equal; i.e.,

$$\sigma_1^2 + \sigma_2^2\big|_{H_1} = \sigma_1^2 + \sigma_2^2\big|_{H_2}$$

Solving, $G = (N_1/N_2)\, R\Gamma^{1/2}$, independent of both ST/N_0, and

$$P(e) = \frac{1}{4}\left(2 - \operatorname{erf}\left\{\frac{1}{2}\frac{\left[N_1 - N_2\Gamma^{1/2} + \Gamma\left(\frac{ST}{N_0}\right)_2\right]^2}{N_1 + N_2\Gamma + 2\Gamma\left(\frac{ST}{N_0}\right)_2}\right\}^{1/2} - \operatorname{erf}\left\{\frac{1}{2}\frac{\left[N_2\Gamma^{1/2} - N_1 + \Gamma^{1/2}\left(\frac{ST}{N_0}\right)_2\right]^2}{N_1 + N_2\Gamma + 2\Gamma\left(\frac{ST}{N_0}\right)_2}\right\}^{1/2}\right) \tag{5}$$

Note that dependence on A_i in Eqs. (4) and (5) has been removed. In addition, if $R = \Gamma = 1$, and $N_1 = N_2 = N$, from Eqs. (3–5) the balanced case

$$P(e) = \frac{1}{2}\left(1 - \operatorname{erf}\left\{\frac{\left(\frac{ST}{N_0}\right)_2^2}{4\left[N + \left(\frac{ST}{N_0}\right)_2\right]}\right\}^{1/2}\right) \tag{6}$$

is obtained, which checks with Ref. 1.

4. AC-Coupling

If the integrator is now ac-coupled, with the given gain settings established above, different performance is achieved. Ac-coupling removes the mean (dc) component, which is $(\mu_{H_1} + \mu_{H_2})/2$, and the means become

$$(\mu_{H_1})_{ac} = \mu_{H_1} - \frac{\mu_{H_1} + \mu_{H_2}}{2} = \frac{\mu_{H_1} - \mu_{H_2}}{2}$$

$$(\mu_{H_2})_{ac} = \mu_{H_2} - \frac{\mu_{H_1} + \mu_{H_2}}{2} = -\frac{\mu_{H_1} - \mu_{H_2}}{2}$$

The variances are unaffected. Making the necessary changes in Eqs. (3–5), the probability of error, for $G = 1$ (imbalance) becomes

$$P(e) = \frac{1}{4}\left\{2 - \operatorname{erf}\left[\frac{1}{2}\frac{\left(\frac{RN_1\Gamma + N_2}{2}\right)^2\left(\frac{ST}{N_0}\right)_2^2}{N_1^3R^2 + N_2^3 + 2\Gamma R^2 N_1^2\left(\frac{ST}{N_0}\right)_2}\right]^{1/2} - \operatorname{erf}\left[\frac{1}{2}\frac{\left(\frac{RN_1\Gamma + N_2}{2}\right)^2\left(\frac{ST}{N_0}\right)_2^2}{N_1^3R^2 + N_2^3 + 2N_2^2\left(\frac{ST}{N_0}\right)_2}\right]^{1/2}\right\} \tag{7}$$

For $G = (N_1/N_2)R\Gamma$ (signal balance),

$$P(e) = \frac{1}{4}\left\{2 - \operatorname{erf}\left[\frac{1}{2}\frac{\Gamma^2\left(\frac{ST}{N_0}\right)_2^2}{N_1 + N_2\Gamma^2 + 2\Gamma\left(\frac{ST}{N_0}\right)_2}\right]^{1/2} - \operatorname{erf}\left[\frac{1}{2}\frac{\Gamma^2\left(\frac{ST}{N_0}\right)_2^2}{N_1 + N_2\Gamma^2 + 2\Gamma^2\left(\frac{ST}{N_0}\right)_2}\right]^{1/2}\right\} \tag{8}$$

For $G = (N_1/N_2)R\Gamma^{1/2}$ (variance balance),

$$P(e) = \frac{1}{2}\left\{1 - \operatorname{erf}\left[\frac{1}{2}\frac{\left(\frac{\Gamma + \Gamma^{1/2}}{2}\right)^2\left(\frac{ST}{N_0}\right)_2^2}{N_1 + N_2\Gamma + 2\Gamma\left(\frac{ST}{N_0}\right)_2}\right]^{1/2}\right\} \tag{9}$$

Some typical results are shown in Figs. 3 and 4. Because $N = BT$, N is expressed in decibels (10 log [ ]) to be compatible with conventional noise bandwidth notation; Γ and R are also expressed in 10 log [ ] form. Given $P_{unbal}(e) = f_{unbal}(ST/N_0)$ and $P_{bal}(e) = f_{bal}(ST/N_0)$ for the unbalanced and balanced ($N = N_2$) receivers, respectively, degradation (in dB) is defined by

$$\text{Degradation}\left(\frac{ST}{N_0}\right) = \frac{ST}{N_0} - f_{bal}^{-1}\left[P_{unbal}(e)\right]$$

Decaying curves with ST/N_0 mean that $P_{unbal}(e) \to P_{bal}(e)$ at high ST/N_0, etc. One may observe from these plots that a balance amplifier can measurably improve performance in some cases, but not in others. Comparing Figs. 3a and 3b, it appears that if the ratio log (N_1/N_2) is of opposite sign to log R_1 a "compensation" is effected, improving the unbalanced receiver's performance.

Ac-coupling appears to offer consistent improvement in performance. This is particularly significant in a case as shown in Fig. 5 where the usual $P(e)$ is plotted (degradation plots lose significance when the two curves are substantially different). The reason for such catastrophic dc performance is that the mean term implicit in $P(e \mid H_2)$ in Eq. (3) has become positive and with a "decide-on-zero" detector $P(e \mid H_2) \to 0.5$. However, it is less positive than the mean term in $P(e \mid H_1)$ so that ac-coupling provides usable performance. Balance amplifiers are not of much use in this case. Particularly alarming is that this case could conceivably be built using crystal filters with ±0.5-dB tolerance in insertion loss R, passband ripple Γ, and noise bandwidth N_1N_2. Clearly, care must be taken when specifying filters for this application.

