NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report 32-1583

The Theory, Design, and Operation of the Suppressed Carrier Data-Aided Tracking Receiver

M.K Simon
J.C. Springett

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

June 15, 1973

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report 32-1583

The Theory, Design, and Operation of the Suppressed Carrier Data-Aided Tracking Receiver

M.K. Simon
J.C. Springett

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

June 15, 1973

Preface

The work described in this report was performed by the Telecommunications Division of the Jet Propulsion Laboratory.

Acknowledgment

The authors gratefully acknowledge Dr. William C Lindsey, co-inventor of the data-aided loop (DAL), for his many technical contributions which assisted in the preparation of this report. Much appreciation is also due Mr. Christopher Carl, who was primarily responsible for the early implementation of the DAL in software. Finally, the authors wish to acknowledge Mr Frank Kollar, who carried the early software package over into the final hardware design and who also is responsible for a large part of the experimental results given in support of the theory.

Contents

Contents (contd)

Tables

Figures

Contents (contd)

Figures (contd)

Contents (contd)

Figures (contd)

Abstract

This report describes a viable, efficient, and easily mechanized carrier regenerating receiver for use in suppressed carrier-tracking systems. The receiver referred to as a data-aided receiver (DAR) incorporates a data-aided loop (DAL) which provides the required carrier reference signal. The DAL employs the principle of decision feedback and as such is more efficient than other forms of suppressed carrier-tracking loops

The analysis, design, and implementation of the DAR are covered in detail Performance comparisons and mechanization tradeoffs are made, wherever possible, with discrete carrier systems and other suppressed carrier systems presently in use. Experimental performance verification is given throughout in support of the theory presented.

The Theory, Design, and Operation of the Suppressed Carrier Data-Aided Tracking Receiver

I. Introduction

Time and progress have made tremendous inroads on the space-oriented communication system technology that existed in 1958. Data bit rates have increased some four orders of magnitude, and communication distances presently reach to the outer planets. Ground antennas have grown from the relatively small helical types used for the early "Microlock" system to giant 64-m parabolic structures with gains of 72 dB at X-band Receiver noise temperatures have dropped from over 1000 K to less than 25 Data processors have evolved from hand reduction based on strip-chart records to large computer systems.

This progress, of course, took time and research. Ideas had to be generated and developments made until the technology reached fruition. Often, the implementation and acceptance of a new advancement has been paced by the availability of a good, reliable, low-cost circuit technology. The use of block and convolutional codes certainly fits this pattern as their properties and advantages were known well before their actual use beginning in the late '60s, when integrated circuits and minicomputers made decoding practical.

The cost of further improvements in present communications capability, in general, appears to be very high. A great amount is invested in equipment, so much so that to do other than make small modifications within the existing systems is economically unfeasible. With this as a premise, emphasis is placed on ways by which the system performance can be improved without major modification or replacement.

One such area for improvement involves the amount of transmitter power allocated to the carrier. From the inception of the U.S. space program in the late 1950s, the use of carrier coherent receivers has formed the basis for tracking and communicating with the space vehicle At the heart of the receiver is the phase-locked loop (PLL), which even today continues to be constructed and used almost identically to that used for the first Explorer satellite

The concept of using what is called "suppressed carrier" signaling to improve overall system performance, as compared to "discrete carrier" signaling, is very old The advantages of such an approach as it pertains to modern digital communications are presented in Section II of this report

The acceptance of suppressed carrier signals for space vehicles has generally been thwarted by the lack of a complete cohesive theory which explains how well such a

signaling system, including its "carrier regenerating receiver," works relative to the well established discrete carrier systems. The use of carrier regenerating receivers has been further delayed due to the lack of a viable, efficient, and easily mechanized design. This report, which contains complete analysis, designs and implementations, and applications for a receiver known as the data-aided receiver (DAR), seeks to eliminate the above deficiencies

Section III of the report deals with the general theory of operation for the suppressed carrier DAR, in both the acquisition and steady-state tracking modes Functions and forms which are synonymous with the PLL are fully discussed.

Section IV concentrates on the detailed implementation of the DAR, with mechanization tradeoffs, analysis, and experimental performance verification. Section V, which is closely allied with Section IV, investigates the automatic gain control (AGC) and lock detection aspects of integrating the DAR into a larger receiver system.

Section VI explores a number of practical applications and configurations of the DAR. Included are ways of modifying discrete carrier receivers to become data-aided receivers, applications to telemetry and command links, including subcarrier demodulation, and two-way tracking systems.

The final section, VII, presents a qualitative summary of the advantages and disadvantages of DAR systems.

Although no report can be absolutely complete, answer all questions, or cover all possible applications, the authors sincerely believe that this document represents a solid basis upon which one can decide whether or not the use of suppressed carrier signaling is both practical and economical for a given application

II. Advantages of Suppressed Carrier Signaling

A. Basic Rationale—Power Division and Bandwidth

For space applications, two major functions are generally supplied by the radio link: (1) *navigation* via *coherent tracking* and (2) information in the form of *telemetry* or *commands*. In nearly all past programs, the tracking has been with respect to a *discrete carrier* component of the transmitter signal, and telemetry (or command) has been supplied by biphase-modulating *data* onto one or two *subcarriers*, which in turn *phase-modulate* the transmitter. Expressed mathematically, for the single subcarrier case, the transmitter signal is

$$\begin{aligned} s(t) &= \sqrt{2P_T}\cos\{\omega_C t + \theta_m d(t)\,\mathrm{S}(t)\} \\ &= \sqrt{2P_T}\cos\theta_m\cos\omega_C t - \sqrt{2P_T}\,d(t)\,\mathrm{S}(t)\sin\theta_m\sin\omega_C t \\ &= \sqrt{2P_C}\cos\omega_C t - \sqrt{2P_D}\,d(t)\,\mathrm{S}(t)\sin\omega_C t \end{aligned} \tag{2-1}$$

where

P_T = total transmitter power

$P_C = P_T\cos^2\theta_m$ = discrete carrier power

$P_D = P_T\sin^2\theta_m$ = data signal power

ω_C = carrier angular frequency

$d(t)$ = data stream with ± 1 symbols

$\mathrm{S}(t) = \mathrm{S}\sin(\omega_{SC}t)$ = unit power square-wave subcarrier of angular frequency ω_{SC}

Square-wave rather than sine-wave subcarriers are generally used because of their higher efficiency with respect to *sideband power* utilization and ease of implementation on the spacecraft, thus, in the following discussion, square-wave subcarriers are assumed. From Eq. (2-1), it can be noted that the transmitter signal consists of two distinct parts: the discrete carrier (left-hand term), and the data signal (right-hand term), which is also referred to as the sideband term. From the information signaling point of view, the power devoted to the discrete carrier term is wasted. However, conventional discrete carrier-tracking receivers require a certain amount of carrier power in order to satisfactorily accomplish the navigation and carrier demodulation functions In fact, there exists a power division which optimizes the performance of both the tracking and data detection functions (see Ref. 2-1, Sections 5.4.5, 5 6, and A5.1). Let us define the efficiency of the discrete carrier system as

$$\eta_{DC} \triangleq \frac{P_D}{P_T} = \sin^2\theta_m \tag{2-2}$$

As $\theta_m \to \pi/2$, $P_C \to 0$ and $\eta_{DC} \to 1$. When $\theta_m = \pi/2$, $P_C = 0$ and there is no discrete carrier component, hence the term *suppressed carrier*. Now all of the available transmitter power is in the information-bearing signal, the best of conditions from the data signal point of view. If we define the efficiency of the suppressed carrier system η_{SC} also as the ratio of data signal power to total signal power, then clearly $\eta_{SC} = 1$. Finally, we may define the *data signal improvement* of the suppressed carrier format relative to that of the discrete carrier, with all other parameters held constant, as

$$G_D \triangleq \frac{\eta_{SC}}{\eta_{DC}} = \csc^2\theta_m \tag{2-3}$$

Noting that $1 < G_D < \infty$, some data signal improvement is always achieved by using the suppressed carrier technique.

With the suppressed carrier signal, the need to "carrier track" still exists. Since no discrete carrier signal is available, however, a receiver which tracks the information-bearing signal is required In general terms this is called a *carrier regenerating loop* Several ways in which such a loop may be accomplished are introduced in Section C At this point, let us confine the discussion to carrier-tracking efficiencies A figure of merit for the carrier-tracking loop is the *signal-to-noise ratio* (SNR) ρ in the carrier-tracking *loop noise bandwidth* W_L For the discrete carrier-tracking loop, this signal-to-noise ratio is given by

$$\rho_{DC} = \frac{2P_C}{N_0 W_L} = \frac{2P_T\cos^2\theta_m}{N_0 W_L} \tag{2-4}$$

where N_0 is the one-sided *noise spectral density* in watts per hertz (W/Hz). For the carrier regeneration tracking loop, the loop SNR is given by

$$\rho_{SC} = \frac{2P_T L_R}{N_0 W_L} \tag{2-5}$$

where $L_R < 1$ is an *efficiency loss* associated with the carrier regeneration process Loss L_R can range from

0.99+ for *data-aided loops* to less than 0.2 for *squaring-type* loops (see Section C). We may therefore define the *tracking loop improvement* of the suppressed carrier relative to the discrete carrier case as

$$G_{TR} \triangleq \frac{\rho_{SC}}{\rho_{DC}} = L_R \sec^2 \theta_m \tag{2-6}$$

To give a feeling for the magnitude of the data signal and tracking loop improvements, Table 2-1 shows some typical values. ($L_R = 0.91$ is assumed.)

Table 2-1. Typical values for data signal and tracking loop gains ($L_R = 0.91$)

θ_m, deg	G_D, dB	G_{TR}, dB
30	6 0	0 8
40	3 8	1 8
50	2 2	3 4
60	1.2	6 0
70	0 6	9 4

The foregoing discussion is, of course, somewhat simplified, as other considerations and tradeoffs are necessary; however, it does give a gross feeling for what can be accomplished with suppressed carrier signals

The suppressed carrier concept can easily be extended to the case of two data-modulating signals. The transmitter signal for the *interplex** form of modulation (Ref. 2-1, Section 5.4.3) is given as

$$s(t) = \sqrt{2P_T} \sin [\omega_C t + \theta_{m1} d_1(t) \mathsf{S}_1(t) + \theta_{m2} d_1(t)\, d_2(t) \mathsf{S}_1(t) \mathsf{S}_2(t)] \tag{2-7}$$

The power in each usable component is given in Ref. 2-1 (Table 5-6). When $\theta_{m1} = \pi/2$, the suppressed carrier case is obtained, so that

$$s(t) = \sqrt{2P_T} \cos \theta_{m2} d_1(t) \mathsf{S}_1(t) \cos \omega_C t + \sqrt{2P_T} \sin \theta_{m2} d_2(t) \mathsf{S}_2(t) \sin \omega_C t \tag{2-8}$$

It is seen that there are only two data-dependent components. A special, well known case, referred to as *quadriphase*, is obtained when $\theta_{m2} = \pi/4$, $\mathsf{S}_1(t) = \mathsf{S}_2(t) \equiv 1$, and $d_1(t)$ and $d_2(t)$ are derived from a single data stream.

*This form is necessary to obtain quadrature suppressed carriers

Gains identical to those of the single-channel case are easily obtained as

$$G_{D1} \triangleq \frac{P_T}{P_{D1}} = \csc^2 \theta_{m1} \tag{2-9}$$

$$G_{D2} \triangleq \frac{P_T}{P_{D2}} = \csc^2 \theta_{m1} \tag{2-10}$$

$$G_{TR} \triangleq \frac{P_{SC1}}{P_{DC}} = L_R \sec^2 \theta_{m1} \tag{2-11}$$

Equation (2-11) assumes that the carrier regeneration loop tracks the signal component dependent on $d_1(t)$.

We now turn our attention to spectral occupancy. In the case of telemetry, the discrete carrier signal must be compatible with the tracking stations. For deep space applications interfacing with the Deep Space Network (DSN), there are restrictions on the choice of subcarrier frequency and the *data symbol rate*

$$\mathscr{R}_s \triangleq \frac{1}{T_s}$$

An unwritten but well observed restriction is that the minimum number of square-wave subcarrier cycles per symbol be 1.5. Actually, 3 or more subcarrier cycles per symbol is preferred. Another requirement is that the receiving system be able to recover up to and including the ninth harmonic of the subcarrier in order to maintain reasonable detection efficiency. A typical single-channel *power spectrum*, $S(\omega)$, for the case of 1.5 subcarrier cycles per symbol is shown in Fig. 2-1. The *radio-frequency* (RF) *bandwidth* required to recover the sideband power inclusive of the ninth harmonic of the subcarrier is approximately $20\,\omega_{SC} = 30\,(2\pi\mathscr{R}_s)$.

For the suppressed carrier, single-channel signal, a subcarrier is not required and, in fact, is a detriment to the carrier regenerating loop operation. Thus, the carrier may be directly modulated by the symbols; a typical spectrum

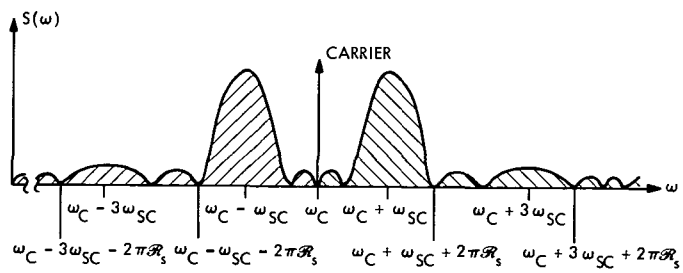

Fig. 2-1. Typical discrete carrier single-channel spectrum

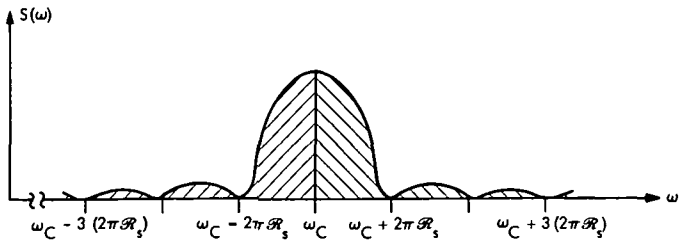

Fig. 2-2. Typical suppressed carrier single-channel spectrum

is shown in Fig. 2-2. For this spectrum, the amount of RF bandwidth required to recover the same amount of sideband power as that from the discrete carrier spectrum of Fig. 2-1 is approximately $5(2\pi\mathcal{R}_s)$. *As a result, an RF bandwidth saving of a factor of 6 is obtained when the suppressed carrier signal is used.*

Similar bandwidth reduction arguments can be made for the two-channel case In general, the highest data-rate stream directly modulates the carrier. However, since there are several useful variations involved in the choice of modulation formats, the details will be omitted here. Section VII provides some insight into the possible variations.

The general conclusion reached from the foregoing discussion is that suppressed carrier modulation is highly efficient from the standpoint of both transmitter power utilization and RF bandwidth required.

B. Two-Way SNR Advantages Relative to Discrete Carrier-Tracking Receivers

At this point, we will briefly consider the merits of using *suppressed carrier receivers* (SCRs) in the two-way tracking system. The tracking and data channel gains discussed in the preceding section apply to both up- and downlinks Our concern here is with the *turnaround function*. In the standard discrete carrier *transponder*, the uplink carrier is tracked by a phase-locked loop (PLL) (Ref. 2-2, Sections 2.3, 4.3). The received carrier frequency is then effectively multiplied by the turnaround ratio G and becomes the exciter for the downlink transmitter. For *S-band* up and down frequencies, the ratio $G = 240/221$, while for S-band up and *X-band* down (S/X) frequencies, $G = 880/221$. The output of the uplink carrier-tracking loop has two phase-modulation components: (1) the in-band random phase noise, which is a function of the tracking loop signal-to-noise ratio ρ, and (2) components of the uplink modulation (e.g., command), which fall within the tracking bandwidth. The intensity of this latter modulation is dependent on the uplink modulation index and the modulated carrier spectrum relative to the tracking loop bandwidth as well as the large frequency characteristics of the low-pass loop filter.

When the phase-modulated output of the uplink tracking loop is frequency-multiplied by the turnaround ratio, the phase deviation of the two modulation components is also increased by G. The resultant suppression of the downlink carrier due to the unwanted modulations, particularly for the S/X configuration, can become significant. The use of an uplink SCR minimizes to the greatest extent possible the turnaround modulation effects. First, since ρ is larger compared to the discrete carrier receiver, the random phase noise will be as small as achievable. Secondly, one of the virtues of a carrier regenerating loop is that it is basically an *anti-modulation device* in the sense that very little of the direct carrier modulation is tracked by the loop. This is particularly true of the data-aided loop (DAL) introduced in the next section. Thus, for all practical purposes, the phase-modulation component due to the uplink sideband frequencies in the vicinity of the carrier is eliminated.

A more detailed discussion of two-way SCRs, along with the modulation-induced downlink phase modulation components, is presented in Section VII-D.

C. Known Types of Suppressed Carrier Receivers

The purpose of a suppressed carrier regenerating loop is to acquire and track a discrete frequency component which is derived in some nonlinear manner from the received suppressed carrier signal. The nonlinear operation must be of even order and, for the case of single-channel signals, is second-order. Three well known regeneration loops will be discussed here: (1) the *squaring loop*, (2) the *Costas loop*, and (3) the data-aided loop.

For the purpose of the following discussion, the received signal that is input to the receiver will have the form

$$x(t) = \sqrt{2P_D}d(t)\cos\omega_C t + n_i(t) \qquad (2\text{-}12)$$

where the input noise $n_i(t)$ is bandpass with one-sided noise spectral density N_0 W/Hz. (See Eq. 3-2 for a more detailed definition.) The suppressed carrier signal component of $x(t)$ has the power density spectrum shown in Fig. 2-2

The basic squaring loop is shown in Fig. 2-3. The signal $x(t)$ is bandpass-filtered and squared (multiplied by itself)

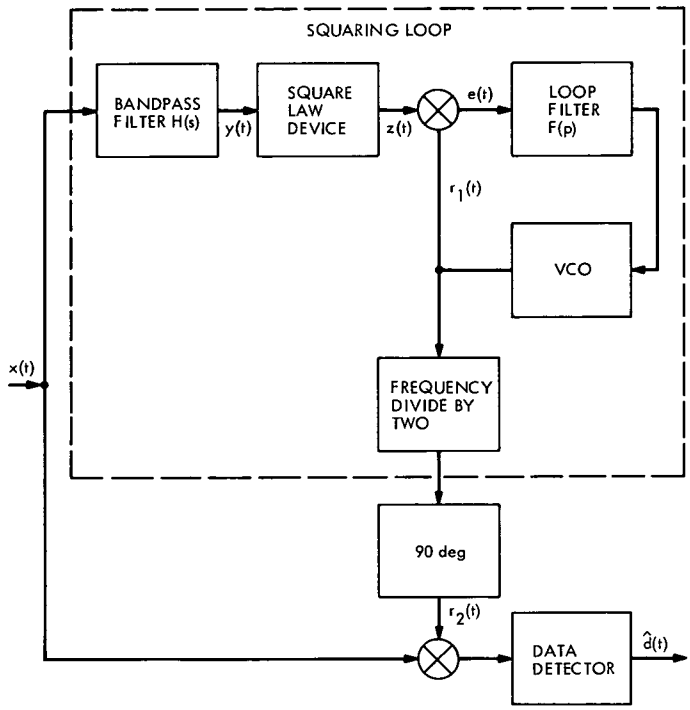

Fig. 2-3. Squaring loop receiver

to produce $z(t)$, which has a discrete frequency component at $2\omega_C$. The loop locks to this component, and the output of the *voltage-controlled oscillator* (VCO) is divided in frequency by two and phase-shifted 90 deg to provide the reference $r_2(t)$ which *coherently demodulates* the carrier. The output of the *data detector* is the maximum-likelihood estimate $\hat{d}(t)$ of the transmitted symbols.

The performance of the squaring loop, as measured by the value of L_R, is largely a function of the SNR into the square-law device (see Ref 2-2) This loss is given by

$$L_R = \left[1 + \frac{K_L}{2\rho_s}\right]^{-1} \tag{2-13}$$

where K_L is related to the autocorrelation function of the input bandpass filter (BPF), $H(s)$, and ρ_s is the SNR measured in the noise bandwidth of the input BPF. As an example of a typical value for L_R, suppose that the input BPF is a two-pole Butterworth and that the 3-dB bandwidth of the filter is set at $4\mathcal{R}_s$. Assuming a nominal symbol error probability of 1×10^{-2} ($P_D T_s/N_0 = 2.7$), $\rho_s = 0.608$. For the two-pole Butterworth, $K_L = 0.75$, thus, $L_R = 0.62$ or -2.1 dB. Even higher losses than this will occur if the data stream is coded, since $P_D T_s/N_0 << 2.7$

We see that the squaring loss is a function of input SNR, which in turn depends upon the bandwidth of the input filter. Obviously, in order to minimize the loss, the filter bandwidth should be made as narrow as possible without attenuating a significant portion of the signal power. This presents certain mechanization problems. First, the center frequency of the BPF may be relatively high (generally the last IF frequency in the receiver), while the symbol rate $\mathcal{R}_s$ may be quite low, requiring that the filter be very high Q. This not only introduces linearity and stability problems but can also result in high costs. Furthermore, each time the data rate is changed, it is necessary to change the filter, further increasing costs, and introducing a degree of inflexibility. Thus, the use of the squaring loop for suppressed carrier, multiple-data-rate applications has severe limitations It has been used for the tracking and demodulation of low-data-rate, biphase-modulated subcarriers of the type used in the Mariner 1969 (and following) two-channel command systems (Ref. 2-3).