5. Experimental Work[2]

Using the FSK link described previously (SPS 37-51, Vol. III, pp. 311–313), verification of the theory (signal balance case) by varying the attenuation term R was attempted. The link was as shown in Fig. 2 with the channel attenuations replaced with

$$A_1 = A_1' \cdot C_1$$
$$A_2 = A_2' \cdot C_2$$

where the A_i' were accounted for in a signal balance amplifier and the C_i added to test performance after balancing. For $G = (B_1\alpha_1A_1')/(B_2\alpha_2A_2')$ and modifying Eq. (4),

$$P(e) = \frac{1}{4}\left\{2 - \operatorname{erf}\left[\frac{1}{2}\frac{\left(N_1C_1 - N_2C_2\Gamma + C_1\Gamma\left(\frac{ST}{N_0}\right)_2\right)^2}{N_2C_2^2\Gamma^2 + N_1C_1^2 + 2\Gamma C_1^2\left(\frac{ST}{N_0}\right)_2}\right]^{1/2} - \operatorname{erf}\left[\frac{1}{2}\frac{\left(N_2\Gamma C_2 - N_1C_1 + \Gamma C_2\left(\frac{ST}{N_0}\right)_2\right)^2}{N_1C_1^2 + N_2C_2^2\Gamma^2 \quad 2\Gamma^2C_2^2\left(\frac{ST}{N_0}\right)_2}\right]^{1/2}\right\} \tag{10}$$

is obtained for dc-coupling; modifying Eq. (8),

$$P(e) = \frac{1}{4}\left\{2 - \operatorname{erf}\left[\frac{1}{2}\frac{\left(\frac{C_1 + C_2}{2}\right)^2\Gamma^2\left(\frac{ST}{N_0}\right)_2^2}{N_1C_1^2 + N_2C_2^2\Gamma^2 + 2C_1^2\Gamma\left(\frac{ST}{N_0}\right)_2}\right]^{1/2} - \operatorname{erf}\left[\frac{1}{2}\frac{\left(\frac{C_1 + C_2}{2}\right)^2\Gamma^2\left(\frac{ST}{N_0}\right)_2^2}{N_1C_1^2 + N_2C_2^2\Gamma^2 + 2C_2^2\Gamma^2\left(\frac{ST}{N_0}\right)_2}\right]^{1/2}\right\} \tag{11}$$

is obtained for ac-coupling. Figures 6 and 7 show the theoretical results compared with the experimental results. Here degradation is referenced to the case $C_1 = C_2 = 1$ (instead of the balanced receiver as above). The laboratory results correspond well with the theory, the differences are attributed to measurement tolerances.

[2]Experimental results used in this study were obtained by J. T. Sumida, JPL Spacecraft Telemetry and Command Section.

References

1. Glenn, A. B., "Analysis of Non-Coherent FSK Systems with Large Ratios of Frequency Uncertainties to Information Rates," *RCA Review*, Vol. 27, pp. 272–314, June 1966.
2. Marcum, J. I., "A Statistical Theory of Target Detection by Pulsed Radar," Rand Research Memo RM-754, reprinted in *IRE Transactions on Information Theory*, Vol. IT-6, pp. 161–164, Apr. 1960.

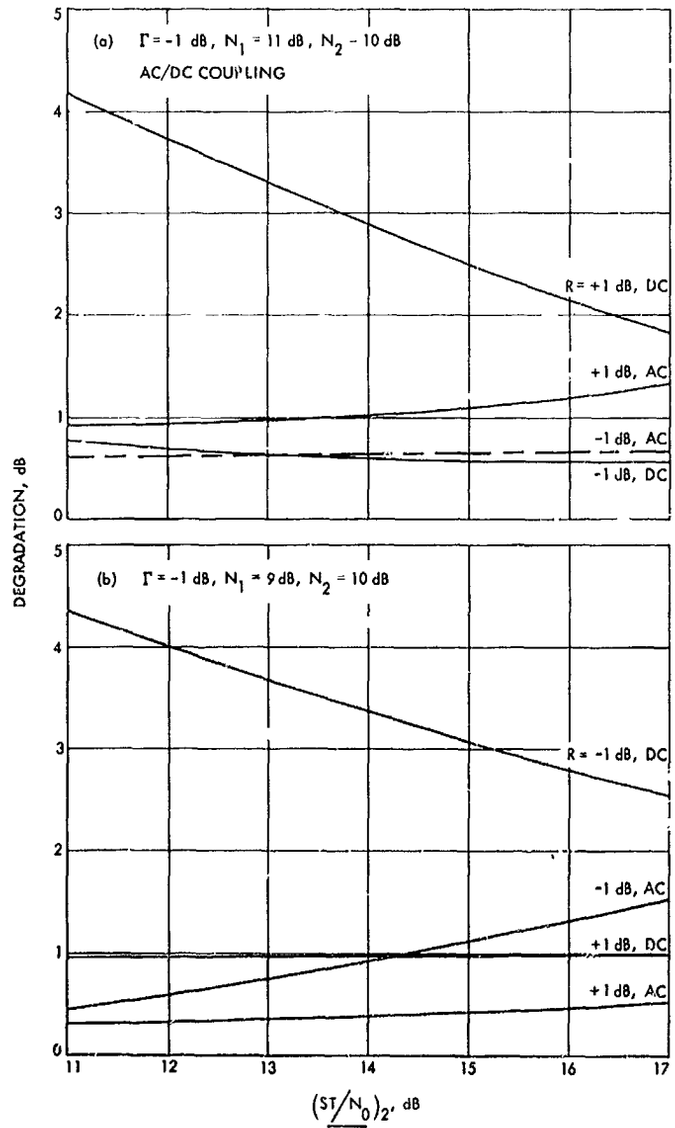


Fig. 3. Degradation of unbalanced receiver

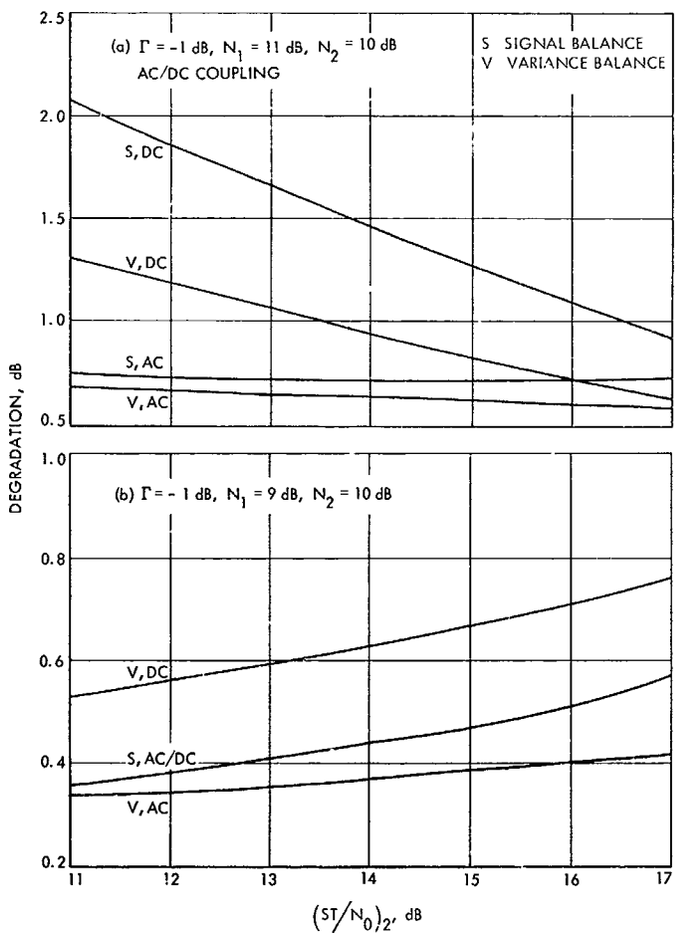


Fig. 4. Degradation of "corrected" receiver

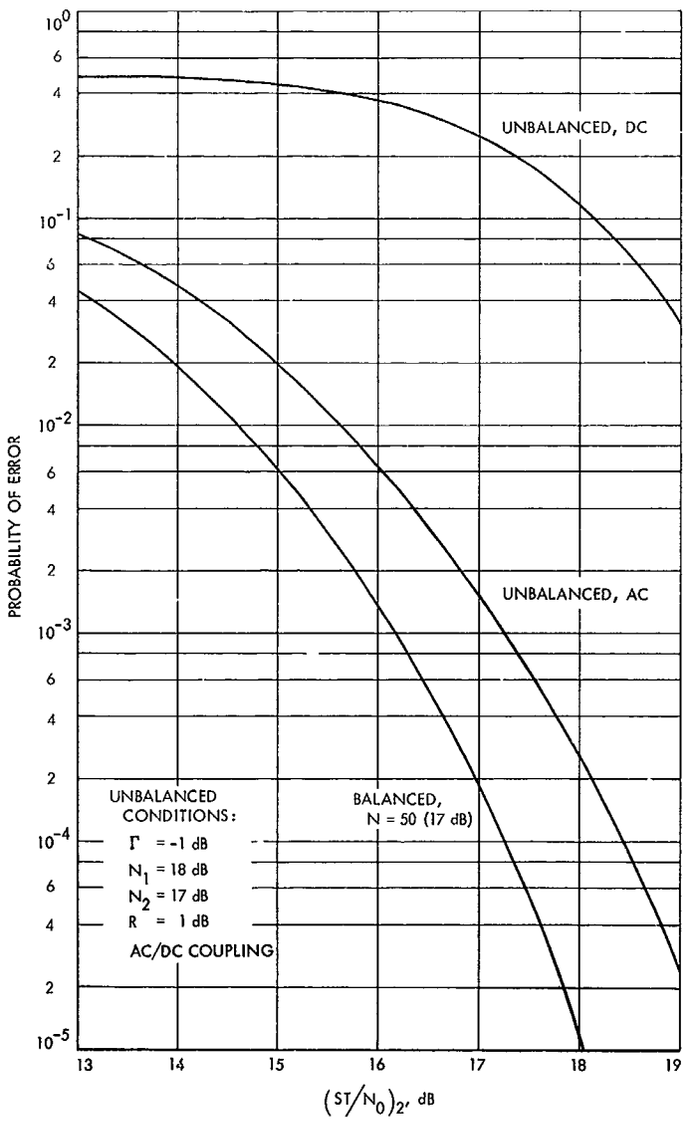


Fig. 5. Comparison of balanced and unbalanced receiver performance

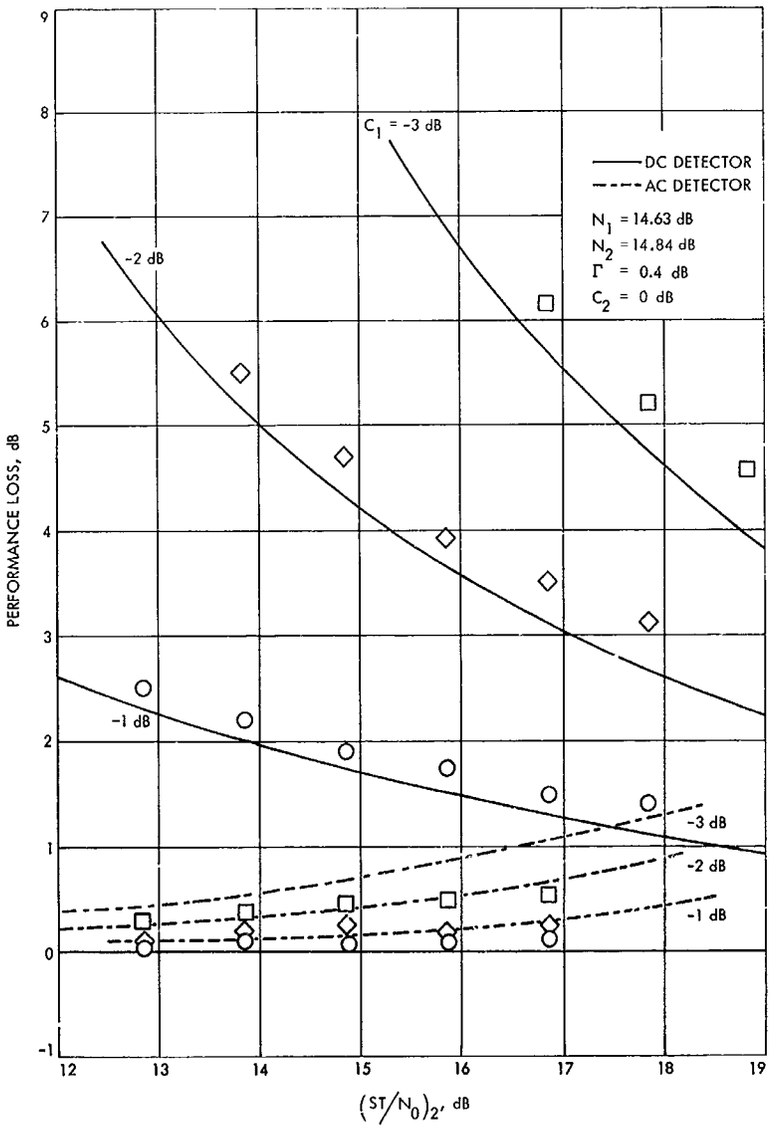


Fig. 6. Comparison of theory and experiment for channel 1 imbalance

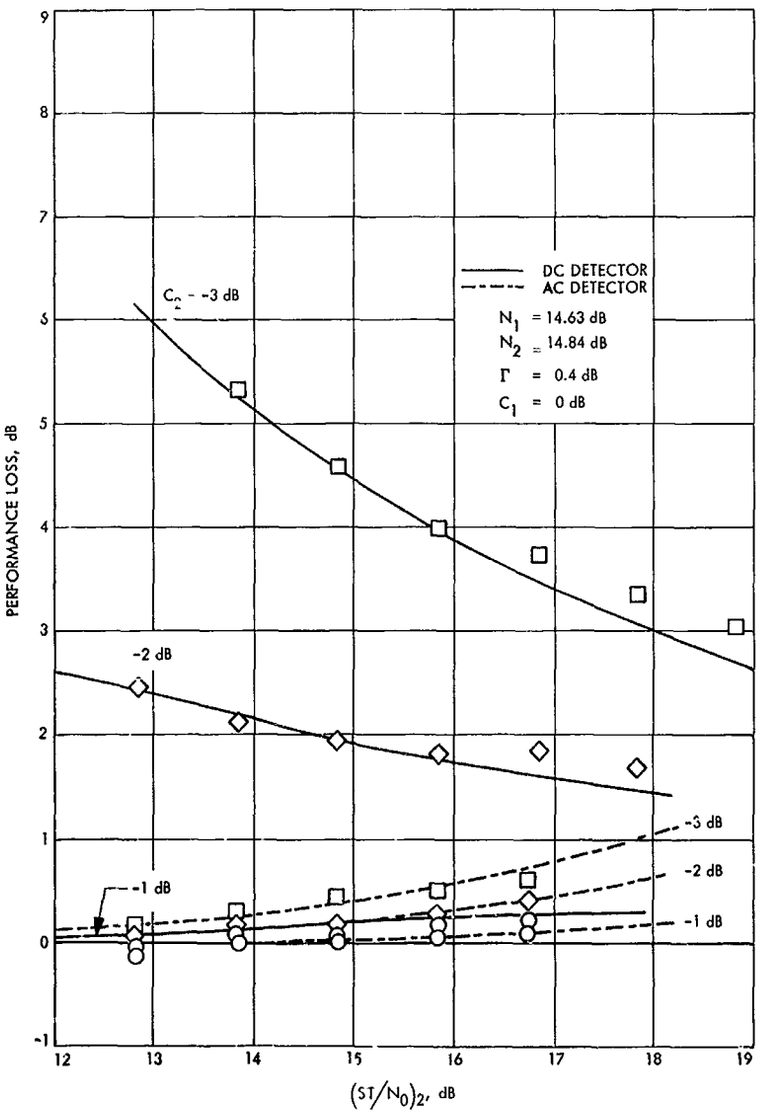


Fig. 7. Comparison of theory and experiment for channel 2 imbalance

XXIII. Spacecraft Radio

TELECOMMUNICATIONS DIVISION

A. Low-Data-Rate Telemetry RF Systems Development, *R. Postal*

1. Introduction

A solid-state multiple frequency shift-keyed (MFSK) 2295-MHz transmitter has been developed as a subassembly for a *m*-ary noncoherent telecommunication system capable of surviving a high impact on a planetary surface. An overall description of the transmitter was given in SPS 37-52, Vol. III, pp. 249–250. Of prime importance is the oscillator frequency stability necessary to achieve the desired communication system efficiency. Although definite stability requirements have not been established, it is desired to have a word separation of only 10 Hz at S-band with a word time of 5 s. In the future, a word time of greater than 20 s may be desired. Until definite requirements can be established, it is an objective of this task to determine the practical limitations of the stability of crystal-controlled oscillators. The discussion in this article pertains to the MFSK oscillator portion of the S-band transmitter.

2. Oscillator Development

In attempting to provide a highly stable MFSK crystal-controlled oscillator, several oscillator design guidelines were found to be necessary. These guidelines are as follows:

(1) Use low noise transistors (2N918 or better).