The Costas loop, shown in Fig 2-4, does not require that the input BPF have a specified bandwidth, rather, its performance depends upon the specification of the two low-pass filters, $G(s)$. Actually, the Costas loop is nothing more than a low-pass version of the squaring loop (see Ref 2-4 for bandpass–low-pass equivalents), and has the identical stochastic differential equation as that of the squaring loop, provided that the noise spectrum formed by the filters for each configuration is the same (Ref. 2-2) Since it is generally easier to build switchable, active, low-pass filters, the Costas loop has a mechanization advantage over the squaring loop for multiple-data-rate applications

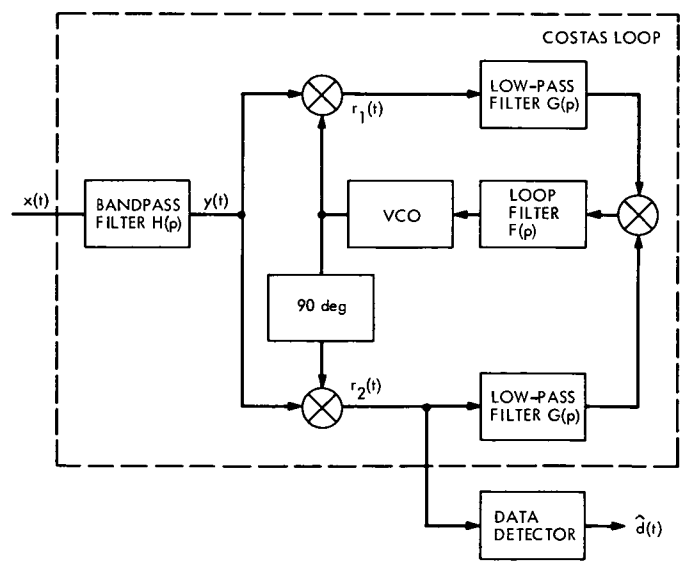

Fig. 2-4. Costas loop receiver

It should be noted that for both the squaring and Costas loops, no attempt is made to optimize the receiver performance with respect to the modulating signal $d(t)$. That is, the nature of $d(t)$, other than its power spectrum, plays no real part in the formation of the loop error signal. An optimum receiver would, however, make proper use of the data estimate $\hat{d}(t)$. A receiver configuration which accomplishes this is the data-aided loop, the basic configuration is shown in Fig 2-5 The topology of the DAL is very similar to that of the Costas loop, there are, however, major differences. First, the data detector is now an integral part of the loop. Secondly, a time delay equal to the symbol duration $T_s = 1/\mathcal{R}_s$ is employed. Thirdly, the low-pass filters $G(s)$ are not required, and the input BPF serves only to limit the total input noise voltage.

The performance analysis of the DAL is the main subject of this report, beginning with Section III. Drawing from one of the results presented in Section III-C, we may compare the DAL performance to that of the squaring (or Costas) loop. For the case where the symbol rate $\mathcal{R}_s$ is large compared to the loop bandwidth W_L, the efficiency loss for the DAL is given by

$$L_R = (1 - 2P_e^s) \tag{2-14}$$

where P_e^s is the average symbol probability of error. Using the previously cited conditions for the squaring loop, we readily find that $L_R = 0.98$ or -0.1 dB. Thus, for the stated example, with respect to loop signal-to-noise ratio, the DAL performance exceeds that of the squaring loop by 2.0 dB. A more parametric comparison of these loops is given in Ref. 2-5.

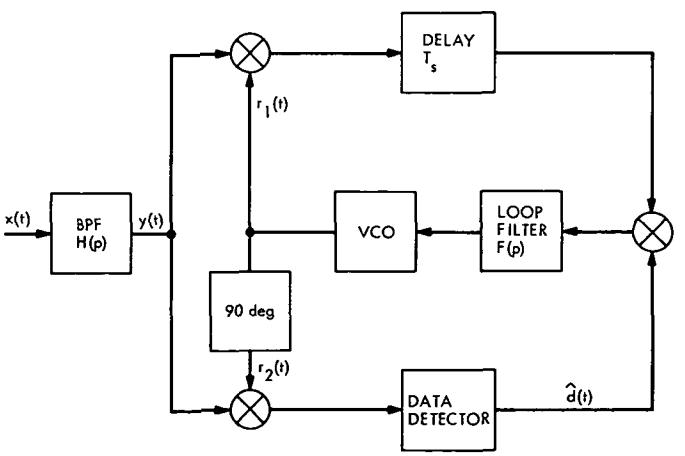

Fig. 2-5. Suppressed carrier data-aided receiver

Finally, it should be pointed out that the DAL offers certain mechanization advantages over, say, the Costas loop. It is very amenable to simple all-digital implementation, which makes the problems of *data-rate switching* virtually disappear. Details on mechanization are given in Section IV.

The data-aided loop as described in this report has yet to find application in any operational system. There are, however, two closely related mechanizations which are presently being employed.

The first application is found in the all-digital *command detector* (Ref. 2-6), where the subcarrier tracking loop is indeed a form of the DAL The second is the subcarrier tracking loop found in the DSN's subcarrier demodulator assembly (SDA) (Refs 2-7 and 2-8) In the SDA, the loop is actually closer to a Costas loop than a DAL, with the exception being that a hard limiter follows the low-pass filter (LPF) which contains the demodulated data symbols, thus making a continuous, nonintegrated estimate of the data symbol polarity

D. Differences Between DAL Suppressed Carrier and Discrete Carrier Systems

For discrete carrier systems, the RF tracking loop need only acquire and track the carrier line in the received spectrum. The tracking loop has no knowledge of the data modulation The carrier loop does, of course, provide references for demodulating the data-modulated subcarrier from the carrier.

The DAL, on the other hand, must have credible estimates of the data symbols before a useful tracking error signal can be formed. However, the recovery of data symbols implies coherent carrier demodulation. Thus, the acquisition of the DAL would seem to involve one of those incredible schemes to determine "which comes first, the chicken or the egg." Actually, this is not the case, and it is quite easy to obtain good data symbol estimates even though the loop is very much out of lock due to the fact that the symbol transitions are not affected by carrier lock, thus permitting the data detector to synchronize and function The effect of the carrier being out of lock is to amplitude-modulate the data symbol stream by the loop error signal

All suppressed carrier regeneration tracking loops, including the DAL, have *double-frequency loop error signals* with respect to those of the discrete carrier loops

This gives rise to differences in the stochastic performance, and introduces a *data phase ambiguity.*

The DAL also has a *finite delay function* which can contribute to instability under certain conditions. In terms of attendant functions, the DAL has an *envelope type of automatic gain control* (AGC), and an *absolute value type of lock detector* (see Section V).

The differences between the DAL and discrete tracking loops are summarized in Table 2-2, along with the text sections in which detailed discussions are found.

Table 2-2. Differences between discrete carrier and data-aided tracking receivers

Item	Discrete carrier loop	DAL	Sections
Loop operation dependent on data modulation	No	Yes	II-D, III-A, III-B, III-C, IV-B
Subcarrier required for single-channel modulation	Yes	No	II-A
Form of linear loop error signal	ϕ	2ϕ	III-C
Data estimate has phase ambiguity	No	Yes	III-C, VI-A
Delay function instability	No	Yes	III-D
AGC system	Coherent amplitude	Envelope	VI-A
Magnitude of modulation feedthrough	Spectrum, loop bandwidth, and modulation index dependent	Essentially zero	II-B, VII-G

References

2-1. *Telecommunications Systems Design Techniques Handbook,* edited by R. E. Edelson, Technical Memorandum 33-571, Jet Propulsion Laboratory, Pasadena, Calif., July 15, 1972.

2-2 Lindsey, W. C., and Simon, M. K., "The Performance of Suppressed Carrier Tracking Loops in the Presence of Frequency Detuning," *Proc IEEE,* Vol. 58, No 9, Sept 1970, pp. 1315–1321

2-3. *Mariner Mars 1969 Handbook,* Document 605-211, Jet Propulsion Laboratory, Pasadena, Calif., July 28, 1969 (JPL internal document).

2-4. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering,* John Wiley & Sons, Inc., New York, 1965.

2-5. Lindsey, W. C, and Simon, M K., "Data-Aided Carrier Tracking Loops," *IEEE Trans. Commun. Technol*, Vol. COM-19, No 2, April 1971, pp 157–168

2-6. Tegnelia, C., and Meahl, M., *Final Report—Digital Command Detector Development,* Document 900-547, Jet Propulsion Laboratory, Pasadena, Calif., April 28, 1972 (JPL internal document).

2-7. Brockman, M. H., "Subcarrier Demodulator Assembly," in *The Deep Space Network,* Space Programs Summary 37-49, Vol. II, p. 100, Jet Propulsion Laboratory, Pasadena, Calif., Jan. 31, 1968.

2-8 Brockman, M. H., "MMTS: Performance of Subcarrier Demodulator," in *The Deep Space Network,* Space Programs Summary 37-52, Vol. II, pp. 127–141, Jet Propulsion Laboratory, Pasadena, Calif., July 31, 1968.

III. A General Theory for the Data-Aided Receiver

A. The Basic Concept and Its Functional Configuration

The basic idea of a *data-aided receiver* (DAR) centers around using the power in the composite received signal (carrier and data-modulated subcarriers) sidebands to enhance the signal-to-noise ratio in the bandwidth of the carrier-tracking loop The vehicle through which this is accomplished employs the principle of decision-directed feedback, and the carrier-tracking loop to which it is applied is referred to as a *data-aided loop* (DAL). The underlying idea behind the data-aided approach is that an enhancement of loop signal-to-noise ratio (SNR) not only improves the carrier-tracking function (by reducing the loop's mean-squared phase noise and average rate of cycle slipping) but also results in an improvement in the data detector performance because of the accompanying reduction in noisy reference loss.

For a single-channel system wherein a discrete carrier is phase modulated by a biphase-modulated data subcarrier, the sideband power being centered around the subcarrier frequency would ordinarily be filtered out by the carrier tracking loop [e g., a phase-locked loop (PLL)], and hence does not affect carrier-tracking performance. In the DAR, an estimate of the biphase-modulated data subcarrier is formed which, when fed back to the DAL, can be used to recover the power in the sideband components for carrier-tracking purposes

More specifically, consider the block diagram illustrated in Fig. 3-1. The upper half of the configuration constitutes a PLL used in present-day tracking receivers. The modification proposed here makes use of a quadrature channel to insert a signal component into the voltage-controlled oscillator (VCO), whose power is proportional to that of the biphase-modulated data subcarrier. One principal point to note is that the shape of the equivalent phase-detector characteristic (S-curve) is affected by the power in the signal's sidebands, the phase jitter in the subcarrier tracking loop, the symbol sync jitter, and the conditional probability of error of the data detector. However, cursory examination of these effects indicates that the symbol sync jitter and subcarrier sync jitter are small relative to the radio loss caused by the carrier-tracking loop phase error; consequently, essentially all of the sideband power can be recovered and used for improving the principal source of system degradation, *viz.*, the noisy reference loss.

In Section II, we established a case for suppressed carrier signaling, that is, where the data are directly biphase-modulated on the carrier, thus, we shall focus our entire attention on the specific version of the DAL referred to as the suppressed carrier DAL (denoted henceforth simply by DAL for the sake of brevity) (see Fig. 3-2).

B. Development of the Stochastic Integro-Differential Equation of Operation

In this section, we develop the loop equation of operation to the extent that it applies to Fig. 3-2. For a single-channel suppressed carrier system, the received signal,

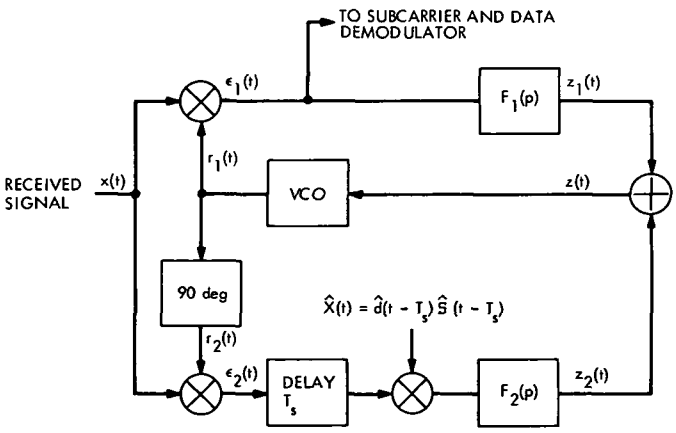

Fig. 3-1. Data-aided carrier-tracking loop

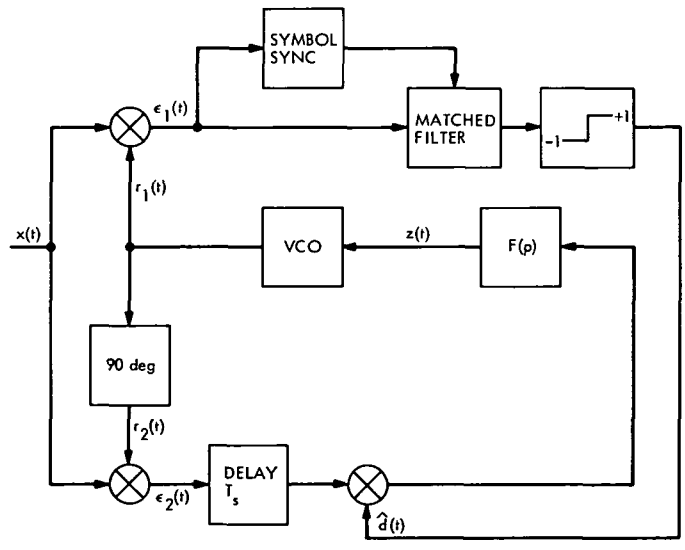

Fig. 3-2. Suppressed carrier data-aided carrier-tracking loop

$x(t)$, is of the form

$$x(t) = \sqrt{2P_D}\sin\left[\omega_C t + \frac{\pi}{2}d(t) + \theta\right] + n_s(t)$$

$$= \sqrt{2P_D}d(t)\cos\Phi(t) + n_s(t) \tag{3-1}$$

where $\theta = \Omega_0 t + \theta_0$ is a slowly varying function of time which is related to the doppler offset Ω_0, the function $\Phi(t)$ is defined by $\Phi(t) = \omega_C t + \theta$, and the random process $\{n_s(t)\}$ is assumed to be narrowband and gaussian. A sample function representation of this process is given by

$$n_s(t) = \sqrt{2}\,[n_1(t)\cos\Phi(t) - n_2(t)\sin\Phi(t)] \tag{3-2}$$

where $n_1(t)$ and $n_2(t)$ are independent gaussian random processes with single-sided spectral densities N_0, W/Hz It is convenient to characterize the reference signals $r_1(t)$ and $r_2(t)$ in Fig. 3-2 by

$$\begin{aligned} r_1(t) &= \sqrt{2}\,K_1\cos\hat{\Phi}(t) \\ r_2(t) &= -\sqrt{2}\,K_2\sin\hat{\Phi}(t) \end{aligned} \tag{3-3}$$

where $\hat{\Phi}(t) = \omega_C t + \hat{\theta}$, and $\hat{\theta}$ is the loop estimate of θ Assuming a lower arm multiplier gain of K_{2m} and neglecting double-frequency components, the output $z(t)$ of the filter $F(p)$ is given by

$$z(t) = K_2K_{2m}F(p)\exp[-pT_s]$$
$$\times\{\sqrt{P_D}\,d(t)\hat{d}(t)\sin\phi(t) + \hat{d}(t)N_2[t,\phi(t)]\} \tag{3-4}$$

where

$$N_2[t,\phi(t)] \triangleq n_1(t)\sin\phi(t) + n_2(t)\cos\phi(t) \tag{3-5}$$

is a baseband "white gaussian noise process" of single-sided spectral density N_0 W/Hz, and

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

is the loop phase error.

The upper arm of Fig. 3-2 forms the data estimate $\hat{d}$ Assuming a multiplier gain of K_{1m} for the upper arm cross-correlator, then

$$\hat{d}(t - T_s) = \mathrm{sgn}\left[K_{1m}\int_0^{T_s} x(t)\,r(t)\,dt\right] \tag{3-6}$$

where the signum function sgn x is defined by

$$\mathrm{sgn}\,x = \begin{cases} +1 & x > 0 \\ -1 & x < 0 \end{cases} \tag{3-7}$$

Expanding $x(t)$ of Eq. (3-1) into its signal and noise components, we can rewrite (3-6) as

$$\hat{d}(t - T_s) = \mathrm{sgn}\left\{K_1K_{1m}\left[\sqrt{P_D}\int_0^{T_s} d(t)\cos\phi(t)\,dt + \int_0^{T_s} N_1[t,\phi(t)]\,dt\right]\right\} \tag{3-8}$$

where

$$N_1[t,\phi(t)] \triangleq n_1(t)\cos\phi(t) - n_2(t)\sin\phi(t) \tag{3-9}$$

is also a baseband white gaussian noise process of single-sided spectral density N_0 W/Hz. Furthermore, $N_1[t,\phi(t)]$ and $N_2[t,\phi(t)]$ are uncorrelated Also, for simplicity of notation, we omit hereafter the dependence on t in all variables which are functions of time

When the two-sided loop bandwidth W_L of the DAL is small relative to the data rate $\mathcal{R}_s = 1/T_s$, then the phase error ϕ can be considered to be constant for many T_s second intervals Under these conditions, the conditional error probability of the data detector, i.e., the probability that $d \neq \hat{d}$, given a phase error of ϕ rad, is essentially independent of t and given by

$$P_e^s(\phi) = \frac{1}{2}\,\mathrm{erfc}\,(\sqrt{R_s}\cos\phi) \tag{3-10}$$

where

$$R_s = \frac{P_D T_s}{N_0} \tag{3-11}$$

and

$$\mathrm{erfc}\,x \triangleq \frac{2}{\sqrt{\pi}}\int_x^\infty \exp(-t^2)\,dt \tag{3-12}$$

Furthermore, insofar as its effect on the loop equation is concerned, the product $d\hat{d}$ can be replaced by its statistical average, i.e., $E\{d\hat{d}\} = (+1)\times Pr(d = \hat{d}) + (-1)\times Pr(d \neq \hat{d}) = (+1)\times[1 - P_e^s(\phi)] + (-1)\times P_e^s(\phi) = 1 - 2P_e^s(\phi)$. The reason for this is that the loop filter being low-pass in nature will tend to extract the time average

(or dc value) of the random signal $d\hat{d}$, and since $d\hat{d}$ is ergodic, its time and statistical averages are equal Using these assumptions, Eq. (3-4) reduces to

$$z = K_2 K_{2m} F(p) \exp[-pT_s] \times \{\sqrt{P_D}\,\mathrm{erf}(\sqrt{R_s}\cos\phi)\sin\phi + \hat{d}N_2(t,\phi)\} \tag{3-13}$$

where $\mathrm{erf}\, x \triangleq 1 - \mathrm{erfc}\, x$.

It is clear that, in general, the transfer function factor $\exp(-pT_s)$, with $p = j\omega$, affects loop stability and reduces the signal acquisition or pull-in range as discussed in Sections D and E. For the moment, we shall ignore these problems and make a simplifying assumption which enables us to neglect this exponential factor in regard to predicting steady-state performance If $W_L T_s << 1$ (i e, the usual case of interest), then $\exp(-j\omega T_s)$ is approximately unity for all ω within the bandwidth of the DAL. Hence, from a steady-state performance standpoint, this factor has negligible effect Mechanistically, however, the delay in the lower loop is important in assuring that d is multiplied by the $\hat{d}$ corresponding to the same symbol interval.

Making the usual assumption that the instantaneous VCO phase θ is related to its input z through

$$\hat{\theta} = \frac{K_V}{p} z \tag{3-14}$$

where K_V is the gain of the VCO, we find, from Eqs. (3-13) and (3-14), and the fact that $\phi = \theta - \hat{\theta}$,

$$\dot{\phi} = \dot{\theta} - KF(p)\left[\sqrt{P_D}\,\mathrm{erf}(\sqrt{R_s}\cos\phi)\sin\phi + \hat{d}N_2(t,\phi)\right] \tag{3-15}$$

where $K \triangleq K_2 K_{2m} K_V$ represents the loop gain Equation (3-15) represents the stochastic integro-differential equation of loop operation for the configuration illustrated in Fig. 3-2.

C. Steady-State Performance

It is simplest at the outset to consider the steady-state performance characteristics of the *first-order* DAL, where the loop filter $F(p)$ is of the form $F(p) = 1$ The extension of the results which follow to the second-order DAL is straightforward based upon the contribution made in Ref. 3-1 The essential details of this extension will be presented at the end of our discussion

The steady-state performance of any carrier-tracking loop operating in its nonlinear region can be characterized by (1) its phase noise σ_ϕ^2 (i.e., the variance of the modulo-2π reduced phase error) and (2) its average rate of cycle slipping $\bar{S}$. With respect to calculation of the first of the above two performance parameters, it is essential that we characterize the probability density function (p.d.f.) of the modulo-2π reduced phase error, which for simplicity we shall also denote by ϕ Hereafter, when we use the term phase error, we shall implicitly mean the modulo-2π reduced version unless otherwise specified

1. Steady-State Phase Error Probability Density Function

Using results given in Ref. 3-1 (Eq. 32), it is easy to show that the p.d f of the phase error process is given by

$$p(\phi) = N \exp[-U_0(\phi)] \int_{\phi}^{\phi+2\pi} \exp[U_0(x)]\, dx, \quad |\phi| \leq \pi \tag{3-16}$$

where N is the normalization constant and the potential function $U_0(\phi)$ is given by

$$U_0(\phi) = -\frac{2}{K_{00}}\left\{\Omega_0\phi + K\sqrt{P_D}\left[\cos\phi\,\mathrm{erf}(\sqrt{R_s}\cos\phi) + \frac{1}{\sqrt{\pi R_s}}\exp(-R_s\cos^2\phi)\right]\right\} \tag{3-17}$$

with

$$K_{00} = \frac{N_0 K^2}{2} \tag{3-18}$$

The loop gain K can be related to the system parameters by

$$K = \frac{4}{\delta\sqrt{P_D T_s}\,\mathrm{erf}\sqrt{R_s}} \tag{3-19}$$

where

$$\delta \triangleq \frac{2}{W_L T_s} \tag{3-20}$$

and

$$W_L \triangleq \frac{\sqrt{P_D}K \operatorname{erf}\sqrt{R_s}}{2} \tag{3-21}$$

is the two-sided bandwidth of the DAL. Substituting (3-18), (3-19), and (3-20) into (3-17) gives, upon simplification,

$$U_0(\phi) = -\frac{4\Omega_0}{N_0K^2}\phi - R_s\delta \operatorname{erf}\sqrt{R_s} \times\left[\cos\phi \operatorname{erf}(\sqrt{R_s}\cos\phi) + \frac{1}{\sqrt{\pi R_s}}\exp(-R_s\cos^2\phi)\right] \tag{3-22}$$

In terms of the loop signal-to-noise ratio,

$$\boxed{\rho \triangleq \frac{4\sqrt{P_D}\operatorname{erf}\sqrt{R_s}}{N_0K} = \frac{2P_D\operatorname{erf}^2\sqrt{R_s}}{N_0W_L} = R_s\delta\operatorname{erf}^2\sqrt{R_s}} \tag{3-23}$$

and the steady-state phase error (often called *static phase error* or SPE),

$$\boxed{\phi_{ss} = \sin^{-1}\left[\frac{\Omega_0}{\sqrt{P_D}K\operatorname{erf}\sqrt{R_s}}\right]} \tag{3-24}$$

Eq. (3-22) may be rewritten as

$$U_0(\phi) = -(\rho\sin\phi_{ss})\phi - \rho\left[\frac{\cos\phi\operatorname{erf}(\sqrt{R_s}\cos\phi) + (1/\sqrt{\pi R_s})\exp(-R_s\cos^2\phi)}{\operatorname{erf}\sqrt{R_s}}\right] \tag{3-25}$$

We note that when $\Omega_0 = 0$, Eq. (3-25) reduces to

$$U_0(\phi) = -\rho\frac{\cos\phi\operatorname{erf}(\sqrt{R_s}\cos\phi) + (1/\sqrt{\pi R_s})\exp(-R_s\cos^2\phi)}{\operatorname{erf}\sqrt{R_s}} \tag{3-26}$$

For a first-order PLL, the potential function $U_0(\phi)$ is simply

$$U_0(\phi) = -\rho\cos\phi = \rho\int\sin\phi\,d\phi \tag{3-27}$$

Thus, the effective loop nonlinearity (or S-curve) of the DAL is given by

$$\mathcal{S}(\phi) = -\frac{d}{d\phi}\left[\frac{\cos\phi\operatorname{erf}(\sqrt{R_s}\cos\phi) + (1/\sqrt{\pi R_s})\exp(-R_s\cos^2\phi)}{\operatorname{erf}\sqrt{R_s}}\right]$$

or

$$\boxed{\mathcal{S}(\phi) = \frac{\operatorname{erf}(\sqrt{R_s}\cos\phi)\sin\phi}{\operatorname{erf}\sqrt{R_s}}} \tag{3-28}$$

and is plotted vs ϕ in Fig. 3-3, with R_s as a parameter. We note that in the limit of infinite data signal-to-noise ratio, (3-28) becomes

$$\boxed{\mathcal{S}(\phi) = \operatorname{sgn}(\cos\phi)\sin\phi = S_q(\phi)\sin\phi} \tag{3-29}$$

where $S_q(\phi)$ is a unit power square wave in ϕ with period equal to 2π, i.e.,

$$S_q(\phi) = \begin{cases} +1 & -\frac{\pi}{2} + 2n\pi \leq \phi \leq \frac{\pi}{2} + 2n\pi \\ -1 & \frac{\pi}{2} + 2n\pi \leq \phi \leq \frac{3\pi}{2} + 2n\pi \end{cases} \qquad n = 0, \pm1, \pm2, \cdots \tag{3-30}$$

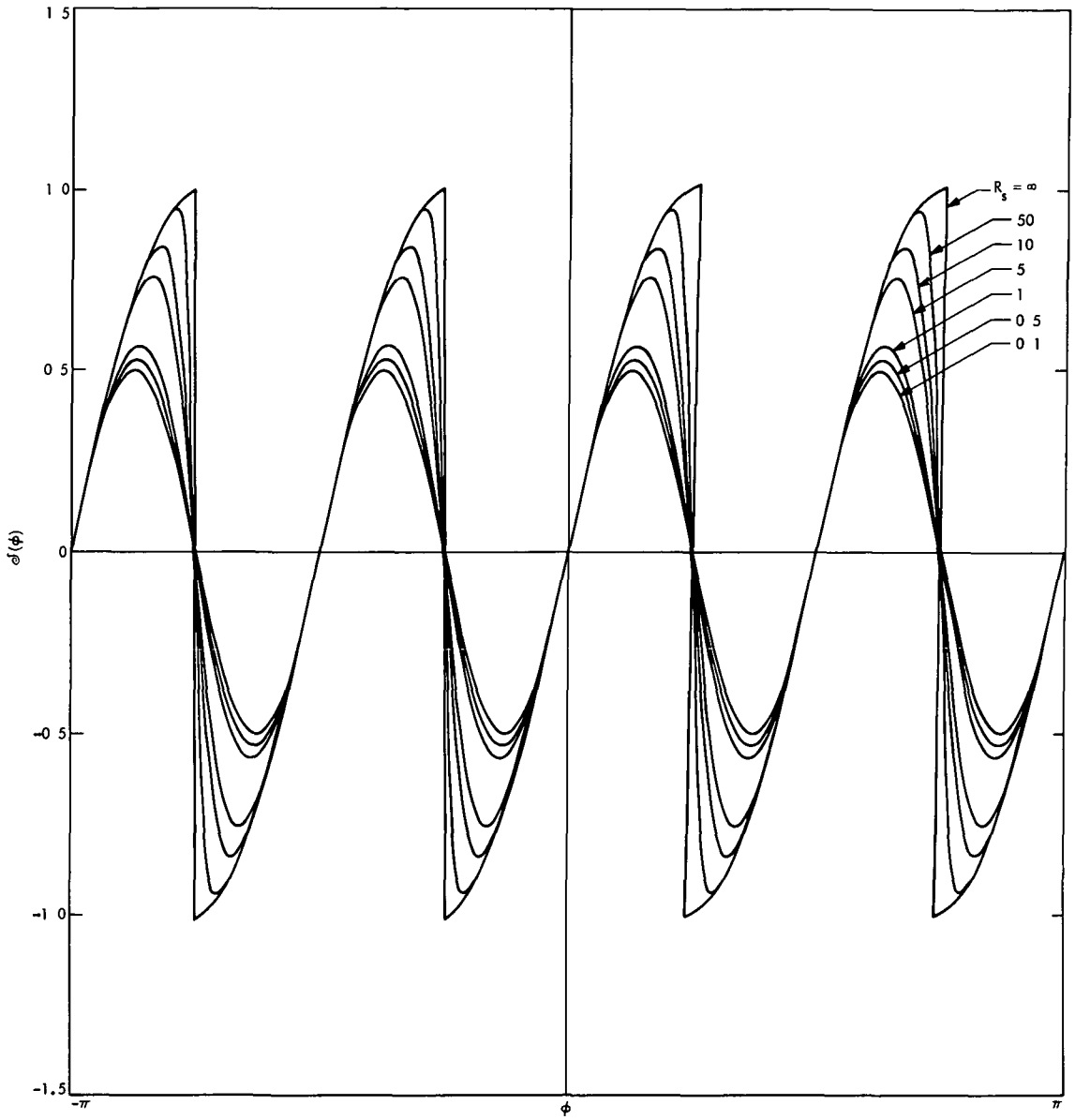

Fig. 3-3. Equivalent loop nonlinearity (S-curve) of the DAL

For $R_s \to 0$, we get

$$\mathcal{S}(\phi) = \frac{1}{2} \sin 2\phi$$

Figure 3-4 illustrates the phase error p.d f., $p(\phi)$, for $\Omega_0 = 0$, $\delta = 10$, and R_s as a parameter. Notice from this figure that the DAL exhibits the same 180-deg phase ambiguity as the squaring or Costas loop, because the stable lock points are separated by only 180 deg. This ambiguity can be resolved in several different ways: (1) differentially encode and decode the symbols, (2) search for and recognize the phase of unique sync words, or (3) for uplinks (ground to spacecraft), use the procedure outlined in Section VI-A.

2. Phase Noise Performance

Assuming that the phase ambiguity is perfectly resolved, then the phase noise σ_ϕ^2 can be computed numerically from

$$\sigma_\phi^2 = 2\int_{-\pi/2}^{\pi/2} \phi^2 p(\phi)\, d\phi - \left[2\int_{-\pi/2}^{\pi/2} \phi p(\phi)\, d\phi \right]^2 \qquad (3\text{-}31)$$

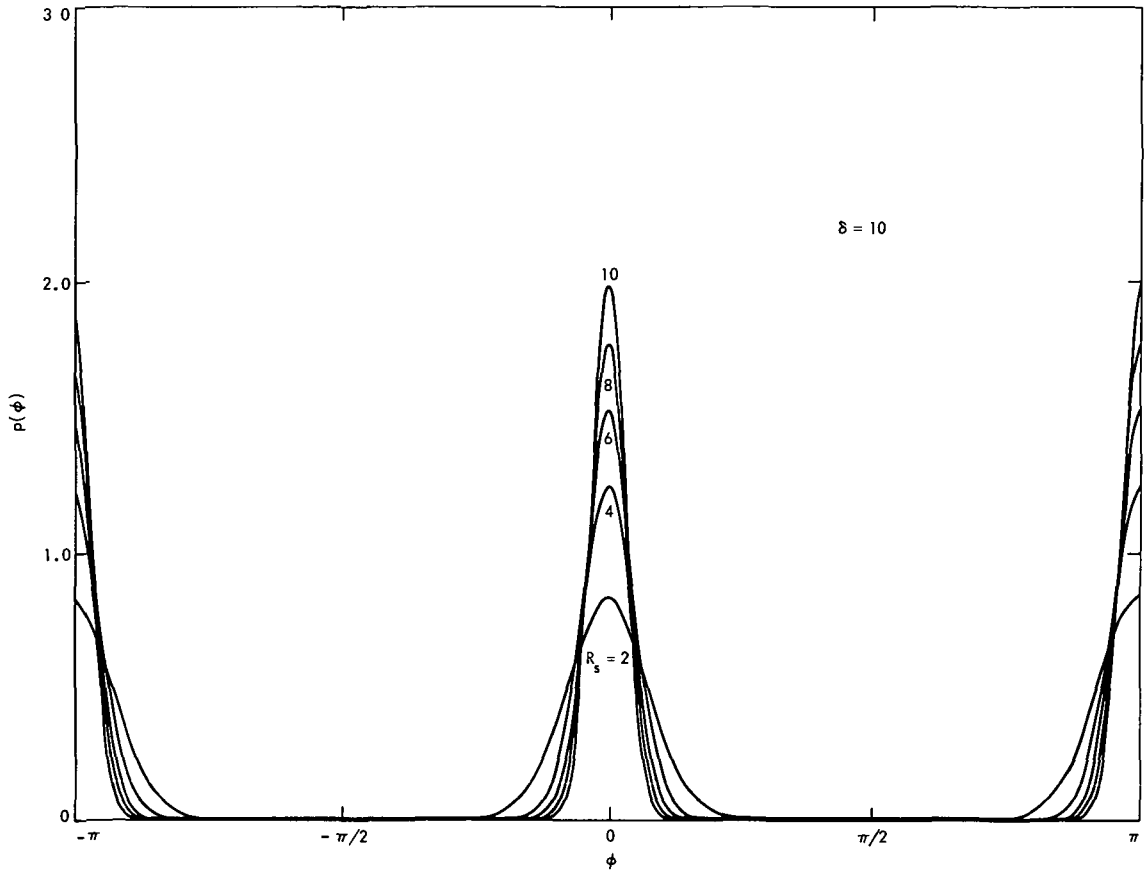

Fig. 3-4. Probability density function of the loop phase error

where $p(\phi)$ is given by (3-16) together with (3-25). When R_s becomes large, σ_ϕ^2 behaves as follows:

$$\sigma_\phi^2 = \frac{\int_0^{\pi/2} \phi^2 \exp(\rho \cos\phi)\, d\phi}{\int_0^{\pi/2} \exp(\rho \cos\phi)\, d\phi} \tag{3-32}$$

which, for large ρ, becomes

$$\sigma_\phi^2 = \frac{1}{\rho} \tag{3-33}$$

Thus, insofar as phase noise is concerned, the *first-order DAL has identical behavior in its linear region of operation as the PLL.* This result is in contrast to the other types of suppressed carrier-tracking loops discussed in Section II, whose phase noise in the linear region is degraded by the squaring loss.

In comparing the carrier-tracking efficiency of discrete and suppressed carrier systems, it is important that the comparison be made at a fixed value of total transmitter energy-to-noise ratio $P_T T_s/N_0$ and symbol time-bandwidth product δ. For the suppressed carrier system (employing a DAL for carrier tracking), the data signal-to-noise ratio R_s is identical to $P_T T_s/N_0$. Thus, the loop signal-to-noise ratio ρ is a direct function of δ and $P_T T_s/N_0$ through Eq. (3-23). For the discrete carrier system (employing a PLL for carrier tracking), the loop signal-to-noise ratio $\rho = 2P_c/N_0 W_L$ is related to $P_T T_s/N_0$, R_s, and δ by

$$\rho = [(P_T T_s/N_0) - R_s]\,\delta \tag{3-34}$$

Also,

$$\sigma_\phi^2 = \int_{-\pi}^{\pi} \phi^2 \frac{\exp(\rho \cos\phi)}{2\pi I_0(\rho)}\, d\phi$$

$$= \frac{\pi^2}{3} + \frac{4}{I_0(\rho)} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} I_k(\rho) \tag{3-35}$$

Figures 3-5 and 3-6 plot σ_ϕ^2 vs $P_T T_s/N_0$ for fixed values of $\delta = 10$ and $\delta = 50$, respectively. For the discrete carrier system, σ_ϕ^2 is computed from Eq. (3-35) together with (3-34), where R_s has been set equal to unity and the range of $P_T T_s/N_0$ is selected to correspond to a range of values for ρ of 0 to 10 dB. The value of R_s was selected to correspond to the minimum acceptable error probability performance in the presence of a perfect carrier reference For the suppressed carrier system, σ_ϕ^2 is computed from Eqs. (3-31) and (3-16) and plotted over the same range of $P_T T_s/N_0$ as the curves for the discrete carrier case We note from the figures that at each value of $P_T T_s/N_0$, the DAL has a smaller σ_ϕ^2 than the PLL. Furthermore, since for the PLL, the relation for R_s given by (3-11) implies a zero value of loop signal-to-noise ratio (see Eq. 3-34) or equivalently, an infinite value of σ_ϕ^2, then for any finite value of σ_ϕ^2 and a fixed value of $P_T T_s/N_0$, the suppressed carrier system allows for a larger value of R_s than the discrete carrier system. This increase in data signal-to-noise ratio manifests itself in an improved error probability performance.

In conclusion, the suppressed carrier system with a DAL has a better error probability performance than the discrete carrier system with a PLL at any given value of transmitter energy-to-noise ratio because (1) the carrier reference has less jitter, resulting in a smaller noisy reference loss, and (2) the data signal-to-noise ratio is larger

3. Average Probability of Error of the Data Detector

When the loop phase error can be considered essentially constant over several symbol intervals, then the conditional probability of error of the data detector is given by Eq. (3-10). The average error probability (assuming that the 180-deg phase ambiguity is perfectly resolved) is then obtained by averaging $P_e^s(\phi)$ over the p d.f. $p(\phi)$ of the phase error as given by (3-16), i.e.,

$$P_e^s = 2\int_{-\pi/2}^{\pi/2} P_e^s(\phi)\, p(\phi)\, d\phi$$

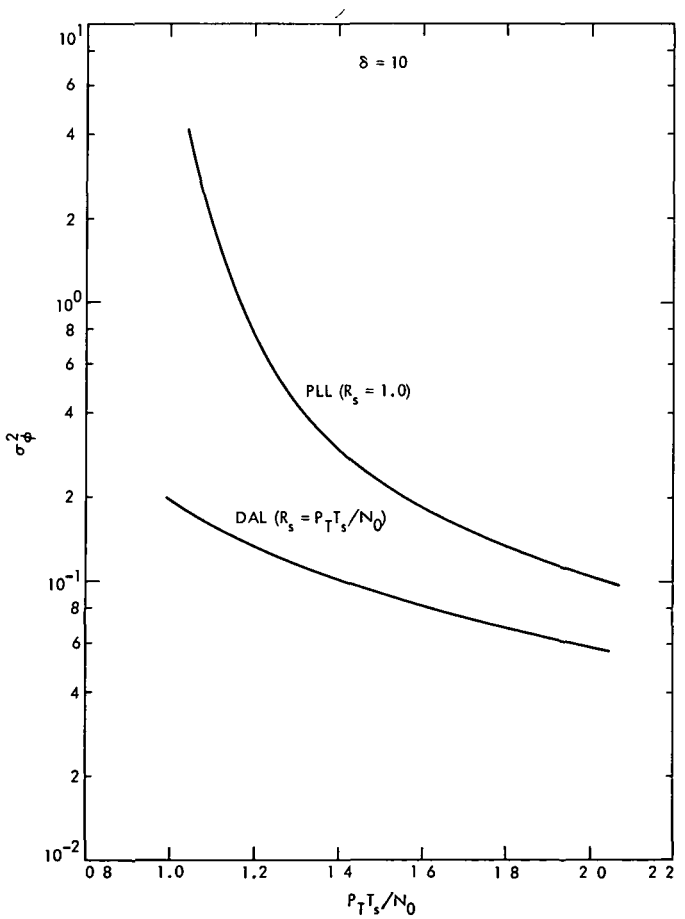

Fig. 3-5. Variance of the loop phase error as a function of total energy-to-noise ratio ($\delta = 10$)

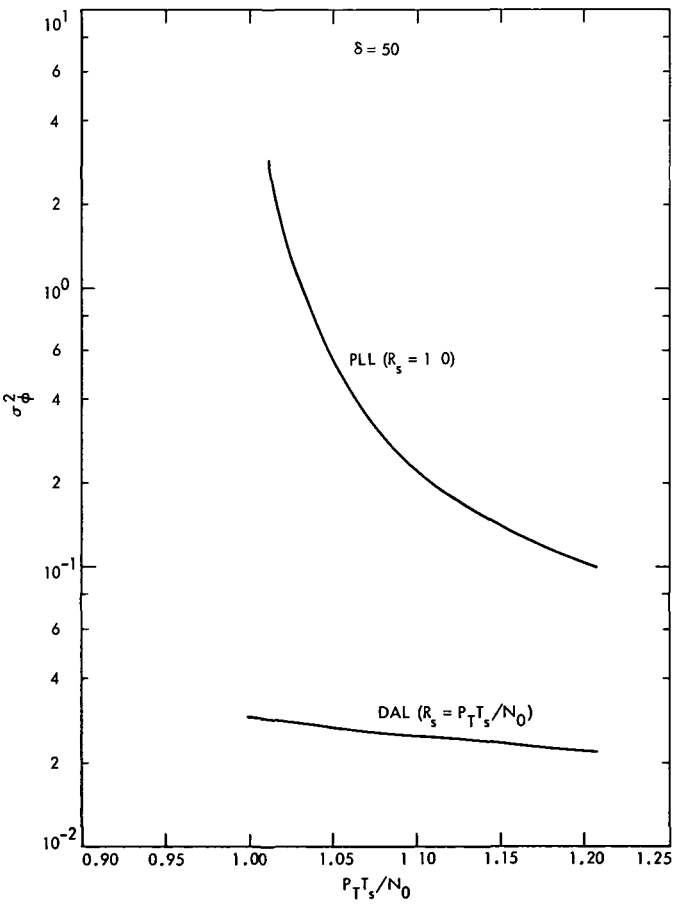

Fig. 3-6. Variance of the loop phase error as a function of total energy-to-noise ratio ($\delta = 50$)