(2) Maximize isolation of transistor reactances from frequency determining circuitry.

(3) Use low leakage varactor in the control circuit.

(4) Minimize coupling to provide at least 20 dB isolation between the oscillator and the following stage.

(5) Use a low noise, third overtone crystal as a series resonant element in the range of 20–32 MHz.

(6) Set crystal drive level in the range of 50–300 μW.

(7) Minimize variations in crystal drive level.

One oscillator configuration that reflects these guidelines is a modified Colpitts circuit. A description of this circuit was included in SPS 37-40, Vol. III, pp. 198–201.

The additional requirement of high impact resistance resulted in a contract with the Valpey–Fisher Corporation to develop a stable, ruggedized crystal assembly. The necessary manufacturing processes, however, were found to be very stringent and resulted in low yields at a very high cost per unit. A recently available ruggedized

TO-5 crystal assembly was also considered as a possible candidate for the MFSK oscillator. Units manufactured by Monitor Products Company and McCoy Crystal Incorporated are shown in Fig. 1. Both crystals used a third overtone 31.875 MHz AT cut resonator mounted on a four-post TO-5 header. Preliminary tests indicate the TO-5 units would meet the MFSK oscillator requirements. The development contract with Valpey–Fisher has been terminated in view of the readily available, lower cost TO-5 units.

Fig. 1. TO-5 crystal units

3. Test Results

Evaluation of the TO-5 crystal assemblies included high impact tests and measurement of frequency versus temperature, phase noise, and spectral purity. Samples of 12 units each from Monitor Products and McCoy Inc. were used for these tests. Shock tests were performed in three planes at levels ranging from 2000–10,000 g. All units survived 2000 g; however, five McCoy units failed at 5600 g and one Monitor unit failed during the last plane at 10,000 g. Shock caused a $\Delta f/f$ ranging between 1×10^{-7} and 5×10^{-6} per shock. It is believed the ruggedness of the Monitor unit is due to the four-post direct resonator support as opposed to the McCoy ribbon support.

Phase noise and frequency versus temperature data of several TO-5 units are given in Table 1. Phase noise is given as measured phase tracking error at S-band in a phase-locked loop receiver with a noise bandwidth of 20 Hz. Crystal temperature performance is listed as maximum slope of $\Delta f/f$ per °C as measured between 15 and 40°C. Monitor sample 10 and McCoy sample 34 crystals were then selected for spectral purity measurements.

The MFSK crystal-controlled oscillator module was placed in an inoperative oven to minimize frequency drift due to temperature variations. To provide measurements at S-band, the oscillator output frequency was multiplied to 2295 MHz and then translated to 60 KHz for processing in a computer-programmed spectrum analyzer. Spectral analysis is accomplished by sampling the signal with

Table 1. Phase noise and temperature data

Crystal	Phase noise, deg peak	Maximum slope, ppm/°C
Monitor		
10	3.0	0.270
14	3.0	—
15	9.0	0.125
20	9.0	0.188
McCoy		
31	1.5	0.038
34	1.0	0.038
38	2.5	0.132
40	3.5	0.110

an analog-to-digital converter. After N groups of $20T$ samples are collected (where T is equal to sample time and N is greater than 1000), each group is auto-correlated and Fourier transformed to find its spectral peak frequency and value. The program then computes the average value of the amplitude of each peak and computes a loss factor in terms of spectral degradation from an ideal oscillator. Spectral degradation data for the two crystals are:

T, s	Average degradation, dB	
	Monitor 10	McCoy 34
20	5.4	1.38
10	2.3	—
5	0.4	—

Since the McCoy sample 34 showed an improvement of 4 dB over the Monitor sample 10 for the $T = 20$-s case, there was no interest in testing the McCoy unit at smaller sample times.

4. Conclusions

The MFSK oscillator will meet the sterilization and high-impact requirements.

The stability performance indicates that crystal-controlled oscillators have the potential for achieving good MFSK communication system efficiency for word time to at least 20 s. The ultimate performance depends on how well the drift rate of such oscillators can be controlled within the specific mission constraints. Information on this subject will be presented in future articles.

XXIV. *Mariner* Telecommunications

TELECOMMUNICATIONS DIVISION

A. Photon-Actuated Solid-State Switch Development, *D. Bergens*

1. Introduction

The development of a photon-actuated solid-state switch for signal transmission with electrical isolation was introduced in SPS 37-44, Vol. IV, pp. 320–325. The switch's electrical and environmental requirements and general configuration were given. Results were discussed for phase I, which consisted of the design, fabrication, and testing of gallium-arsenide diode–phototransistor combinations, and the electrical design of the gate and driver circuit. The results of phase II, in which the gate–driver was fabricated as an integrated circuit, and complete switches were assembled and tested, are summarized in this article. The photon-switch development was performed by Texas Instruments, Inc. (TI), under a NASA–JPL subcontract.

2. Driver

a. Design. Figure 1 is the schematic diagram of the complete photon switch. The gallium-arsenide diode is

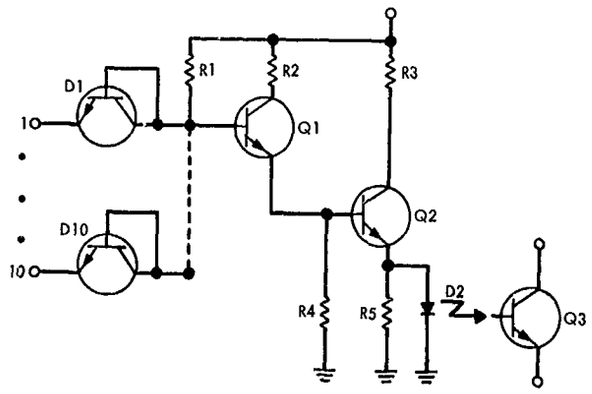


Fig. 1. Photon-switch schematic diagram

D2 and the silicon phototransistor is Q3. The remainder of the circuit is the gate and driver. The circuit is straightforward and needs no explanation, except for R5, which was added to provide a greater off-margin when the switch input is at 1 V, by shunting current around D2. The shunting effect of R5 is not serious at high currents when the driver is on because of the logarithmic characteristic of the diode. The supply voltage requirement was 4 $\pm$½ V.

The gate–driver design was worst-case-designed with parameters of integrated-circuit transistors and resistors. This required the establishment of margins on the semiconductor processing (resistivity ranges, diffusion depths, etc.). A transistor configuration with which TI had previous experience and data was used for all transistors. Its emitter-base junction was used for the input diodes. The base diffusion was used for all the resistors because it has the lowest temperature coefficient. The resistivity range-of-routine base diffusion was wider than needed for the worst-case design. This problem was resolved in the resistor layout by tapping the resistors and providing a choice of interconnect metalizations to pick up the appropriate taps.

The layout of the gate–driver integrated circuit is shown in Fig. 2. This is also a composite drawing of the mask for the integrated circuit. Several features of the layout require an explanation. Four of the standard transistors were used in parallel for Q2, because its emitter current can be as high as 44 mA and the optimum emitter current of the standard transistor is only 10 mA. For future expansion and yield considerations, 14 input diodes were designed in and around the perimeter of the chip. The resistors were tapped to provide the different lengths needed to account for the resistivity range. A choice of two metalizations was sufficient to match the expected resistivity range to the required resistor ratios. The most critical resistor, R3, is actually two resistors of different dimensions in parallel to provide a finer resolution adjustment. Additional taps were also included for experimental reasons.

b. Processing. Routine triple diffusion processing for integrated circuits was used for the gate–driver circuit. Seven photographic masks were required–3 for diffusions and 4 for contact windows and metalizations. A number of evaluations was made during the processing of the prototype runs. Transistor gains were measured. The resistors were probed on the slice before the final interconnect metalization to establish the resistivity range and the choice between the two metalizations. The base resistivity was within the expected range for the prototype runs.

3. Photon-Switch Assembly

The complete photon switch consists of 7 parts: the gate–driver chip, phototransistor, GaAs diode, 2 ceramic insulators, the 14-lead to TO-84 package, and the package lid. Other materials used in assembling the switch are the SeSAs-glass, epoxy, gold-wire, and gold-solder preforms. The major assembly steps are shown in Fig. 3. The assembly procedure is:

(1) Solder the phototransistor and gate–driver chips to the ceramic insulators.

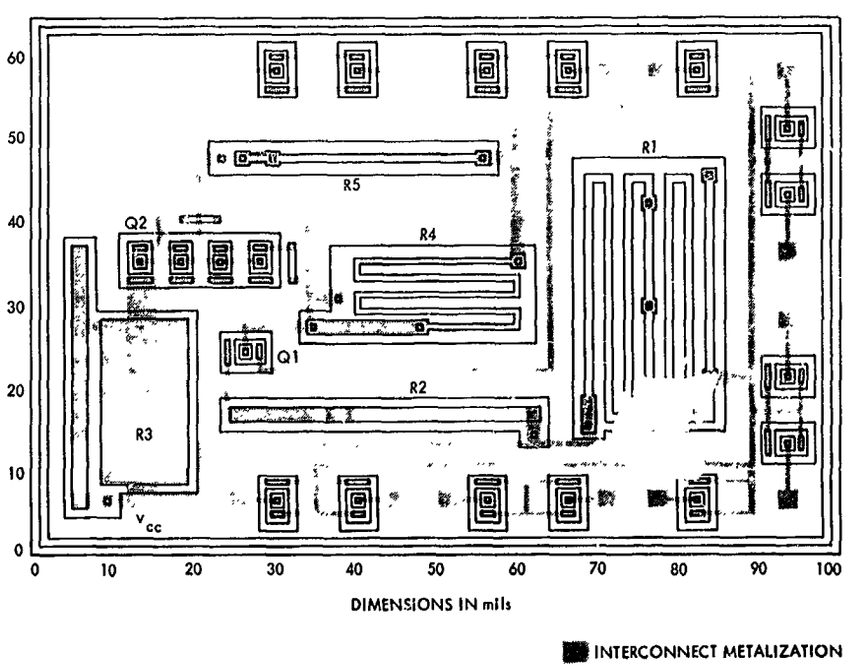
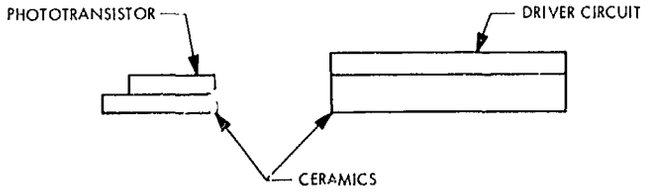
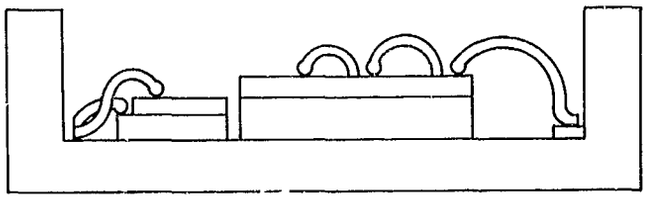


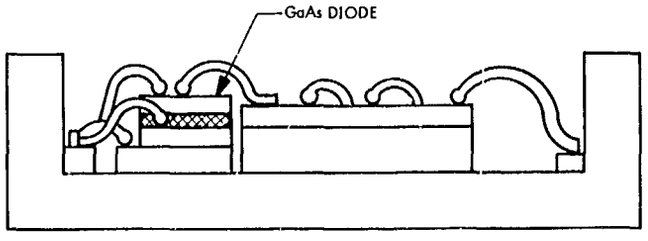
Fig. 2. Gate–driver circuit component layout



1. MOUNT SI WAFERS ON CERAMICS



2. MOUNT CERAMICS IN HEADER AND BOND LEADS



3. MOUNT GaAs DIODE AND BOND LEADS

Fig. 3. Major assembly steps for photon switch

(2) Solder the ceramic subassemblies into the flat package.