4. Average Rate of Cycle Slipping

A general expression for the average rate of cycle slipping $\bar{S}$ of a first-order suppressed carrier-tracking loop is (Ref 3-2)

$$\bar{S} = -\frac{1}{2} C_0 D_0 K_{00} \coth \frac{\pi \Omega_0}{K_{00}} \tag{3-36}$$

where K_{00} is defined by (3-18), and

$$C_0 \triangleq \left[\int_{-\pi/2}^{\pi/2} \exp\left[-U_0(\phi)\right] \left\{ 1 + D_0 \int_{-\pi/2}^{\phi} \exp\left[U_0(x)\right] dx \right\} d\phi \right]^{-1}$$

$$D_0 \triangleq \frac{\exp\left[-U_0(-\pi/2)\right] - \exp\left[-U_0(\pi/2)\right]}{\exp\left[-U_0(\pi/2)\right] \int_{-\pi/2}^{\pi/2} \exp\left[U_0(\phi)\right] d\phi} \tag{3-37}$$

Since the potential function $U_0(\phi)$ can always be expressed in the form

$$U_0(\phi) = -\frac{2\Omega_0}{K_{00}} \phi + \rho \int \mathcal{S}(\phi)\, d\phi \tag{3-38}$$

then, after much simplification,

$$\bar{S} = \frac{K_{00} \cosh\left(\pi\Omega_0/K_{00}\right)}{\exp\left(\pi\Omega_0/K_{00}\right) \int_{-\pi/2}^{\pi/2} \exp\left[-U_0(\phi)\right] d\phi \int_{-\pi/2}^{\pi/2} \exp\left[U_0(\phi)\right] d\phi - 2 \sinh\left(\pi\Omega_0/K_{00}\right) \int_{-\pi/2}^{\pi/2} \exp\left[-U_0(\phi)\right] d\phi \int_{-\pi/2}^{\phi} \exp\left[U_0(x)\right] dx} \tag{3-39}$$

For zero detuning, Eq. (3-39) reduces to

$$\bar{S} = \frac{K_{00}}{\int_{-\pi/2}^{\pi/2} \exp\left[-U_0(\phi)\right] d\phi \int_{-\pi/2}^{\pi/2} \exp\left[U_0(\phi)\right] d\phi} \tag{3-40}$$

Applying (3-39) to the DAL, we get

$$\frac{\bar{S}}{W_L} = \frac{(4/\rho) \cosh\left[(\pi/2)\, \rho \sin \phi_{ss}\right]}{\exp\left[(\pi/2)\, \rho \sin \phi_{ss}\right] \int_{-\pi/2}^{\pi/2} \exp\left[-U_0(\phi)\right] d\phi \int_{-\pi/2}^{\pi/2} \exp\left[U_0(\phi)\right] d\phi - 2 \sinh\left[(\pi/2)\, \rho \sin \phi_{ss}\right] \int_{-\pi/2}^{\pi/2} \exp\left[-U_0(\phi)\right] d\phi \int_{-\pi/2}^{\phi} \exp\left[U_0(x)\right] dx} \tag{3-41}$$

which, for $\Omega_0 = 0$, reduces to

$$\frac{\bar{S}}{W_L} = \frac{4}{\rho \int_{-\pi/2}^{\pi/2} \exp\left[-\rho \int^{\phi} \mathcal{S}(x)\, dx\right] d\phi \int_{\pi/2}^{-\pi/2} \exp\left[\rho \int^{\phi} \mathcal{S}(x)\, dx\right] d\phi}$$

$$= \left\{ \rho \int_0^{\pi/2} \exp\left[-\rho \int^{\phi} \mathcal{S}(x)\, dx\right] d\phi \int_0^{\pi/2} \exp\left[\rho \int^{\phi} \mathcal{S}(x)\, dx\right] d\phi \right\}^{-1} \tag{3-42}$$

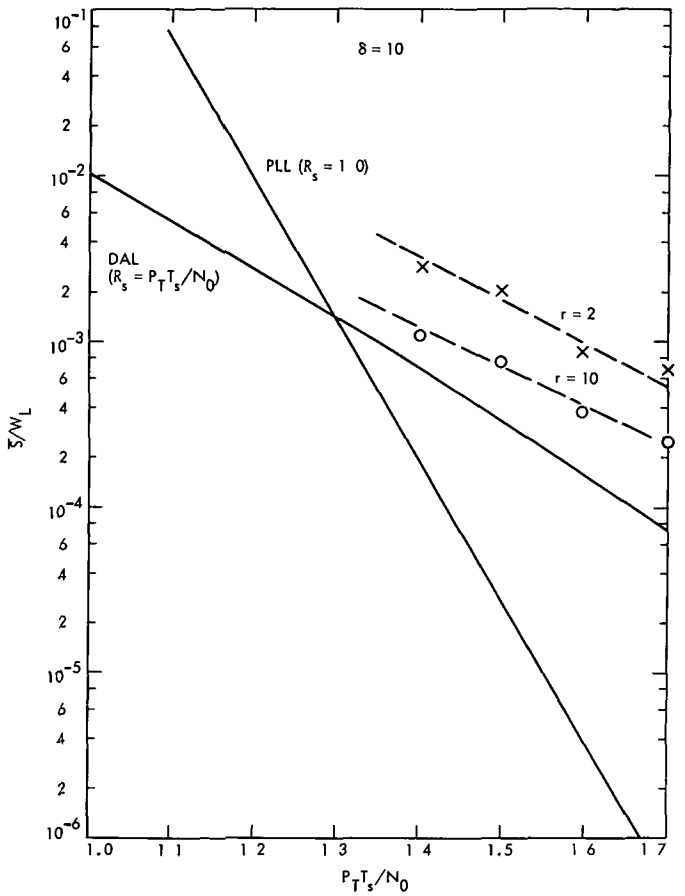

Fig. 3-7. Normalized average rate of cycle slipping as a function of total energy-to-noise ratio ($\delta = 10$)

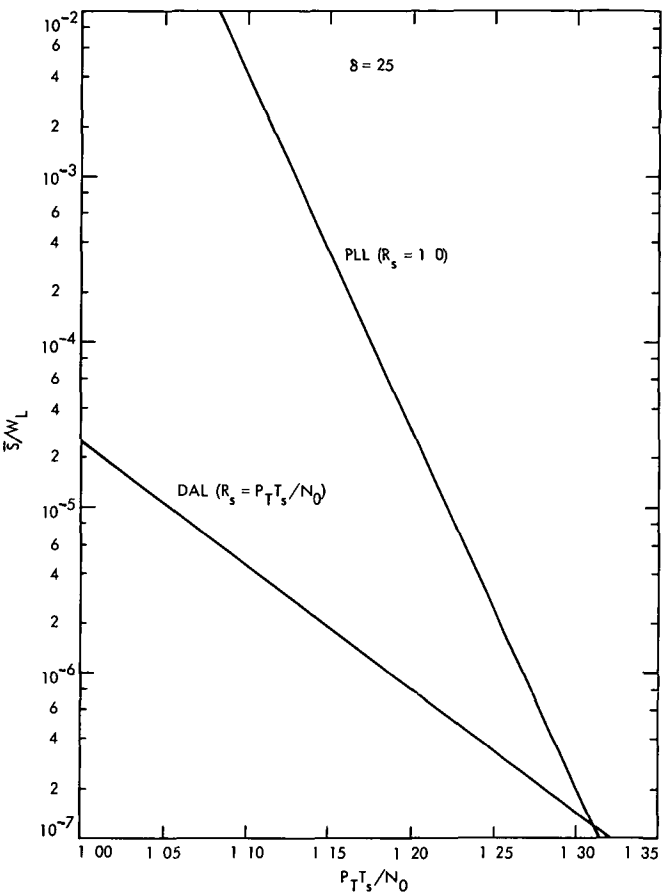

Fig. 3-8. Normalized average rate of cycle slipping as a function of total energy-to-noise ratio ($\delta = 25$)

Figures 3-7 and 3-8 illustrate the variation of $\bar{S}/W_L$, with $P_T T_s/N_0$ at fixed δ for the same parameters as in Figs, 3-5 and 3-6, respectively. The appropriate equation for the PLL is

$$\frac{\bar{S}}{W_L} = \frac{1}{\pi^2 \rho I_\delta^2(\rho)} \tag{3-43}$$

Also included in Fig 3-7 are some experimental points taken from a second-order DAL for the cases of $r = 10$ and $r = 2$. These data are included to verify the slope of the $\bar{S}/W_L$ curve for the DAL, as well as to provide an assessment of the validity of Eq. (3-36), when applied to second-order loops. Noting that the first-order DAL performance is approached as $r \to \infty$, the experimental data ($r = 10$) strongly confirm the analytical results.

Asymptotic (large ρ) approximations to (3-42) and (3-43) are possible For the PLL,

$$\frac{\bar{S}}{W_L} = \frac{2}{\pi} \exp(-2\rho) \tag{3-44}$$

while for the DAL,

$$\frac{\bar{S}}{W_L} = \sqrt{\frac{\rho\pi}{2}} \exp\left[-\left(2 - \frac{\pi^2}{8}\right)\rho\right] \tag{3-45}$$

5. Extension to Case of Second-Order Loops

For the second-order DAL with loop filter

$$F(s) = \frac{1 + s\tau_2}{1 + s\tau_1} \tag{3-46}$$

the p.d.f. of the phase error is still characterized by (3-16); however, the potential function of (3-25) must be modified as follows:

$$U_0(\phi) = -\beta\phi - \alpha\left[\frac{\cos\phi \operatorname{erf}(\sqrt{R_s}\cos\phi) + (1/\sqrt{\pi R_s}) \exp(-R_s\cos^2\phi)}{\operatorname{erf}\sqrt{R_s}}\right] \tag{3-47}$$

where

$$\alpha \triangleq \left(\frac{r+1}{r}\right)\rho - \frac{1-F_1}{r\overline{G^2(\phi)}}$$

$$\beta \triangleq \left(\frac{r+1}{r}\right)^2 \frac{\rho}{2W_L} \times [\Omega_0 - \sqrt{P_D}K \operatorname{erf}\sqrt{R_s}(1-F_1)\overline{\mathcal{S}(\phi)}] + \alpha\overline{\mathcal{S}(\phi)}$$

$$F_1 \triangleq \frac{\tau_2}{\tau_1}$$

$$W_L \triangleq \frac{r+1}{2\tau_2}$$

$$r \triangleq \sqrt{P_D}K\tau_2 F_1 \operatorname{erf}\sqrt{R_s}$$

$$G(\phi) \triangleq \mathcal{S}(\phi) - \overline{\mathcal{S}(\phi)} \tag{3-48}$$

and the overbar denotes statistical expectation with respect to the p.d.f. $p(\phi)$ We note that the above reduces to the results previously given for the first-order DAL by letting r approach infinity and $F_1 = 1$.

Using the p d.f. as characterized by the above parameters, the phase noise performance of the second-order DAL is still characterized by Eq. (3-31), and its asymptotic performance for large loop signal-to-noise ratio is given by (3-33), with ρ replaced by α. Insofar as the average rate of cycle slipping is concerned, Eq. (3-36) can be used to approximate the performance of the second-order DAL for large r provided that the error, as indicated by the experimental data in Fig 3-7 is tolerable. For small r, Eq. (3-36) is not very useful, as shown by the experimental data for $r = 2$ in Fig. 3-7 A similar variation with r has been shown for the standard PLL in Ref. 3-2 (Fig 10-22)

D. Effects of the Delay Function

Unlike the PLL or the other types of suppressed carrier loops discussed in Section II, the DAL requires an ideal delay element (see Fig. 3-2 and Section III-B) in its lower arm for proper operation. The primary purpose of this delay element is to delay the signal emanating from the lower arm cross-correlator by an amount equal to one symbol time T_s so that in the multiplier which follows, a particular data symbol is multiplied by the estimate of that same symbol; i.e., the matched filter in effect introduces a delay of one symbol time Aside from its necessary value, the presence of a delay element in a closed-loop configuration produces several deleterious effects on the loop's performance. The first of these to be considered is the effect of the delay element on loop stability.

1. Effect of Delay Element on Loop Stability

In the past, many techniques (e.g., the Nyquist criterion, root-locus diagram, Routh–Hurwitz criterion, phase-plane, Lyaponov function) have been developed for performing a stability analysis on a closed-loop system Unfortunately, all of these available techniques are applicable only in a noise-free situation In the case of the second-order PLL, this restriction on applicability does not present a problem since a linear, noise-free analysis predicts unconditional stability, whereas the nonlinear techniques demonstrate the existence of an adequate pull-in range For the DAL, the problem of loop stability becomes more poignant because of the presence of a delay element in the open-loop transfer function. This delay, which is ideally equal to the reciprocal of the data rate, will, at sufficiently low data rates, cause instability problems even on a linear, noise-free basis. Furthermore, the nonlinear stability problem is complicated by the fact that the loop S-curve $\mathcal{S}(\phi)$ is now a function of signal-to-noise ratio, and hence a noise-free analysis does not tell the whole story To get around this problem, we shall consider only the nonlinear stability problem in the absence of noise, in which case the loop S-curve is given by Eq (3-29). The results obtained from this special case can then be thought of as the "best possible" nonlinear stability performance and, together with those obtained from the linear analysis, shall be used to provide an approximate answer to the question of how low in data rate one can go before the DAL becomes unstable.

The principal tool used in the linear stability analysis which follows is the Nyquist diagram (and the accompanying Nyquist criterion) as applied to the open-loop, noise-free transfer function of a generalized second-order tracking loop. For the nonlinear stability performance, we resort to a method based upon an approximation to the phase-plane behavior

a. Linear Stability Analysis. Consider a generalized phase-locked tracking system whose noise-free differential equation of operation is given by

$$\phi = \theta - AK\frac{F(p)}{p}\exp(-pT_d)\mathcal{S}(\phi) \tag{3-49}$$

where θ is the input phase to be tracked, $\phi = \theta - \hat{\theta}$ is the loop phase error, $\hat{\theta}$ is the loop's estimate of θ, $F(p)$ is the loop filter, $\mathcal{S}(\phi)$ is the system nonlinearity normalized

to unit slope at the origin, AK is the open-loop gain, and T_d is an arbitrary ideal delay. For the DAL, $\mathcal{S}(\phi)$ is defined in (3-28), $A = \sqrt{P_D}\,\mathrm{erf}\sqrt{R_s}$, and $T_d = T_s$, i.e., the reciprocal of the data symbol rate. The open-loop transfer function $T(p) \triangleq \hat{\theta}(p)/\theta(p)$ is from Eq. (3-49).

$$T(p) = AK\frac{F(p)}{p}\exp(-pT_s) \tag{3-50}$$

where $\mathcal{S}(\phi)$ has been linearized to ϕ Substituting (3-46) into (3-50) and letting $p = j\omega$, we get

$$T(j\omega) = \frac{AK(1 + j\omega\tau_2)}{j\omega(1 + j\omega\tau_1)}\exp(-j\omega T_s) \tag{3-51}$$

Introducing the normalization $x = \omega\tau_1$ and letting $r \triangleq AK\tau_2^2/\tau_1$, Eq. (3-51) simplifies to

$$T(jx) = \frac{r(\tau_1/\tau_2)^2[1 + jx(\tau_2/\tau_1)]\exp[-jx(T_s/\tau_1)]}{jx(1 + jx)} \tag{3-52}$$

The quantity r is related to the loop damping ξ at zero delay by

$$r = 4\xi^2 \tag{3-53}$$

It is desired to study the magnitude and phase characteristics of $T(jx)$ as defined in Eq. (3-52) and from these, using the Nyquist criterion, determine the set of normalized frequencies $\{x_n\}$ at which the phase characteristic, arg $T(jx)$, equals $(2n - 1)\pi$ rad, $n = 0, 1, 2,$. Thus, the set $\{x_n\}$ is the solution to

$$\tan^{-1}\left[x_n\left(\frac{\tau_2}{\tau_1}\right)\right] - \tan^{-1}x_n - \frac{\pi}{2} - x_n\frac{T_s}{\tau_1} = (2n - 1)\pi \tag{3-54}$$

In particular, the solution of Eq. (3-54) for $n = 0$ (i.e., the first crossing of the horizontal axis by the Nyquist plot) must have a corresponding magnitude of $T(jx)$ which is less than unity in order to ensure stability. Thus, if the delay T_s is such that

$$-\tan^{-1}\left[x_0\left(\frac{\tau_2}{\tau_1}\right)\right] + \tan^{-1}x_0 + x_0\frac{T_s}{\tau_1} = \frac{\pi}{2} \tag{3-55a}$$

and

$$|T(jx)|^2 = \frac{r^2(\tau^1/\tau_2)^4[1 + x_0^2(\tau_2/\tau_1)^2]}{x_0^2[1 + x_0^2]} < 1 \tag{3-55b}$$

then the second-order tracking loop will be absolutely stable in the linear sense. Assuming $r\tau_1 >> \tau_2$, the solution of Eq. (3-55b) for x_0 is

$$x_0 > \left(\frac{r\tau_1}{\tau_2}\right)\sqrt{\frac{1 + \sqrt{1 + 4/r^2}}{2}} \tag{3-56}$$

and hence,

$$\tan^{-1}x_0 \cong \frac{\pi}{2} \tag{3-57}$$

Substituting expressions (3-56) and (3-57) in Eq (3-55a) gives

$$\boxed{\frac{T_s}{\tau_2} < \frac{\text{P.V}\ \{\tan^{-1}[r\sqrt{(1 + \sqrt{1 + 4/r^2})/2}]\}}{r\sqrt{(1 + \sqrt{1 + 4/r^2})/2}}} \tag{3-58}$$

where P.V. denotes the principal value. Equation (3-58) then represents a bound on normalized loop delay in order to guarantee linear stability. As an example, for $r = 2$ (i e , $\xi = 0.707$), T_s/τ_2 must be less than 0.52 (see Fig. 3-9)

The quantitative results given above rely heavily on the ability to make the small phase error assumption which allowed us to use the linear model and a linear stability criterion. In fact, we notice that the linear stability analysis determines only whether the steady-state phase error

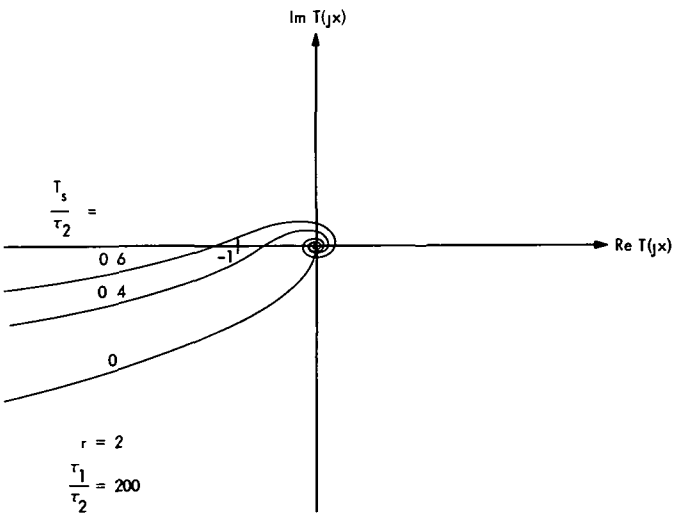

Fig. 3-9. Nyquist plot for second-order tracking loop with delay

becomes zero or unbounded and hence does not predict the existence of stable and unstable limit cycles Thus, one should regard the above results as a qualitative guide (possibly even as a loose upper bound) to the real behavior which must be predicted by a nonlinear model.

b. Nonlinear Stability Analysis. Despite the fact that we must simplify the nonlinear stability problem by assuming a noise-free situation—i e , the loop S-curve is given by (3-29)— an analytical solution for the critical value of delay above which the loop becomes unstable is difficult to come by In fact, the only solutions which have thus far been obtained for problems of this nature pertain only to the PLL. More specifically, for a PLL with a loop filter given by (3-46) and an arbitrary delay T_d in the loop, the absolute instability region is given by (Ref. 3-2, Eq 10-67)

$$\boxed{\frac{T_d}{\tau_2} > \sqrt{\frac{19.34\tau_2}{\tau_1 r^2}} \triangleq \eta_c} \tag{3-59}$$

In other words, for values of $T_d < \tau_2\eta_c$, the loop may or may not reach a stable lock point, depending, of course, on the initial frequency offset. If the initial frequency offset is too large, then the loop goes into a stable limit cycle These considerations on the loop's behavior in its region of conditional stability will be discussed in Section E.

It is interesting to note that the region of absolute instability as described by (3-59) also applies to a tracking loop with a sin 2ϕ characteristic, e.g , a squaring loop Thus, despite the fact that we cannot find an expression for the absolute instability region of the noise-free DAL, we realize that the S-curve of (3-29) is also a nonlinear function of 2ϕ, and on this basis alone, we might conjecture that (3-59) applies at least in a qualitative sense Further results of a quantative nature must at this point come from experimental evidence. Experimentally it has been found that for a DAL with $r = 2$ and $\tau_1/\tau_2 = 200$, a stable situation is achieved when $\mathcal{R}_s > 2W_L$, where W_L is the loop bandwidth at zero delay as given by (3-21). Since $\mathcal{R}_s = 1/T_s$ and $T_d = T_s$, then expressing the inequalities of (3-58) and (3-59) in terms of $\mathcal{R}_s$ and W_L, the corresponding inequalities necessary to guarantee stability become

$$\left.\begin{aligned} \mathcal{R}_s &> 1.3\,W_L \quad \text{(linear theory)} \\ \mathcal{R}_s &> 4.3\,W_L \quad \text{(nonlinear theory)} \end{aligned}\right\} \tag{3-60}$$

2. Effect of Delay Function on Loop Bandwidth and Natural Frequency

In order to assess the effect of the delay function on loop bandwidth and natural frequency, it is essential that we characterize the *linear* model of the DAL in terms of its closed-loop transfer function Letting $\mathcal{S}(\phi)$ be approximated by ϕ, then the closed-loop transfer function which relates the phase estimate $\hat{\theta}$ to the input phase θ is

$$H(s) \triangleq \frac{\hat{\theta}(s)}{\theta(s)} = \frac{AKF(s)\exp(-sT_s)}{s + AKF(s)\exp(-sT_s)} \tag{3-61}$$

where we have set $A = \sqrt{P_D}\,\mathrm{erf}\sqrt{R_s}$. Substituting (3-46) for $F(p)$ into the above yields, upon simplification,

$$H(s) = \frac{1 + s\tau_2}{1 + \{\tau_2 + [\exp(sT_s)]/AK\} + \{[\tau_1\exp(sT_s)]/AK\}\,s^2} \tag{3-62}$$

Letting

$$\begin{aligned} s' &= s\tau_2 \\ T' &= T_s/\tau_2 \end{aligned} \tag{3-63}$$

then using the definition of r as given in (3-48), we have the normalized form

$$H(s') = \frac{1 + s'}{1 + \{1 + [\exp(s'T')]/(r\tau_1/\tau_2)\} + \{[\exp(s'T')]/r\}\,s'^2} \tag{3-64}$$

If $r\tau_1 >> \tau_2$ (the usual case of interest), then

$$H(s') = \frac{1 + s'}{1 + s' + \{[\exp(s'T')]/r\}\,s'^2} \tag{3-65}$$

The linear loop bandwidth W_L is defined by

$$W_L = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} H(s)\,H(-s)\,ds \tag{3-66}$$

or equivalently,

$$W_L\tau_2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} H(s')\,H(-s')\,ds' \tag{3-67}$$

Letting $s' = j\omega'$,

$$W_L\tau_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(j\omega')|^2\,d\omega' \tag{3-68}$$

Substituting (3-65) into (3-68), we get

$$W_{L\tau_2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left| \frac{1 + j\omega'}{[1 - (\omega'^2/r)\cos\omega' T'] + j\omega'[1 + (\omega' \sin\omega' T')/r]} \right|^2 d\omega'$$

or

$$W_{L\tau_2} = \frac{1}{\pi}\int_{0}^{\infty} \frac{1 + \omega'^2}{1 + \omega'^2[1 - (2/r)\cos\omega' T'] - \omega'^3[(2/r)\sin\omega' T'] + \omega'^4/r^2} d\omega' \tag{3-69}$$

At zero delay, from (3-48),

$$W_{L\tau_2} = \frac{r+1}{2} \tag{3-70}$$

Plotted in Fig 3-10 is the ratio of $W_{L\tau_2}$ as given by (3-69) to its zero-delay value (Eq. 3-70) vs normalized delay $T' = T_s/\tau_2$, with r as a parameter We note from the figure that the presence of delay in the loop can significantly increase the effective linear loop bandwidth when the value of this delay (reciprocal of the data rate) relative to the filter time constant τ_2 becomes appreciable.

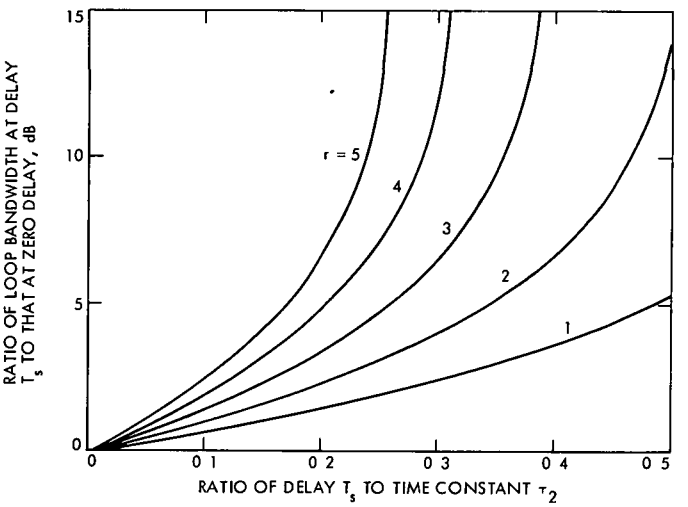

Fig. 3-10. Normalized loop bandwidth as a function of normalized delay

The natural frequency of the linear model of the DAL (or for that matter, any carrier-tracking loop) can be measured as follows (Fig. 3-11). A carrier (at frequency ω_C) is

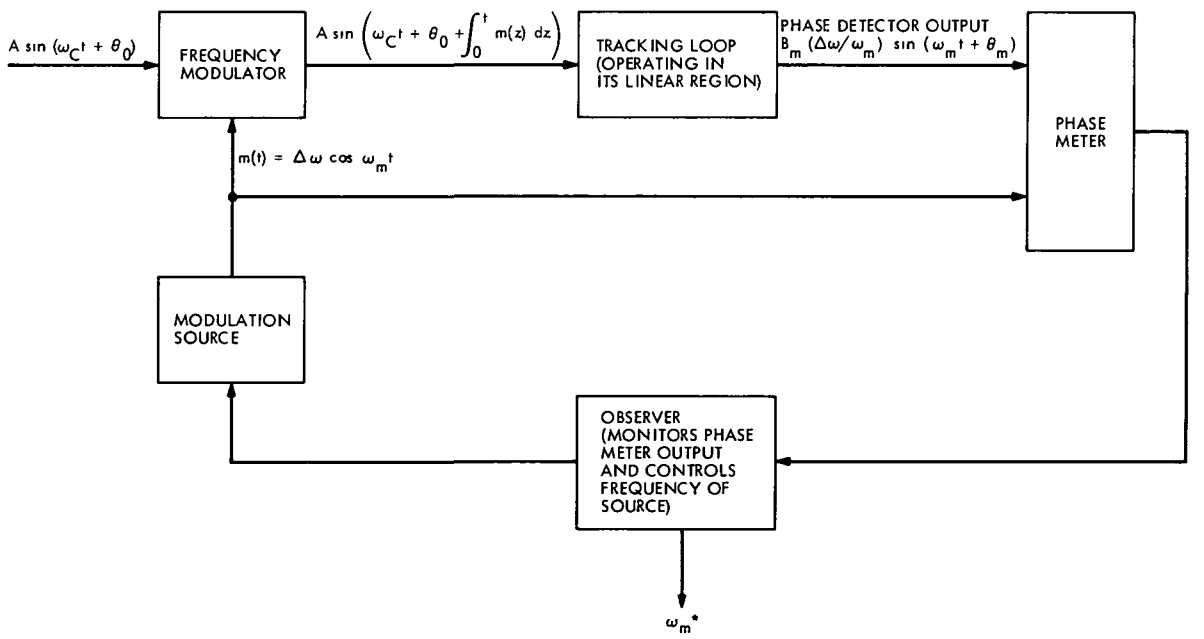

Fig. 3-11. Technique for measuring natural frequency of a tracking loop

frequency-modulated by a sinusoidal modulation $m(t) = \Delta\omega \cos \omega_m t$ to produce the signal

$$s(t) = A \sin [\omega_C t + \theta_1(t)] \tag{3-71}$$

where

$$\theta_1(t) = \theta_0 + \int_0^t m(z)\, dz = \theta_0 + \frac{\Delta\omega}{\omega_m} \sin \omega_m(t) \tag{3-72}$$

When this signal is applied to the DAL, the phase detector output (the phase error in the linear model) will be of the form

$$\epsilon(t) = B_m \frac{\Delta\omega}{\omega_m} \sin(\omega_m t + \theta_m) \tag{3-73}$$

The natural frequency is then that frequency ω_m^* at which the phase detector output is in phase with the input modulation $m(t)$, i.e., $\theta_m = \pi/2 + 2n\pi$. In mathematical terms, we begin by finding the transfer function which relates the phase error to the input phase modulation. For the linear loop model,

$$\Phi(s) = \frac{\theta_1(s)}{1 + AK \exp(-sT_s) F(s)/s} \tag{3-74}$$

where

$$\theta_1(s) = \frac{\theta_0}{s} + \frac{\Delta\omega}{s^2 + \omega_m^2} \tag{3-75}$$

Substituting (3-75) for $\theta_1(s)$ and (3-46) for $F(s)$ into (3-74) gives, upon simplification,

$$\Phi(s) = \frac{\omega_m}{s^2 + \omega_m^2} T(s) \tag{3-76}$$

where

$$T(s) = \frac{(1/\omega_m \tau_1)[\theta_0(s^2 + \omega_m^2) + s\Delta\omega][1 + \tau_1 s]}{s^2 + s\{[1 + AK\tau_2 \exp(-sT_s)]/\tau_1\} + (AK/\tau_1)\exp(-sT_s)} \tag{3-77}$$

Thus, from (3-76) and the above definition of natural frequency, we wish to find the frequency ω_m^* at which

$$\arg [T(j\omega_m^*)] = \frac{\pi}{2} + 2n\pi \tag{3-78}$$

Letting

$$\omega_m' = \omega_m^* \tau_2$$
$$T' = T_s/\tau_2$$

and assuming that $r\tau_1 >> \tau_2$, we get

$$\arg [T(j\omega_m')] = \frac{\pi}{2} + \tan^{-1}\left[\omega_m'\left(\frac{\tau_1}{\tau_2}\right)\right] + \omega_m' T' - \tan^{-1}\left\{\omega_m'\left[\frac{1 - (\omega_m'/r)\sin \omega_m' T'}{1 - (\omega_m'^2/r)\cos \omega_m' T'}\right]\right\} \tag{3-79}$$

or equivalently, ω_m' is the solution of the transcendental equation

$$\tan^{-1}\left[\omega_m'\left(\frac{\tau_1}{\tau_2}\right)\right] + \omega_m' T' - \tan^{-1}\left\{\omega_m'\left[\frac{1 - (\omega_m'/r)\sin \omega_m' T'}{1 - (\omega_m'^2/r)\cos \omega_m' T'}\right]\right\} = 2n\pi \tag{3-80}$$

For $T' = 0$ (zero delay) and $\tau_1/\tau_2 >> 1$, i e ,

$$\tan^{-1}\left[\omega_m' \frac{\tau_1}{\tau_2}\right] \cong \frac{\pi}{2}$$

the solution to (3-80) is

$$\omega_m' = \sqrt{r} \tag{3-81}$$

or

$$\omega_m^* = \frac{\sqrt{r}}{\tau_2}$$

which checks with a well known result for second-order loops (Ref. 3-2, Eq. 4-24). For $T' > 0$ and $\tau_1/\tau_2 >> 1$, ω_m' will be greater than $\sqrt{r}$, and we may still assume that $\tan^{-1}[\omega_m(\tau_1/\tau_2)] \cong \pi/2$. Using this assumption in (3-80), we find, after considerable simplification, that ω_m' must satisfy

$$\tan^{-1}\left[\frac{\sin \omega_m' T' - \omega_m' \cos \omega_m' T'}{\cos \omega_m' T' + \omega_m' \sin \omega_m' T' - (\omega_m'^2/r)}\right] = 2n\pi - \frac{\pi}{2} \tag{3-82}$$

or equivalently,

$$\boxed{\begin{aligned} &\sin \omega_m' T' - \omega_m' \cos \omega_m' T' < 0 \\ &\cos \omega_m' T' + \omega_m' \sin \omega_m' T' - (\omega_m'^2/r) = 0 \end{aligned}} \tag{3-83}$$

For small T' (small values of delay relative to τ_2), the solution to (3-83) is

$$\omega'_m = \frac{1}{\sqrt{(1/r) - T' + (T'^2/2)}} \tag{3-84}$$

More generally, the solution to (3-83) can be expressed implicitly as

$$\omega'_m = \frac{x}{T'} \tag{3-85}$$

where x is the inverse solution of

$$T' = \frac{x}{2}\left[\sqrt{1 + \frac{4}{r}\left(\frac{\cos x}{\sin^2 x}\right)} - 1\right]\tan x \tag{3-86}$$

Plotted in Fig 3-12 is the ratio of ω'_m as given by (3-85) to its zero delay value (Eq. 3-81) vs normalized delay $T' = T_s/\tau_2$, with r as a parameter. As was true in the case of loop bandwidth, the presence of delay in the loop can significantly increase the loop's natural frequency

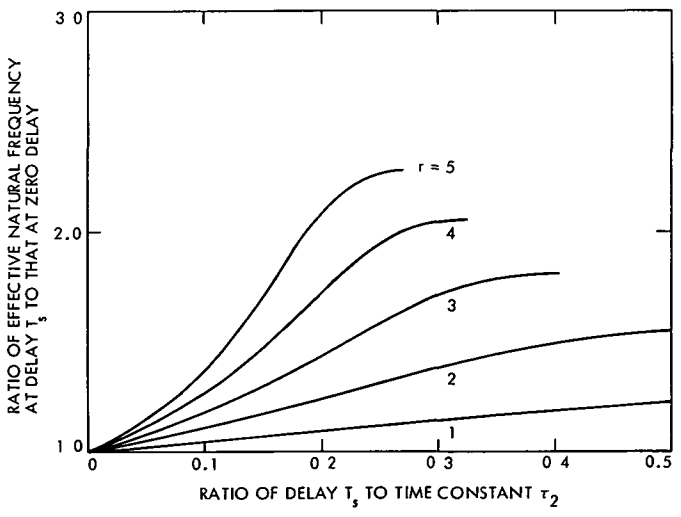

Fig. 3-12. Normalized loop natural frequency as a function of normalized delay

E. Data-Aided Loop Acquisition

The problem of characterizing the acquisition properties of the DAL bears directly on the loop's nonlinear stability performance discussed previously. In other words, in the region of stability, the loop will either acquire (achieve a stable lock point) or go into a stable limit cycle Unfortunately, because of the extreme difficulties of the analysis, we cannot quantitatively characterize the DAL's acquisition performance, however, as before, we can infer certain qualitative measures from the acquisition theory for PLLs with delay developed in Ref. 3-2. In particular, the function

$$\gamma(\omega_f) \cong \frac{\omega_f}{AK} + \frac{rT_s}{2\tau_2}\left[\frac{\cos \omega_f T_s}{\omega_f T_s} + \frac{\sin \omega_f T_s}{\omega_f T_s}\right] \tag{3-87}$$

defines the *synchronization boundary*, where ω_f is the fundamental beat note in a Fourier series expansion of $\phi(t)$. For any given value of normalized delay

$$T' \triangleq \frac{T_s}{\tau_2}$$

setting

$$\frac{\partial \gamma(\omega_f)}{\partial \omega_f} = 0 \tag{3-88}$$

gives in general several values of ω_f, say $\{\omega_{f_n}\}$ (see Fig. 3-13), which satisfy Eq. (3-88). Calling the smallest of these values ω_{f_0}, and denoting the initial frequency offset by Ω_0, the following conclusions may be drawn.

(1) If $\Omega_0/AK < \gamma(\omega_{f_0})$, the loop will lock.

(2) If $\Omega_0/AK > \gamma(\omega_{f_0})$, frequency pushing occurs, and the loop goes into a stable limit cycle

Expanding upon the second conclusion, we note that for any given small value of normalized delay T' and a value of $\Omega_0/AK > \omega_{f_0}$, there will be two solutions to the equation

$$\frac{\Omega_0}{AK} = \gamma(\omega_f) \tag{3-89}$$

namely, $\gamma(\omega_{f_1})$ and $\gamma(\omega_{f_2})$, which represent equilibrium states with corresponding beat frequencies ω_{f_1} and ω_{f_2}. The significance of these equilibrium states is that they represent the conditions of *false lock* of the loop. If we order the beat frequencies such that $\omega_{f_1} > \omega_{f_2}$, then the state ω_{f_1} which occurs at a negative slope of the synchronization boundary is unstable, and the state ω_{f_2}, which is located on a positive slope of the synchronization boundary, is stable. As the delay in the loop is increased, we see that for the same value of loop detuning, more than two equilibrium states can exist. These states alternate between stable and unstable, much like the lock points of a

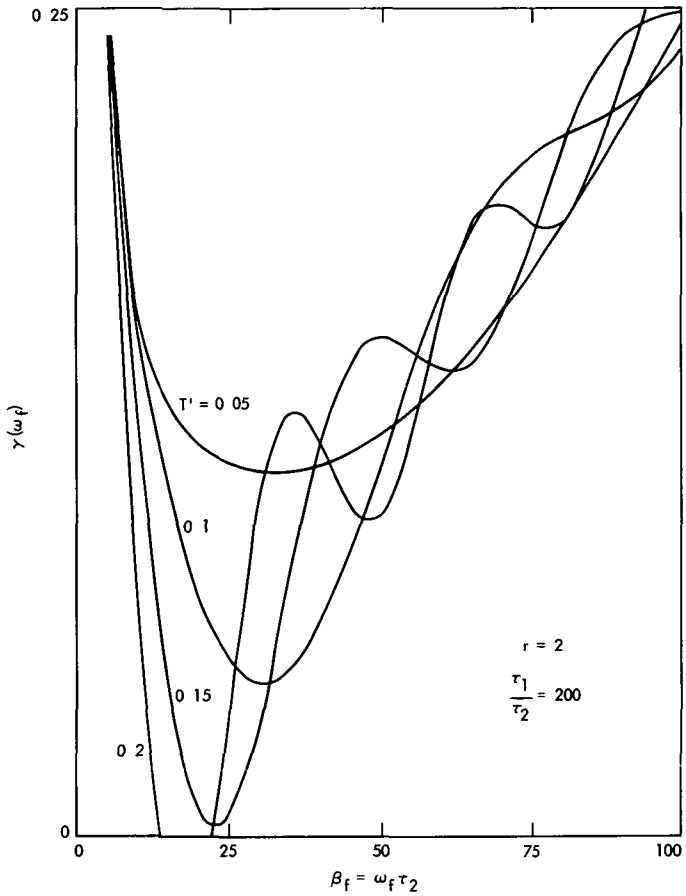

Fig. 3-13. Synchronization boundary of a second-order phase-locked loop with arbitrary delay

PLL with no delay As the delay is increased still further, the loop ultimately will not lock even if the VCO is at its quiescent frequency and the offset in frequency is zero This corresponds to the situation where the equation

$$\gamma(\omega_f) = 0 \tag{3-90}$$

has at least two solutions. For example, in Fig. 3-13, a value of $T' = 0.2$ represents such a situation. In this instance, the loop is said to be absolutely unstable; i.e., it can never lock. The critical value of delay at which this occurs is that value which simultaneously satisfies the equations

$$\gamma(\omega_f) = 0$$
$$\frac{\partial \gamma(\omega_f)}{\partial \omega_f} = 0 \tag{3-91}$$

This critical value of delay was given in Eq. (3-59) in connection with our discussion on nonlinear stability.