(3) Bond leads between pads on the phototransistor and gate–driver chips, and the header pads.

(4) Bond leads to the unmounted GaAs diode.

(5) Mount the diode to the photon-sensitive region of the phototransistor with the SeSAs glass.

(6) Bond the diode leads to the header and gate–driver pads.

(7) Cover the periphery of the diode with epoxy for added support at high temperature.

(8) Clean and hermetically seal the package.

Figure 4 is a sketch of the completed assembly before sealing. The relative size of an unsealed switch is shown in Fig. 5.

4. Tests

a. Environmental. Sample lots of photon switches were subjected to variable frequency vibration, physical shock, centrifuge, and temperature cycling to determine if there were any problems that would preclude their use in spacecraft equipment. There were no failures.

b. Electrical. Thirty photon switches fabricated with prototype drivers were delivered to JPL at the end of phase II. The electrical characteristics of these switches are summarized in Table 1. The parameter I_c is a more sensitive measure of the output than V_{ces}, because I_c is taken at $V_{ce} = 0.6$, which is at the edge of transistor saturation. An explanation of the noise transmissibility parameter was given in SPS 37-44, Vol. IV. Table 1 shows that all the switches met design specifications.

The primary requirement, however, is that the switches meet these specifications after 10,000 h of operation. For this reason, TI established guard bands on I_c and I_{ceo} of 5 mA and 10 μA, respectively, to allow for aging. These two parameters are the ones most likely to change with time. A gradual decrease in I_c is expected due to a very slow degradation of photon output from the diode. This degradation, which is thought to occur only when the photon-emitting diode is on, determines the ultimate life of the switch. Some drift in transistor I_{ceo} with time is not unusual, but it should not change more than 2–1 in good transistors over 10,000 h of operation. The guard band on initial parameters allows for these changes.

c. Life. Twenty-nine of the 30 photon switches were placed on life test in a daisy chain configuration, with the output of one connected to the input of the next, etc. When power is applied, the configuration oscillates because there is an odd number of stages. Each switch is turned on and off at a 1.75-kHz rate. The duty cycle is about 50%. The primary parameter of interest in life testing is I_c because it is the most sensitive indicator of change. In 3,500 h of operation, the degradation of I_c has been minor.

Twelve of the phase I diode–phototransistor pairs were operated with a constant diode current of 30 mA for 14,000 h to determine the aging rate. During this time, the average decrease in I_c was 3 mA, and the maximum decrease was 7½ mA. In 7 of the 12 switches, the decrease