Returning now to the first conclusion, for a PLL with small delay $T_s << 1$, the loop will lock if

$$\boxed{\frac{|\Omega_0|}{AK} < \sqrt{2F_1 - \frac{2rW_L T_s}{r+1}}} \tag{3-92}$$

and $F_1 << 1$. Clearly, when no delay is present in the loop, (3-87) reduces to the well known result for the acquisition range of a PLL, namely,

$$\frac{|\Omega_0|}{AK} < \sqrt{2F_1} \tag{3-93}$$

Thus, we observe from (3-92) and (3-93) that the presence of delay in the loop reduces the acquisition range

Experiments have been conducted on the second-order sampled data hardware version (Section IV) of the DAL to determine the severity and number of false-lock points that can occur under a specified set of conditions To this end, the extreme case of $\delta = 10$ was chosen, together with the conditions that $r = 2$ and $AK = 10^4$. Acquisition tests were then conducted by offsetting the loop VCO from the nominal transmitted frequency by a certain amount when the loop was open*, and then closing the loop and observing the transition to the steady-state condition These tests were conducted in the absence of additive noise

Figure 3-14a shows the results for the normal DAL S-curve. The offset frequency Ω_0 (rad/s) has been normalized to the loop bandwidth False-lock points are indicated by the solid vertical lines topped by numbers. The vertical dashed lines represent instability region boundaries. Some instability boundaries are coincident with the false-lock points The arrows indicate the direction taken by the loop VCO once the loop is closed, assuming that the initial open-loop frequency offset falls somewhere between the relevant instability boundaries To illustrate this, for all initial $|\Omega_0/2\pi W_L| < 0.19$, the proper lock point is reached. For $0.19 < |\Omega_0/2\pi W_L| < 0.42$, the first false-lock point becomes the steady-state condition. For $0.5 < |\Omega_0/2\pi W_L| < 0.68$, the loop will tend to false-lock at point 3 The alternate regions of false lock are indicated by shaded arrows. The reasons why the false-lock points are distributed as shown, or why some instability region boundaries are coincident with lock points, is unknown Notice the definite subregions where frequency pushing

*The condition of open loop involves removing any residual memory from the loop filter $F(p)$

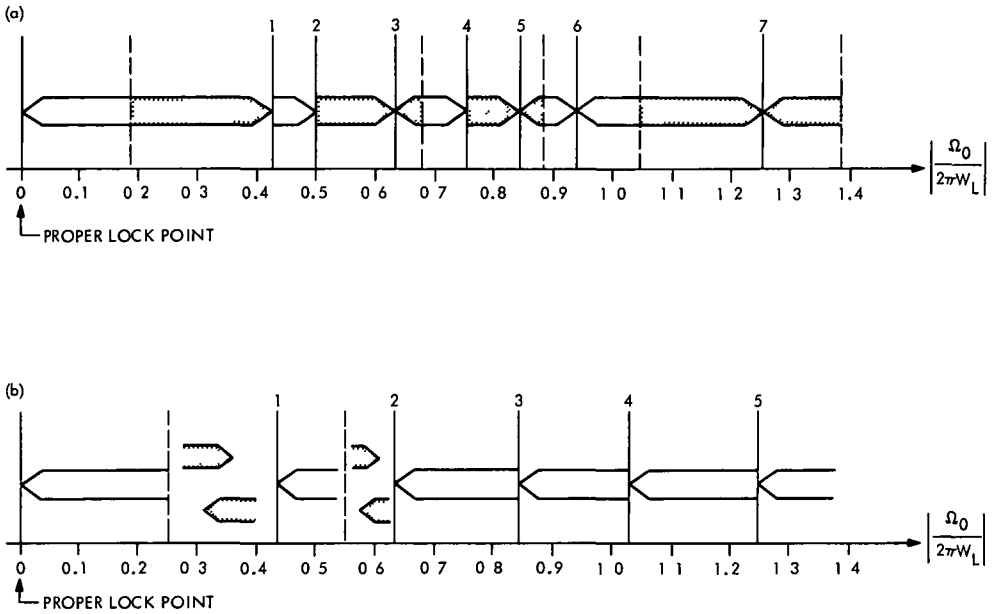

Fig. 3-14. DAL false-lock ranges for $\delta = 10$: a. normal S-curve, b. inverted S-curve

away from the proper lock point occurs. This would be indicated by any arrows that point to the right

A means for minimizing the number of false-lock points for a given δ and eliminating the frequency pushing is advanced by the following arguments

The acquisition range for a tracking loop with a rectangular phase detector characteristic, e.g., $\mathcal{J}(\phi) = \text{sgn}(\sin\phi)$ and no delay in the loop, requires that

$$\frac{|\Omega_0|}{AK} < 2\sqrt{F_1} \tag{3-94}$$

when $F_1 << 1$ Thus, we conclude from (3-93) and (3-94) that a loop with a rectangular characteristic has a wider acquisition range by a factor of $\sqrt{2}$ than a PLL with a sinusoidal characteristic

An expression analogous to (3-92) for a loop with a rectangular S-curve and arbitrary delay T_d is presently not available However, from the zero delay results of (3-93) and (3-94), we might conjecture that for small delays, (3-92) applies if it is multiplied by $\sqrt{2}$. Finally, if the loop S-curve is of the form $\mathcal{J}(\phi) = \sin 2\phi$ or $\mathcal{J}(\phi) = \text{sgn}(\cos\phi)$, then the results obtained for the acquisition range of the PLL and loop with $\mathcal{J}(\phi) = \text{sgn}(\sin\phi)$ apply, respectively, if Ω_0 is replaced by $2\Omega_0$, i e , the acquisition range of a tracking loop whose phase detector characteristic is compressed by a factor of 2 is reduced by a factor of 1/2.

Since the noise-free S-curve of the DAL behaves like $\sin 2\phi$ in the vicinity of the origin, and since the S-curve $\mathcal{J}(\phi) = \text{sgn}(\cos\phi)$ appears to give improved acquisition range, it is natural to ask if we can make the DAL behave like $\text{sgn}(\cos\phi)$ in the vicinity of the origin with a simple configuration change. Indeed, this is so and is essentially accomplished by inserting a minus one gain into the loop. This inverts the phase detector characteristic and causes it to shift by $\pi/2$ rad to achieve stable lock. Thus, the resulting noise-free S-curve becomes

$$\mathcal{J}(\phi) = -\sin\left(\phi + \frac{\pi}{2}\right)\text{sgn}\left[\cos\left(\phi + \frac{\pi}{2}\right)\right]$$

$$= \cos(\phi)\,\text{sgn}(\sin\phi)$$

After the loop has obtained a locked condition [as determined by a *lock detector* (Section V)], the sign inversion would be removed

Experimental evidence tends to indicate that the inverted and phase-shifted S-curve indeed results in an improved acquisition performance

For the conditions stated above, experiments were conducted using an inverted S-curve In addition, it was found that the open-loop gain had to be reduced by a factor of 150, and τ_2 increased by a factor of 3 for satisfactory performance The results are shown in Fig. 3-14b. Note first that the range for which the proper lock point is achieved is increased by a factor of about 1 3 Note

also that the trend for large offsets is to go in the direction toward lock, even though the loop comes to rest at false-lock points It has been observed that these false-lock points are not very strong, and it is quite easy to make the loop leave a false-lock point and seek one to the left. This suggests that if a swept acquisition ramp into the VCO is used, the loop could be acquired easily. Our final note is with respect to the two regions where small, shaded arrows point in both directions This is intended to convey the fact that in these regions, the loop tends to wander between the instability boundary and the false-lock point Thus, there appears to be little loop gain, and therefore control, in these regions.

Clearly, very little is known about acquisition for small δ Determining the best way to achieve lock for large frequency offsets in these cases will require further research and experimentation.

F. Performance of the DAL When Preceded by a Bandpass Limiter

In the practical operation of a coherent receiver employing a carrier-tracking loop, the loop is often preceded by a bandpass limiter (BPL) because the loop bandwidth is a function of the signal power, and any fluctuation in this quantity (owing to a change in range, for instance) causes the loop bandwidth to fluctuate similarly. BPLs are also used to protect various loop components, the multiplier in particular, where signal and noise levels can vary over several orders of magnitude and exceed the dynamic range of these components In this section, the theoretical results needed to explain the behavior of the DAL when preceded by a BPL are discussed The detail theory and operation of the BPL are covered more completely in Refs 3-2 and 3-3.

A block diagram of the DAL preceded by a BPL is shown in Fig 3-15. Since the success of the DAL depends on the signal sidebands, the IF filter must be sufficiently wide to pass the total modulation. Letting W_i denote the two-sided bandwidth of this filter, the signal-to-noise ratio at the BPL input is

$$\rho_i = \frac{2P_D}{N_0 W_i} \tag{3-95}$$

In designing phase-coherent receivers preceded by BPLs, it is convenient to characterize loop performance at a given operating point relative to its value at a fixed design point. In the past, it has been customary to specify this design point by the carrier signal-to-noise ratio in the

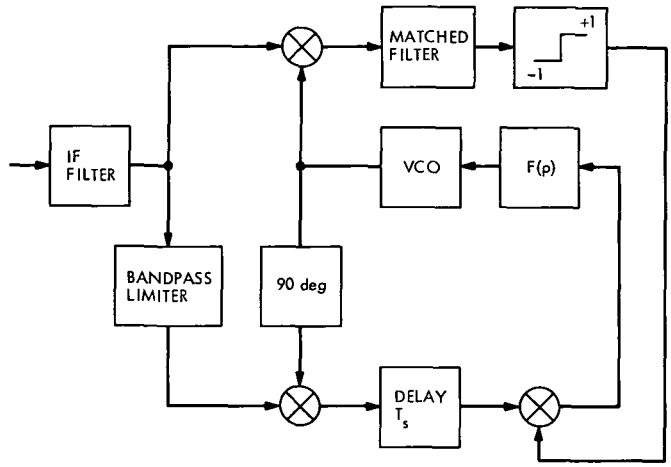

Fig. 3-15. The DAL preceded by a BPL

design point loop bandwidth. In the case of suppressed carrier transmission, the appropriate expression is

$$\Upsilon_0 = \frac{2P_{D0}}{N_0 W_{L0}}$$

where the zero subscript denotes the value of the parameter at the design point. Defining the *BPL suppression factor* $\tilde{\alpha}_1$ by

$$\tilde{\alpha}_1 = \left(\frac{\pi}{2}\right)^{1/2}\left(\frac{\rho_i}{2}\right)^{1/2} \exp\left(-\frac{\rho_i}{2}\right)\left[I_0\left(\frac{\rho_i}{2}\right) + I_1\left(\frac{\rho_i}{2}\right)\right] \tag{3-96}$$

then, using (3-25), the DAL signal-to-noise ratio can be expressed as

$$\rho = x\left(\frac{1}{\Gamma_\rho}\right)\left(\frac{1 + r_0}{1 + r_0/\mu}\right)$$

where

$$\begin{aligned} x &\triangleq \frac{2P_D \operatorname{erf}^2\sqrt{R_s}}{N_0 W_{L0}} \\ y &\triangleq \frac{W_{L0}}{W_i} \\ r_0 &\triangleq \tilde{\alpha}_{10}\sqrt{P_1}\,K \operatorname{erf}\sqrt{R_{s0}\,\tau_2 F_1} \\ \mu &\triangleq \frac{\tilde{\alpha}_{10}}{\tilde{\alpha}_1} \end{aligned} \tag{3-97}$$

The symbol $\tilde{\alpha}_{10}$ represents the value of $\tilde{\alpha}_1$ at the design point, where, by definition, $x = \Upsilon_0$. Also in the above,

$P_1 = 8/\pi^2$ is the total power in the first spectral zone and Γ_p is defined in Ref. 3-4 by*

$$\Gamma_p = \frac{1 - \exp(-\rho_s)}{(\pi/2)(\rho_s/2)\exp(-\rho_i)[I_0(\rho_s/2) + I_1(\rho_s/2)]^2\{1 + [(4/\Gamma_0\pi) - 1]\exp[-\rho_i(1 - \pi/4)]\}} \tag{3-98}$$

The p.d.f. of the phase error is still given by (3-16), together with the potential function defined in (3-47). However, the parameters α and β are now expressed in terms of the design point quantities by

$$\begin{aligned} \alpha &= \left(\frac{r_0/\mu + 1}{r_0/\mu}\right)\rho - \frac{1 - F_1}{(r_0/\mu)\,\overline{G^2(\phi)}} \\ \beta &= \left(\frac{r_0/\mu + 1}{r_0/\mu}\right)\left[\frac{\rho}{2W_{L0}}\left(\frac{r_0 + 1}{r_0/\mu + 1}\right)\right][\Omega_0 - \sqrt{P_D}K\,\mathrm{erf}\sqrt{R_s(1 - F_1)\,\overline{\mathcal{J}(\phi)}}] + \alpha\,\overline{\mathcal{J}(\phi)} \end{aligned} \tag{3-99}$$

where ρ is given by (3-97) Also note that since the BPL appears only in the lower arm of the DAL, the data signal-to-noise ratio is unaffected by its presence The entire penalty in data detection performance due to the presence of the BPL comes from the increased noisy reference loss.

*Γ_0 is the value of Γ_p at $\rho_i = 0$, which, for an ideal bandpass filter, is approximately equal to 1/0 862

References

3-1. Lindsey, W. C., "Nonlinear Analysis of Generalized Tracking Systems," *Proc. IEEE*, Vol. 57, No. 10, Oct. 1969, pp. 1705–1722.

3-2 Lindsey, W. C., *Synchronization Systems in Communication and Control,* Prentice-Hall Inc., Englewood Cliffs, N.J., 1972.

3-3. Springett, J. C., and Simon, M K., "An Analysis of the Phase Coherent–Incoherent Output of the Bandpass Limiter," *IEEE Trans. Comm. Technol*, Vol COM-19, No 1, Feb. 1971, pp. 42–49

IV. DAR Implementation

The subsequent sections deal with the implementation and performance of the *symbol detector* (matched filter and decision device) and *sychronizer*, the *delay function*, and the general organization of the data-aided loop (DAL) We are concerned with effective, economic implementations that will result in a DAR performance reasonably close to that obtained under optimum conditions

A. Digital vs Analog Mechanizations of the Entire DAR

Before proceeding with implementation details, it is necessary to develop a rationale which permits the DAR to operate over a large range of conditions such as input frequency, data rate, loop bandwidth, etc , without undue complications, the need for constant redesign, and with a minimum of actual component or block changes Cost and reliability are uppermost considerations.

One approach would make use of modern, operational-based analog circuits to achieve the desired functions Another would employ as many digital or saturated devices as possible, and produce the desired results by computationally operating on samples of the input signal(s)—in effect, a special-purpose computer. The exclusive use of analog circuits requires a large amount of analog signal and component switching when it becomes necessary to change conditions. Analog circuits are also subject to drift and tolerance problems. On the other hand, the digital embodiment of second- and third-order loop filters, voltage-controlled oscillators (VCOs), and RF phase detectors is quite complex. As a result, the present state-of-the-art, coupled with practical experience, indicates that a hybrid solution is best.

The phase detectors, which must generally operate at RF input frequencies in the range from 1 MHz to several tens of megahertz (or even higher for very wideband systems), are most easily built in the form of a ring-bridge diode configuration. At lower frequencies, it is possible to use MOSFET switching elements. The VCO is also conveniently constructed using analog circuitry—usually some form of voltage-controlled crystal oscillator. The loop filter, whether it be for a second- or third-order loop, consists of resistors and capacitors, together with some linear amplification, and is quite easily mechanized using operational configurations. All other functions of the basic DAR are readily implemented using logical elements. Figure 4-1 shows the DAR functional blocks for a hybrid mechanization; digital portions are indicated by broken lines. The analog-to-digital and digital-to-analog converters are labeled, respectively, ADC and DAC.

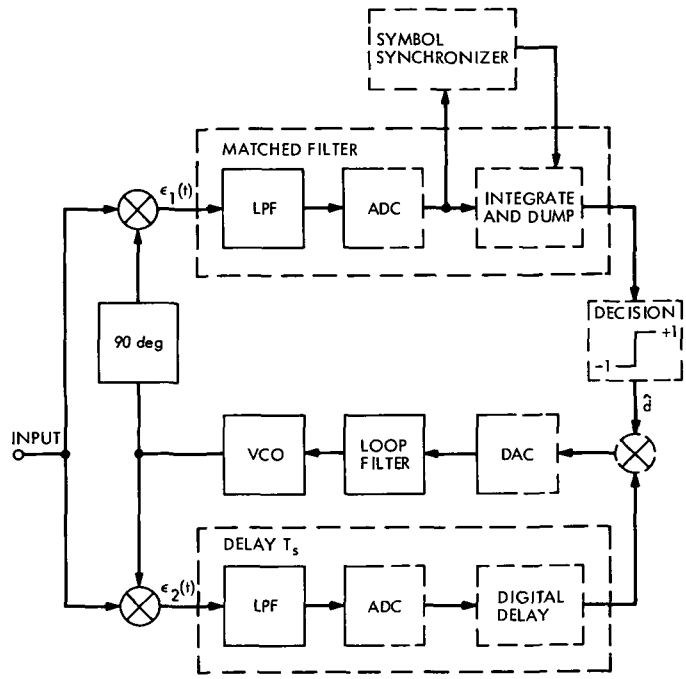

Fig. 4-1. Functional DAR implementation

B. Matched Filter and Symbol Synchronizer

When the DAL is locked, the stochastic input to the matched filter, $\epsilon_1(t)$, consists of the low-pass data symbols plus noise. Mathematically, the output of the matched filter and decision device configurations for any symbol d_i may be written as

$$\hat{d}_i = \operatorname{sgn}\left\{\int_{(i-1)T_s+\hat{\epsilon}}^{iT_s+\hat{\epsilon}} \epsilon_1(t)\,dt\right. \tag{4-1}$$

where $\hat{\epsilon}$ is the symbol synchronizer's estimate of the true symbol epoch ϵ. For convenience, and without any loss of generality, we shall assume that $\epsilon = 0$, and thus, $\hat{\epsilon}$ represents the relative timing offset due to the symbol synchronization process. The symbol synchronizer is designed to maintain $E\{\hat{\epsilon}\} = 0$.

The digital equivalent of the matched filter and decision operations is shown in Fig. 4-2. The low-pass data

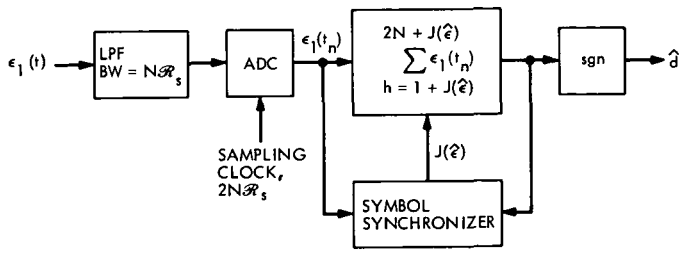

Fig. 4-2. Digital equivalent of matched filter and decision device

symbols plus noise are low-pass filtered (specification of bandwidth is discussed in Section D) and sampled at the Nyquist rate to obtain the samples $\epsilon_1(t_n)$ The ADC comprises a number of uniform levels $L = 2^K$, with K the number of bits, such that a nearly linear sample estimate of the input is made. (See Section E for a discussion of the number of ADC levels) The summer accumulates the sample values taken in one symbol period as directed by the timing information derived from the symbol synchronizer. The epoch of the sum is determined by an integer $J(\hat{\epsilon})$, with $E\{J(\hat{\epsilon})\} = 0$ The data estimate is equal to the sign of the accumulator at the end of the sum process. The accumulator is reset to zero after each symbol.

The performance of the digital matched filter can be made to operate within 0.1 to 0.2 dB of that for an ideal matched filter. The major contributors to a few tenths of a decibel in performance loss are the jitter on the symbol synchronizer timing estimate [i e , the variance of $J(\hat{\epsilon})$], and any signal-to-noise ratio (SNR) degradation due to the finite number of ADC levels Obviously, the closer we wish to come to the ideal case, the greater the complexity and cost of the mechanization. The problem of symbol synchronization is treated first

The method whereby the phase of the symbol synchronizer's local oscillator is discretely updated to match that of the incoming bit sequence makes use of the digital phase-locked loop principle. In particular, the phase detector topology of the data-transition tracking loop (Ref. 4-1) was found by experimental simulation to provide the best error signal, which is used either to bump the phase of the VCO clock $\pm\Delta T$ from its present position, or to leave the VCO at its phase position

The time increment $\Delta T = T_s/2N$ is the time between ADC samples. The probabilities with which these three possible timing updates occur depend on the magnitude of the timing error between the incoming symbol sequence and the clock, and the presence or absence of a transition in the data.