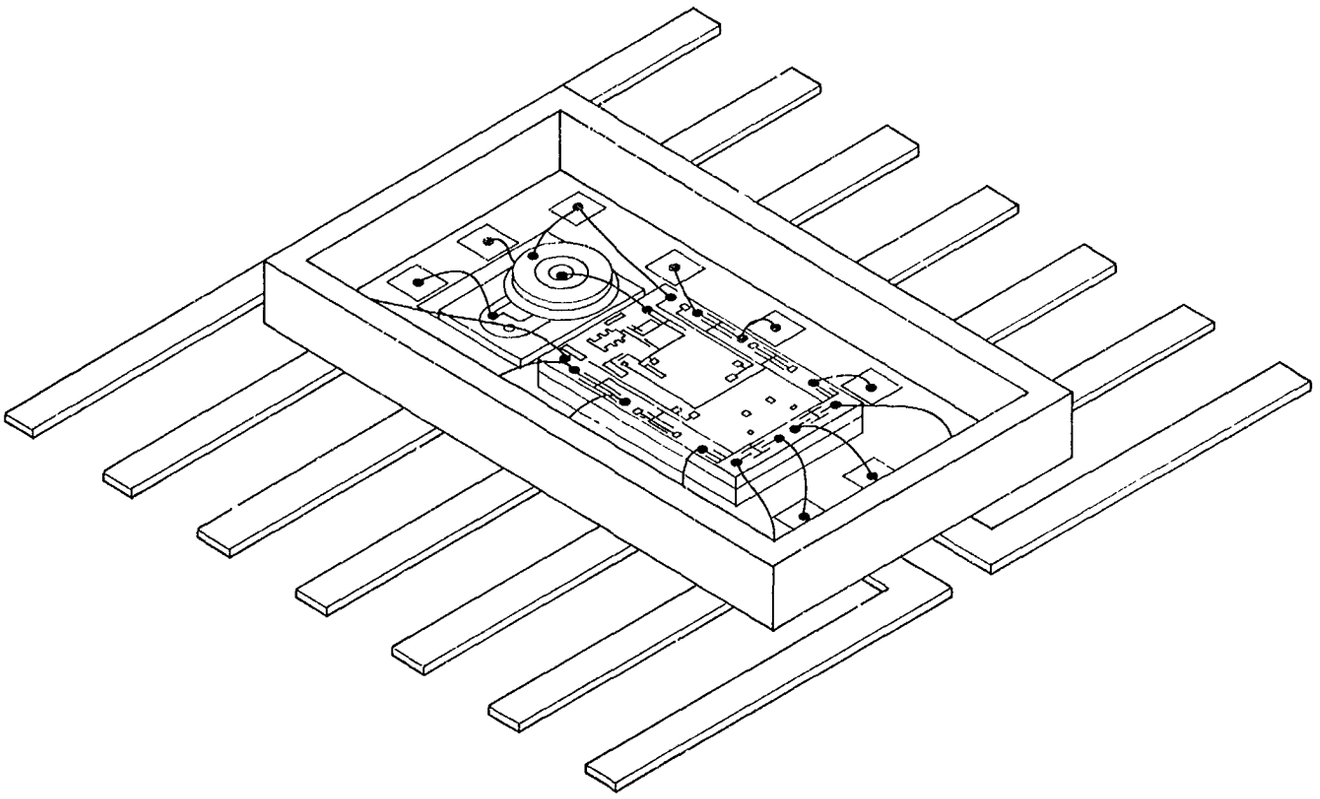


Fig. 4. Completed photo-switch assembly before sealing

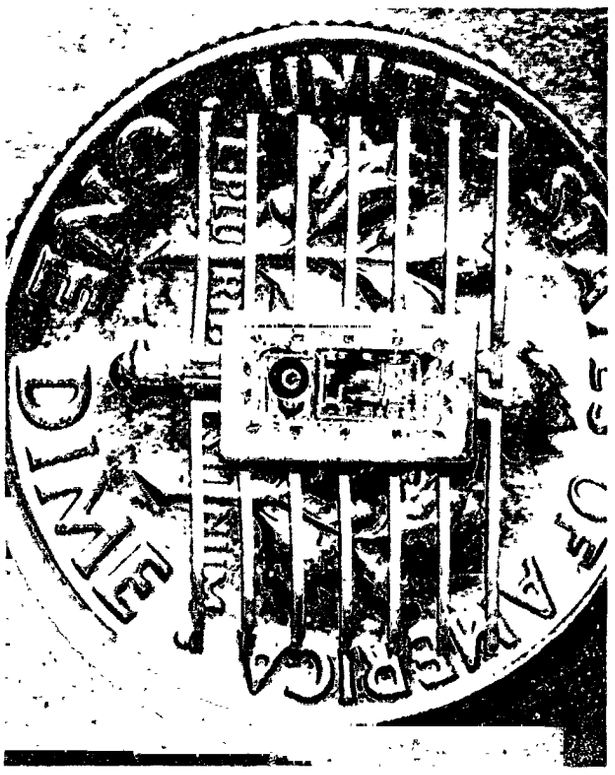


Fig. 5. Relative size of photon switch

Table 1. Electrical characteristics of 30 photon-actuated isolation switches

Parameter	Test conditions[a,b]	Measured value			Specified value	
		Min	Av	Max	Min	Max
Output on-voltage V_{on}, V	I_c = 10 mA, V_{cc} = 3.5 V, V_{in} = 3.0 V					
	Temperature = −20°C	—	—	0.115	—	0.6
	+25°C	-	—	0.150	—	0.6
	+100°C	--	—	0.240	—	0.6
Output on-current I_c, mA	V_{cc} = 0.6 V, V_{cc} = 3.5 V, V_{in} = 3.0 V					
	Temperature = −20°C	28	43	57	10	—
	+25°C	26	33	43	10	—
	+100°C	15	19	26	10	—
Output breakdown voltage BV_{ceo}, V	I_c = 100 μA	60	68	—	35	—
Output leakage current I_{ceo}, μA	V_{in} = 1.0 V, V_{ce} = 20 V, V_{cc} = 4.5 V					
	Temperature = +25°C	—	0.0013	0.0043	—	0.1
	+100°C	—	2.7	9.9	—	20
Isolation capacitance C_{i-o}, pF	Frequency = 1 kHz	—	2.0	2.5	—	10
Input breakdown voltage BV_{DI}, V	I_{in} = 10 μA	7.2	8.0	—	6.5	—
Input off-current I_{in}, mA	V_{in} = 0, V_{cc} = 4.5 V	—	0.19	0.21	—	1
Input on-current, nA	V_{in} = 6.0 V, V_{cc} = 4.5 V	—	<1	2.2	—	3×10^4
Power dissipation-on P_{on}, mW	V_{in} = 6.0 V, V_{cc} = 4.5 V					
	Temperature = −20°C	155	164	180	—	200
	+25°C	148	158	172	—	200
	+100°C	136	142	153	—	200
Power dissipation-off P_{off}, mW	V_{in} = 0, V_{cc} = 4.5 V					
	Temperature = −20°C	—	—	0.97	—	1
	+25°C	—	0.89	0.96	—	1
	+100°C	—	—	0.89	—	1
Switching time-on t_1, μs	I_c = 10 mA, V_{cc} = 4.5 V, $V_{ce(off)}$ = 20 V	2.8	4.0	5.8	—	10
Switching time-off t_2, μs	I_c = 10 mA, V_{cc} = 4.5 V, $V_{ce(off)}$ = 20 V	43	71	98	—	100
Noise transmissibility (emitter to collector) V_n, V	I_c = 2 mA, V_{in} = 0	—	1.8	2.0	—	2

[a] Free air temperature: +25°C unless specified otherwise.

[b] V_{cc} = supply voltage, V_{in} = input voltage, V_{ce} = collector-to-emitter voltage.

was 2 mA or less. The lowest I_c of the 12 switches was 15 mA, after 14,000 h. The rate of change of I_c during the test was approximately linear with time. Eight other diode–transistor pairs that were not operated during the 14,000 h show no change in I_c. The TI engineers believe the aging rate can be directly extrapolated upward in time for switch duty cycles less than 100%. For example, if the duty cycle were 10%, it would take ten times as long for I_c to decrease the same amount.

The saturation voltage of the 12 switches increased an average of 11 mV. The maximum increase was 25 mV. Seven of the 12 increased less than 8 mV. These changes are relatively minor.

One hundred photon switches have been ordered from TI for space flight qualification testing at JPL. When the qualification tests have been successfully completed, the photon switch will be ready for use as interface circuits in the command, central computer and sequencer, and other spacecraft subsystems. They can also be used to interface the ground support equipment with the spacecraft, thus eliminating many of the noise problems encountered during checkout.

TI is marketing the phase I diode–transistor pair in a TO-5 can under the designation TIXL 103. It is called an optically coupled isolator.

A patent application has been filed on the photon switch by JPL. The innovation claimed is "a digital logic gate whose output is electrically isolated from the inputs."