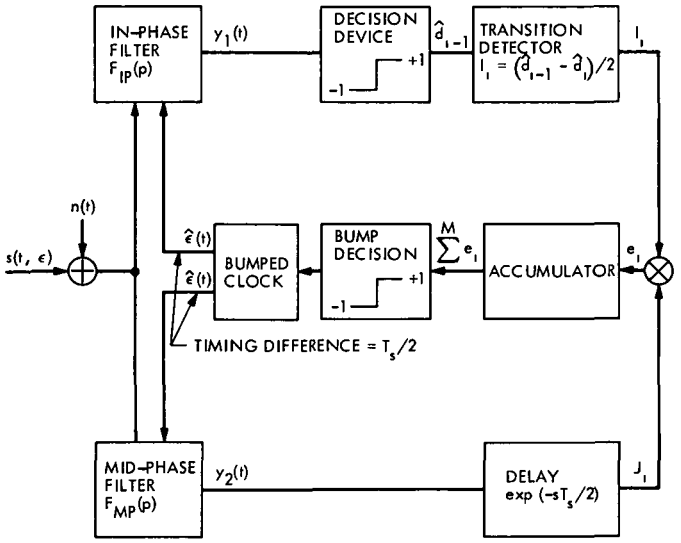

Fig. 4-3. Model for data-transition type of digital symbol synchronizer

Referring to Fig. 4-3, the input signal $s(t, \epsilon)$ is a random pulse train, representing the data symbols; i.e.,

$$s(t, \epsilon) = \sum_i d_i u[t - iT_s - \epsilon] \tag{4-2}$$

where $u(t)$ is the unit rectangular pulse defined by $u(t) = 1$ for $0 \leq t \leq T_s$, $u(t) = 0$ for all other t, T_s is the symbol period; the symbol amplitude d_i takes on the values $\pm\sqrt{P_D}$ with equal probability, and ϵ is the random epoch to be estimated, which, as previously stated, will be assumed to equal zero. The input additive noise process $n(t)$ is assumed to be white gaussian, with a two-sided spectral density $N_0/2$ W/Hz The in-phase and mid-phase filters, $F_{IP}(s)$ and $F_{MP}(s)$, respectively, are approximated by the accumulator shown in Fig. 4-2 and a second accumulator which begins its sum of samples in the mid-symbol position [i.e , when $n = N + 1 + J(\epsilon)$]. The decision devices simply extract the sign of the indicated accumulator

The transition detector then examines two adjacent decisions, $\hat{d}_{i-1}$, $\hat{d}_i$, and records an output I_i according to the following rule:

$$\left.\begin{aligned} &\text{If } \hat{d}_i = \hat{d}_{i-1}, \text{ then } I_i = 0 \\ &\text{If } \hat{d}_i = -1, \hat{d}_{i-1} = +1, \text{ then } I_i = +1 \\ &\text{If } \hat{d}_i = +1, \hat{d}_{i-1} = -1, \text{ then } I_i = -1 \end{aligned}\right\} \tag{4-3}$$

The output J_i of the mid-phase filter is also sampled at intervals of T_s and must be delayed by an amount $\tau = T_s/2$ before multiplication with the appropriate I_i. The error signal, $e_i = I_i \times J_i$, is in general accumulated

for an interval MT_s (i.e., M values of e_i are summed), and the clock is bumped (i.e., the phase estimate $\hat{\epsilon}_{+1}$ at interval MT_s is updated) according to the following rule:

$$\left.\begin{aligned} \hat{\epsilon}_{+1} &= \hat{\epsilon} - \Delta T, && \text{if } \sum^{M} e_i < 0 \\ \hat{\epsilon}_{+1} &= \hat{\epsilon} && \text{if } \sum^{M} e_i = 0 \\ \hat{\epsilon}_{+1} &= \hat{\epsilon} + \Delta T, && \text{if } \sum^{M} e_i > 0 \end{aligned}\right\} \qquad (4\text{-}4)$$

Now $\hat{\epsilon}_{+1}$ is a discrete random variable which ranges over the set of $2N$ possible values in the interval T_s. If we normalize $\hat{\epsilon}_{+1}$ by T_s, then the normalized random variable

$$J(\hat{\epsilon}_{+1}) \triangleq \frac{\hat{\epsilon}_{+1}}{T_s}$$

ranges over the set of integers $-N, -N+1, \cdots, -1, 1, \cdots, N-1, N$ (see Fig. 4-2) For simplicity of notation, we shall subsequently drop the $+1$ subscript on $\hat{\epsilon}_{+1}$, since it is only relative

In order to find the performance in the presence of noise, it is necessary to determine the steady-state probability density function $p(J)$ of $J(\hat{\epsilon})$ and then average the dependence of the symbol probability of error P_e^s on ϵ over this density. The method used is based on a random walk model, and details may be found in Ref 4-2.

The computation of $p(J)$ is rather tedious An example is given in Fig 4-4 for $M = 1$, $R_s = 3$, $N = 8$ We note that $E\{J\} = 0$, but that $J = 0$ is not allowed. As $R_s \to \infty$, the loop in effect becomes a bang-bang servo, switching between $J = -1$ and $J = +1$ as a result of the frequency drift between the transmitter and receiver timing oscillators and inherent circuit noise.

Figure 4-5 shows $p(J)$ for $M = 64$, with other conditions remaining the same. Comparing this result with that of Fig. 4-4 shows a definite decrease in the variance of J, thus indicating that increasing M acts to decrease the effective bandwidth of the symbol synchronizer loop. A plot of the variance of J, σ_J^2, as a function of R_s for the cases of $M = 1$ and $M = 64$ appears in Fig. 4-6.

The choice of the size of M in practice depends on (1) the desired acquisition and transient response time

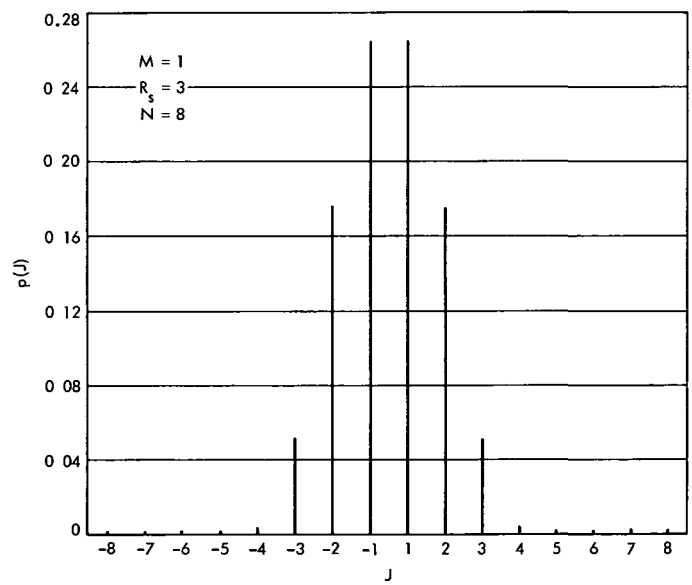

Fig. 4-4. Probability frequency function of $J(\epsilon)$; $M = 1$

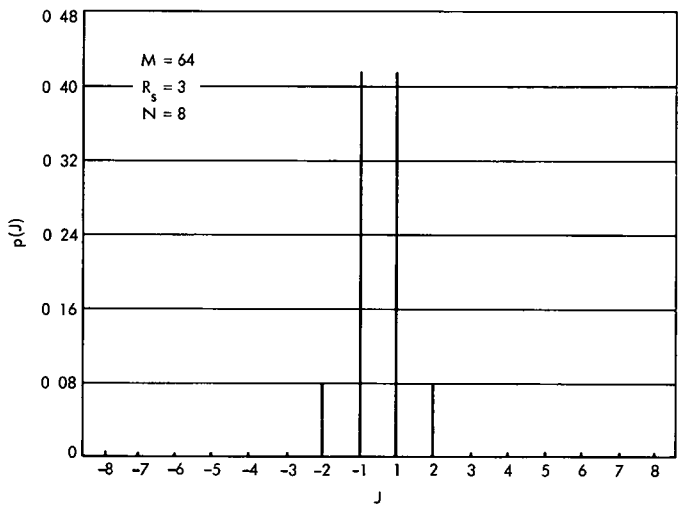

Fig. 4-5. Probability frequency function of $J(\epsilon)$; $M = 64$

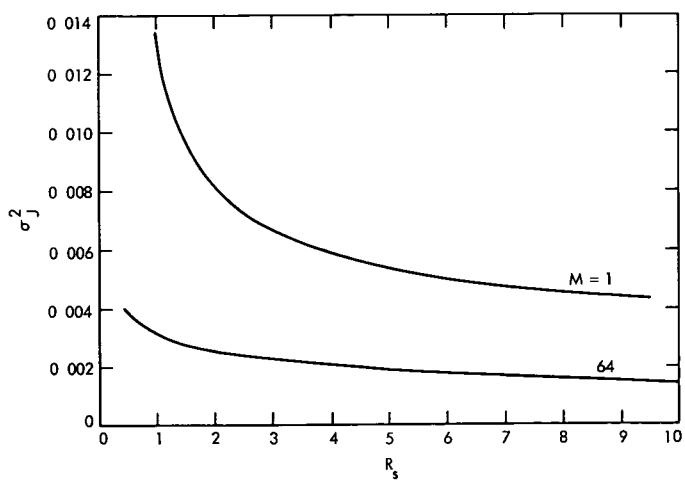

Fig. 4-6. Plots of σ_J^2 vs R_s

(usually measured in symbols) and (2) the allowable degradation in P_e^s performance relative to ideal, due to σ_j^2. Tests have shown that $2 < M < 16$ appears to give the best results from the acquisition standpoint, with $M = 4$ being used for most of the system testing described in this report. Actually, the choice of M for the acquisition case primarily affects the mean time to acquire symbol sync, which is generally very much less than the time required to obtain carrier synchronization. For $M = 4$, $N = 8$, and $R_s = 4$, the mean time to acquire symbol synchronization has been observed to be about 26 T_s.

Table 4-1 shows a list of measured losses resulting from σ_j^2, taken under the stated conditions and inferred from the measured P_e^s. Figure 4-7 presents the error probability performance as a function of signal-to-noise ratio R_s, in dB, for the extreme cases of $M = 1$ and $M = 64$, obtained by averaging $P_e^s(J)$ over the probability density function $p(J)$. The error probability performance of an ideal phase-shift-keyed (PSK) system is indicated in this figure for comparison. Also included is a set of experimental points taken for $M = 4$. Note that they fall closer to the $M = 64$ curve than the $M = 1$ curve, which indicates that rapid improvement with respect to $M = 1$ is possible for small M, and that, practically, M on the order of 4 is sufficient This conclusion has already been demonstrated by the experimental results shown in Table 4-1.

Table 4-1. Effective matched filter loss as a function of M[a]

M	Measured loss from ideal performance, dB
1	1 1
2	0 7
4	0 5
8	0 4
16	0 3
32	0 25
64	0 25

[a] $N = 8$, $R_s = 4$, oscillator offset = 0 01%

One final point should be made before concluding this section. There will normally exist some frequency difference between the transmitter and receiver master symbol timing oscillators, this has been referred to in Table 4-1 and Fig. 4-7 as oscillator offset, specified in percent of the symbol frequency An offset on the order of 0 01% has been found by measurement to cause a measured P_e^s degradation of between 0 1 and 0 2 dB. Some experiments were run to determine how large an offset could be tolerated while still maintaining DAL lock. For the case of $M = 4$, $N = 8$, and $R_s = 4$, the upper limit was found to be in the neighborhood of 0 5%.

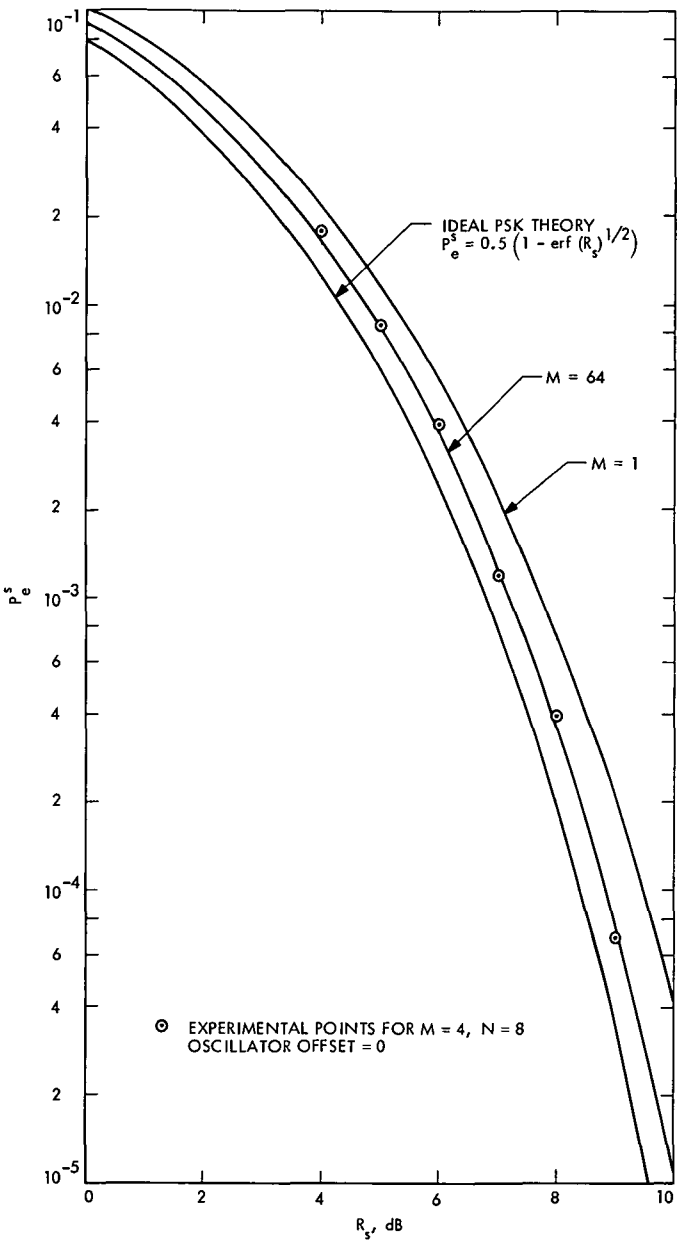

Fig. 4-7. Theoretical and measured symbol synchronizer performance

C. The Delay Function

The delay function, which delays the quadrature phase detector's output by T_s, is conceptually and mathematically a very simple device. Its hardware realization, however, is another matter. Considering for the moment how it might be mechanized without the use of digital components, at very high data rates it could perhaps be realized by using a suitable length of wideband coaxial line. As the data rate becomes progressively lower, the length of the coaxial line will become prohibitively great,

requiring that some form of lumped element delay line equivalent be substituted. Such mechanizations, being composed of R, L, and C components, become physically large and difficult to construct as the delays approach the tens of millisecond region. Another problem is that of adjusting the delay lines to give accurate delays, and maintaining them over large temperature variations. Other types of delay lines, such as magnetostrictive and surface wave, could also be used, however, these types are not very amenable to large delays (i.e., greater than 100 μs). Thus, the analog delay line is not a very practical solution to the delay function problem

Digital implementations of the delay function are straightforward and simple to implement. Consider the diagram shown in Fig 4-8. In a manner identical to that of the matched filter, the low-pass process is sampled by a K-bit ADC, where the sampling clock is identical to that used for the matched filter. The same clock is also used to operate a $2N$-stage shift register, thus providing the requisite delay. Large or small delays present no problem, since they are dependent only on the sampling clock frequency.

An alternate mechanization is shown in Fig 4-9. The $2N$-stage shift register is replaced by an accumulator identical to that used in the matched filter, and with the same timing. The output $\epsilon_3(t_n - T_s)$ now occurs at a rate equal to the symbol rate $\mathcal{R}_s$. Thus, this implementation acts to provide the requisite delay, and also to low-pass filter the loop phase error-related term The filtering process has no real effect on the tracking loop performance, however, as $W_L << \mathcal{R}_s$ for proper loop operation (Section III-D), meaning that the loop filter $F(p)$ determines the loop SNR. Measurements have been unable to discern any performance difference between this mechanization and a true delay line A definite advanatge of the accumulator mechanization over the shift register is that the output signal rate is less by a factor $1/2N$, so that the subsequent circuit blocks, especially the DAC, need work only at the symbol rate We will also see in Section V-B that this is convenient to the implementation of the lock detector

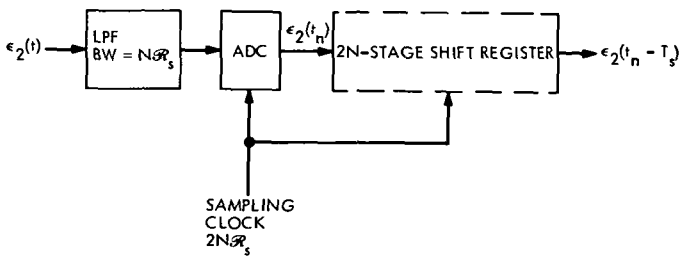

Fig. 4-8. Digital equivalent of delay line

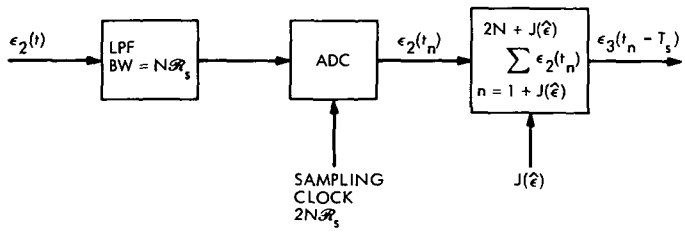

Fig. 4-9. Alternate digital delay line

D. Choice of Sampling Bandwidth

Having established that the matched filter and delay function should be digitally implemented, we next address the problem of sampling bandwidth. As indicated in Figs. 4-2 and 4-8, the sampling bandwidth was arbitrarily chosen to be $N\mathcal{R}_s$ Hence, the question of importance is. How small can we choose N without paying an appreciable penalty in SNR performance? It is desirable to make N as small as practical for the highest symbol rate used, as this has a direct effect on minimizing the speed at which the ADC has to function, and on the accumulator sizes, which in turn minimizes implementation cost.

In terms of system performance parameters, we aim to compute the degradation of SNR with decreasing N and determine from this functional relationship whether or not a distinct lower bound exists.

The performance of the matched filter configuration shown in Fig 4-2 is the subject of the analysis, where it is assumed that the symbol synchronizer acts to provide a perfect timing reference. In terms of the input symbol stream, we reason that with long strings of ones (or minus ones) in the input data, the probability of error of d will not be degraded by the finite bandwidth effect of the low-pass filter (LPF) Clearly, then, the worst case insofar as degradation is concerned will occur for alternating data (i.e , 1, −1, 1, −1, 1, −1,). We shall assume this worst case in the analysis which follows by modeling $d(t)$ as a square wave of period $2T_s$.

Consider the system model of Fig 4-10 and its mathematical equivalent, Fig. 4-11. The signal $s_i(t)$ is a periodic impulse train with Fourier transform

$$S_i(f) \triangleq \mathcal{F}\{s_i(t)\} = \sum_{k=-\infty}^{\infty} \frac{\mathcal{R}_s}{2} \delta\left(f - k\frac{\mathcal{R}_s}{2}\right) \quad (4\text{-}5)$$

This impulse train is passed through a filter whose impulse response is

$$h_0(t) = \begin{cases} \sqrt{S}, & |t| \leq \dfrac{T_s}{2} \\ -\sqrt{S}, & \dfrac{T_s}{2} \leq |t| \leq T_s \\ 0, & T_s \leq |t| \end{cases} \qquad (4\text{-}6)$$

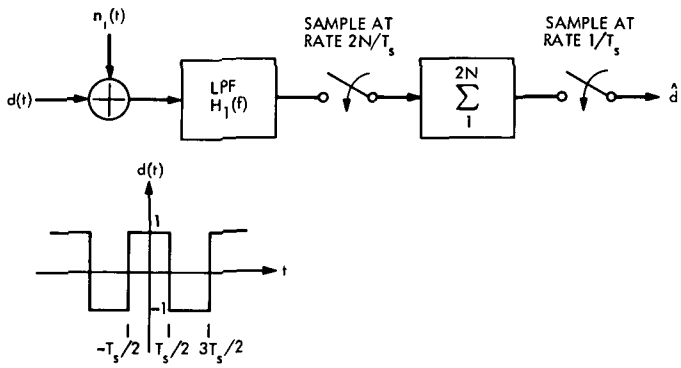

Fig. 4-10. Sampled data system model

Taking the Fourier transform of $h_0(t)$, we have

$$H_0(f) = \frac{2\sqrt{S}}{f\pi} \sin\left(\frac{\pi f}{\mathcal{R}_s}\right)\left[1 - \cos\left(\frac{\pi f}{\mathcal{R}_s}\right)\right] \qquad (4\text{-}7)$$

The signal component of the output of $H_0(f)$ is a periodic square wave, which meets the requirements on $d(t)$.

Passing $s_1(t)$ through the filter $H_2(f)$ provides an impulse sampling waveform of period $1/(2N\mathcal{R}_s)$. The impulse response $h_2(t)$ is characterized by

$$h_2(t) = \sum_{n=-2N+1}^{2N} \delta\left(t - \frac{2n-1}{4N\mathcal{R}_s}\right) \qquad (4\text{-}8)$$

with Fourier transform

$$H_2(f) = 2\sum_{n=1}^{2N} \cos\left(\frac{2\pi f(2n-1)}{4N\mathcal{R}_s}\right) \qquad (4\text{-}9)$$

The filter whose bandwidth is to be determined is

$$H_1(f) = |H_1(f)| \exp[j\Theta(f)] \qquad (4\text{-}10)$$

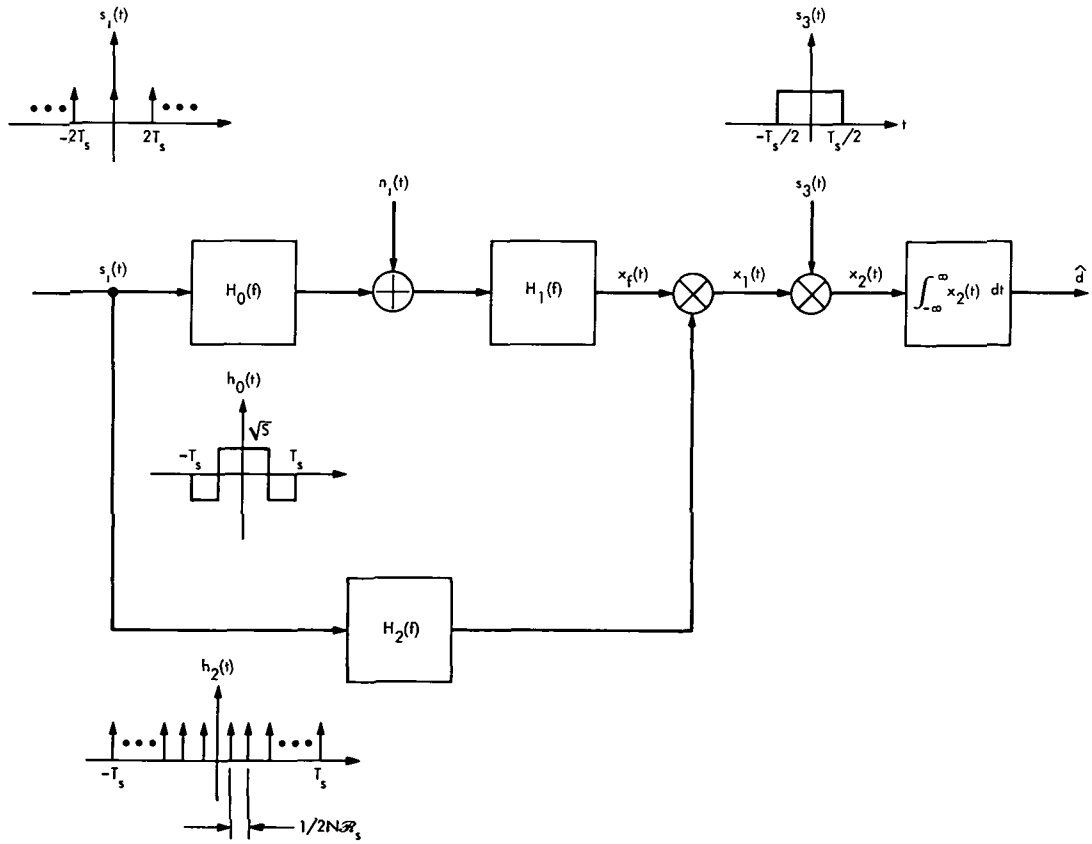

Fig. 4-11. Sampled data system mathematical equivalent

and the output of this filter is denoted by $x_f(t) = s_f(t) + n_f(t)$. The sampled filtered square wave appears in Fig. 4-11 as $s_1(t)$ and has a Fourier spectrum

$$S_1(f) = [S_s(f)\,H_0(f)\,H_1(f)] * [S_s(f)\,H_2(f)] \qquad (4\text{-}11)$$

where the asterisk denotes convolution. A 1-symbol-length segment of this signal, denoted by $s_2(t)$, can be obtained by multiplying $s_1(t)$ by

$$s_3(t) = \begin{cases} 1, & |t| \leq \dfrac{T}{2} \\ 0, & |t| > \dfrac{T}{2} \end{cases} \qquad (4\text{-}12)$$

Hence, in the frequency domain

$$S_2(f) \triangleq \mathcal{F}\{s_2(t)\} = S_1(f) * S_3(f) \qquad (4\text{-}13)$$

Since $x_2(t) = s_2(t) + n_2(t)$ is an impulse sampled version of the filtered input square wave plus noise over a single symbol interval, integrating $x_2(t)$ between $-\infty$ and ∞ (or, equivalently, between $-T_s/2$ and $T_s/2$) is equivalent to summing $2N$ samples of $x_f(t)$. Hence, the data estimate $\hat{d}$ may be expressed as

$$\hat{d} \triangleq \frac{1}{2N}\sum_{n=-N+1}^{N} x_f\left(\frac{2n-1}{4N\mathcal{R}_s}\right) = \frac{1}{2N}\int_{-\infty}^{\infty} x_2(t)\,dt \qquad (4\text{-}14)$$

The next step in the procedure is to obtain the signal-to-noise ratio of $\hat{d}$. The power in the signal component of $\hat{d}$ can be obtained from Eq. (4-14) as

$$P_s = [E\{\hat{d}\}]^2 = \left[\frac{S_2(0)}{2N}\right]^2 \qquad (4\text{-}15)$$

The noise power $\sigma_{n_s}^2$ in the data estimate $\hat{d}$ can be expressed in terms of the variance σ_n^2 of the unsampled zero-mean noise process at the output of $H_1(f)$ and the noise correlation between samples of this process.

Defining the normalized filter

$$H_1'(f) = H_1(N\mathcal{R}_s f) \qquad (4\text{-}17)$$

gives

$$\sigma_{n_s}^2 = \frac{N_0\mathcal{R}_s}{2}\left[1 + \sum_{k=1}^{2N-1}\left(1 - \frac{k}{2N}\right)\rho_k\right] = \frac{N_0}{2T_s}\sigma_{n_s}'^2 \qquad (4\text{-}18)$$

where

$$\rho_k \triangleq 2\int_0^{\infty} |H_1'(f)|^2 \cos \pi f k\,df \qquad (4\text{-}19)$$

The signal-to-noise ratio of $\hat{d}$ is now defined as

$$\boxed{\text{SNR} = \frac{P_s}{\sigma_{n_s}^2}} \qquad (4\text{-}20)$$

where P_s and $\sigma_{n_s}^2$ are found from (4-16) and (4-18), respectively Making these substitutions, we find that

$$\text{SNR} = \left(\frac{2ST_s}{N_0}\right)\left[\frac{P_s'}{\sigma_{n_s}'^2}\right] \qquad (4\text{-}21)$$

The factor $P_s'/\sigma_{n_s}'^2$ represents the SNR degradation with N

Two sample cases have been calculated and plotted to show the effects of N on performance. The filter forms taken are a cascade of K identical poles, for which

$$|H_1'(f)|^2 = \frac{1}{[1 + (f/f_c)^2]^K} \qquad (4\text{-}22)$$

After some straightforward algebraic manipulation,

$$\sqrt{P_s} = 2\sqrt{S}\sum_{k=1}^{2N}\left(\frac{2}{\pi}\right)^2 \frac{1}{(2k-1)^2}\left(\frac{2k-1}{4N}\right)\left\{1 - \left(\frac{2k-1}{4N}\right)\sum_{l_1=1}^{\infty}\frac{2\,[(2k-1)/4N]}{l_1^2 - [(2k-1)/4N]^2}\right\}$$

$$\times \sum_{l_2=0}^{\infty}\left\{\frac{1}{l_2 + [(2k-1)/4N]}\right\}\left|H_1\left[2N\mathcal{R}_s\left(l_2 + \frac{2k-1}{4N}\right)\right]\right| \cos\left\{\Theta\left[2N\mathcal{R}_s\left(l_2 + \frac{2k-1}{4N}\right)\right]\right\}$$

$$= \sqrt{S}\sqrt{P_s'} \qquad (4\text{-}16)$$

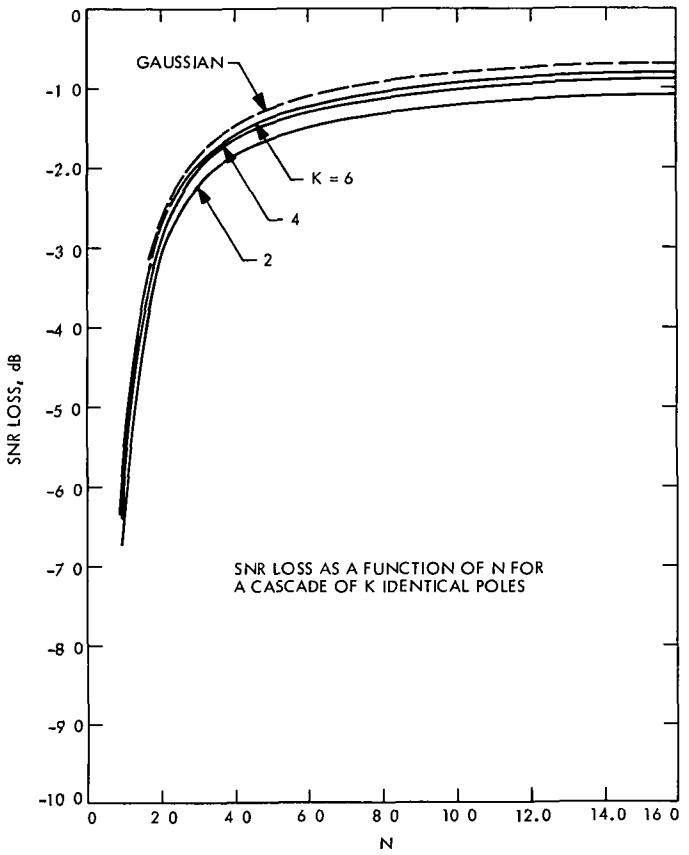

Fig. 4-12. SNR loss as a function of *N* for a cascade of *K* identical poles

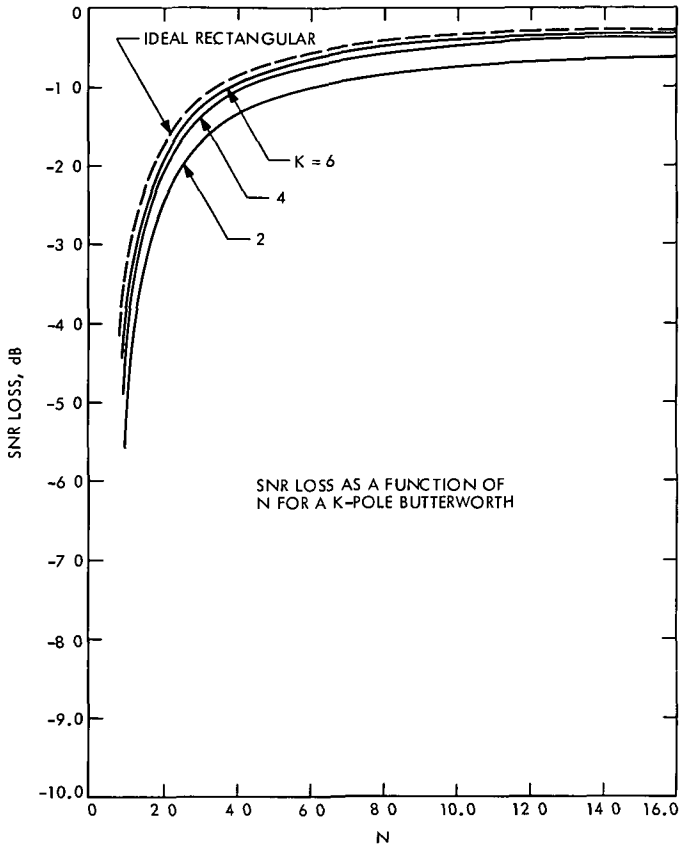

Fig. 4-13. SNR loss as a function of *N* for a *K*-pole Butterworth

and a K-pole Butterworth (K even), where

$$|H_1'(f)|^2 = \frac{1}{1 + (f/f_c)^{K+2}} \tag{4-23}$$

Details pertaining to these filter forms, as well as the values of ρ_k, may be found in Ref. 4-3

Plots of the SNR degradation for the two filter forms appear in Figs 4-12 and 4-13 Clearly, the Butterworth filter is the best choice. Experimental results for $N = 8$ and $K = 6$ Butterworth filters indicate a loss of about 0.3 dB. Obviously, from the curves, N should be taken to be 4 or greater in order to keep the losses at a reasonable level. The implementation described in Section F takes $N = 8$.

E. ADC and DAC Sizes

The process of analog-to-digital conversion is referred to as quantizing of the input signal plus noise ensemble. We wish here to focus our attention on the number of bits required by the ADCs and the DAC.

Two questions arise relative to the quantizing operation.

(1) How many bits K should be used in order that the SNR of the quantized samples be insignificantly degraded relative to ideal samples? This loss will be referred to as the *quantization loss*

(2) What is the effect of the discrete nature of the matched filter accumulator on the probability of symbol error, as a function of K?

The answer to these questions is important, as it essentially determines the word size of the accumulator operations which follow the ADC Quite obviously, the fewer bits per sample, the less the number of total bits of storage required. However, if one lets $K = 0$ (quantizing the sign of the sample only), it can be shown that for low SNRs (and gaussian noise), the SNR is degraded by about 2 dB, an obviously excessive loss We thus must determine a reasonable value for K.

We denote the output of the LPF preceding the ADC as $X(t)$ Since the process of quantizing can be viewed as

a zero memory transformation of the continuous variable $X(t)$, the processes of sampling and quantizing can be interchanged, as shown in Fig. 4-14 (i.e., we let $N\mathcal{R}_s = W$). Further, for the purpose of analysis, since the sampling is taken to be at the Nyquist rate so that the samples represent the continuous process contained in the band $(0, W)$, it is expedient to work with the continuous quantized process. Thus, we consider the model shown in Fig. 4-15.

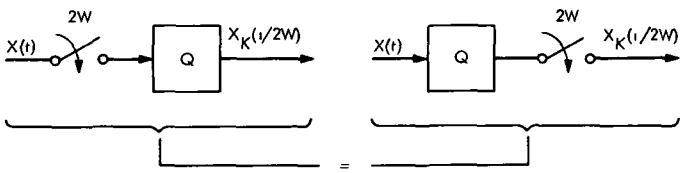

Fig. 4-14. Equivalent quantizers

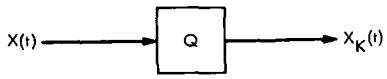

Fig. 4-15. Quantizer analysis model

Now the quantizer Q is represented by the transformation characteristic shown in Fig. 4-16, where the staircase is ideally assumed to have an infinite number of steps. The finite bit quantizer then becomes approximated by the ideal situation when the peak signal plus noise values at the input are constrained (scaled) to be less than the 2^Kth level.

By expanding the derivative of the quantizer characteristic in a Fourier series, it can be easily shown that the desired input/output relationship for Fig. 4-15 is

$$X_K(t) = \frac{A}{2}\left\{\frac{1}{\pi}\sum_{k=0}^{\infty}\frac{\epsilon_k}{k}\sin\left[\frac{2\pi k}{V}X(t)\right] + \operatorname{sgn}[X(t)]\right\} \tag{4-24}$$

where ϵ_k is the Neumann factor.

Now, $X(t)$ can be written approximately as

$$X(t) = \sqrt{S}\,d(t) + n_f(t) \tag{4-25}$$

where the variance of $n_f(t)$ denoted by var $\{n_f(t)\}$ equals N_0W. We have neglected the fact that the LPF preceding

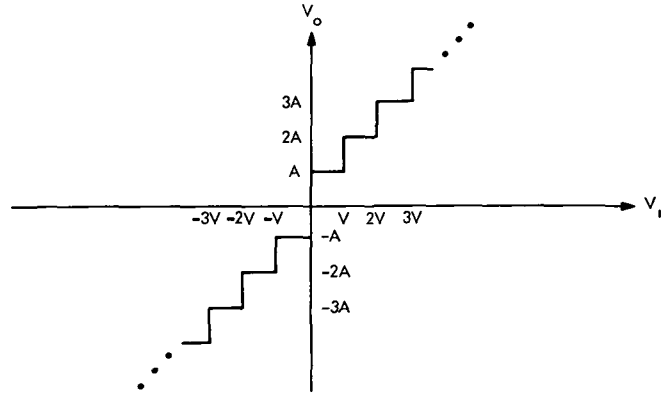

Fig. 4-16. Quantizer characteristic

the ADC filters out some of the high-frequency components of $d(t)$, thus causing a finite transition time between symbols. To do otherwise would make the analytical solution highly intractable.

Our goal is to determine the minimum number of required ADC bits (K) that produce a quantization loss on the order of 0.3 dB Now, if only one ADC level is used (i.e., $K = 0$), such that $|X(t)| < V$ for all t, we find that

$$X_K(t) \triangleq X_0(t) = A\operatorname{sgn}[X(t)]$$

Under this condition, as already mentioned, the SNR loss is $2/\pi$ or 2 dB (Ref 4-4). Thus, we must have a sufficient number of levels to accommodate $|X(t)|_{\max} >> V$. We may then write

$$X_K(t) \cong \frac{A}{2\pi}\sum_{k=0}^{\infty}\frac{\epsilon_k}{k}\sin\left[\frac{2\pi k}{V}X(t)\right]$$

$$= \frac{A}{2\pi}\sum_{k=0}^{\infty}\frac{\epsilon_k}{k}\sin\left[\frac{2\pi k\sqrt{S}}{V}d(t) + \frac{2\pi k}{V}n_f(t)\right] \tag{4-26}$$

Expanding the sin [ ] term, noting that $d(t)$ can take on only $+1$ or -1 values, we obtain

$$X_K(t) = \frac{A}{2\pi}\sum_{k=0}^{\infty}\frac{\epsilon_k}{k}\left[d(t)\sin\left(\frac{2\pi k\sqrt{S}}{V}\right)\cos\left(\frac{2\pi k}{V}n_f(t)\right)\right.$$

$$\left. + \cos\left(\frac{2\pi k\sqrt{S}}{V}\right)\sin\left(\frac{2\pi k}{V}n_f(t)\right)\right] \tag{4-27}$$

Separating the $k = 0$ term from the expression, we have

$$X_K(t) = \frac{A}{V}\left[\sqrt{S}\,d(t) + n_f(t)\right] + \frac{A}{\pi}\sum_{k=1}^{\infty}\frac{1}{k}\left[d(t)\sin\left(\frac{2\pi k\sqrt{S}}{V}\right)\cos\left(\frac{2\pi k}{V}n_f(t)\right) + \cos\left(\frac{2\pi k\sqrt{S}}{V}\right)\sin\left(\frac{2\pi k}{V}n_f(t)\right)\right] \quad (4\text{-}28)$$

where we recognize the first term as being nothing more than the input function scaled by A/V, leaving the summation to represent all of the additional terms which degrade the SNR of the samples. We must therefore obtain the variance of the summation. Evaluating the necessary integral (n_f assumed to be gaussian), we obtain

$$\text{var}\left\{\frac{A}{\pi}\sum_{k=1}^{\infty}\frac{1}{k}[\,]\right\} = \frac{A^2}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2} \times\left\{1 - \cos\left(\frac{4\pi k\sqrt{S}}{V}\right)\exp\left[-\left(\frac{4\pi k}{V}\right)^2\frac{N_0W}{2}\right]\right\} \quad (4\text{-}29)$$

Before proceeding with the final step, it is necessary to relate the peak value of $|X(t)|$ to the maximum level taken for the ADC. For a K-bit ADC, the maximum quantization level is 2^KV. Now, if ρ_{ADC} is defined as the input SNR to the ADC, then assuming that $\rho_{ADC} < 1$, the peak value of $|X(t)|$ will be on the order of the peak value of the noise, which we will take to be

$$4\sigma_{n_f} = 4\sqrt{N_0W} = 2^KV \quad (4\text{-}30)$$

Thus, we obtain

$$V = \frac{4\sqrt{N_0W}}{2^K} \quad (4\text{-}31)$$

which, when substituted into Eq (4-29), gives

$$\text{var}\left\{\frac{A}{\pi}\sum_{k=1}^{\infty}\frac{1}{k}[\,]\right\} = \frac{A^2}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2} \times\left\{1 - \cos\left(2^K\pi k\sqrt{\rho_{ADC}}\right)\exp\left[-\frac{1}{2}(2^K\pi k)^2\right]\right\} \quad (4\text{-}32)$$

We must remember that Eq. (4-32) is valid for, say, $K \geqq 2$. Now, a quick examination of the exponential term for $K = 2$ shows that it can be neglected even for $k = 1$, thus,

$$\text{var}\left\{\frac{A}{\pi}\sum_{k=1}^{\infty}\frac{1}{k}[\,]\right\} \cong \frac{A^2}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{A^2}{12} \quad (4\text{-}33)$$

a well known result for the variance of the quantization error

Evaluating the signal mean and noise variance from Eq. (4-28), the final result for the sample SNR, defined by the symbol ρ_K, is

$$\boxed{\rho_K = \frac{\rho_{ADC}}{1 + 2^{-2K}} \triangleq \rho_{ADC}\,L_{ADC}} \quad (4\text{-}34)$$

Table 4-2 gives several values of L_{ADC}.

Table 4-2. ADC SNR losses

K	2	3	4	5
L_{ADC}, dB	−0 26	−0.06	−0.02	−0 01

It can be seen from the table that if $K > 3$, the SNR loss is negligible.

We now turn our attention to the discrete nature of the accumulator distribution and its effect on symbol probability of error. Unfortunately, no simple analytical solution for this problem has been readily obtained. Consider the following dilemma For a K-bit ADC, the discrete probability density function (p.d.f.) of the $S + N$ samples consists of $L = 2^{K+1} - 1$ discrete values. If $M + 1$ samples are accumulated, then the resulting discrete p.d.f. of the accumulator has $2^ML - (2^M - 1)$ discrete values. The actual numerical magnitude of the accumulator p.d.f. values depends on the M-fold convolution of the input p.d.f , or alternatively, on raising the generating function to the Mth power. In either case, the procedure is very complex, as it involves a large number of combinations for realistic K and M. For example, if $K = 5$ and $M = 15$, which is typical for an actual mechanization, the number of possible accumulator values is slightly in excess of 2×10^6 Actually, it is not necessary to calculate this many

values, as it can be shown that only a few thousand are really significant to the desired outcome, however, even this is an arduous job As a result, the answer to the problem has been obtained empirically.

Table 4-3 presents some representative R_s losses inferred from measured P_e^s for specified values of R_s, when $M = 15$ We note that the largest losses occur for the higher values of R_s. This is not at all surprising, since the probability of error is determined by the extremes of the distribution of the noise, which is significantly modified by the fact that beyond some $k\sigma_n$ (say, $k=4$), all of the values of the noise are mapped into the largest quantized value. The only way to reduce the loss for small values of K is to increase M (and therefore increase the LPF bandwidth and the sampling rate). This has been experimentally proven for the digital command detector (Ref. 4-5), where $K = 3$ and $M = 127$, and the loss is on the order of -0.1 dB

Table 4-3. Measured R_s losses inferred from P_e^s

K \ R_s, dB	4.0	5.0	6.0	7.0	8.0	9.0
4	−0.6	−0.6	−0.6	−0.7	−0.8	−0.9
5	< −0.1		−0.1	−0.1	−0.3	−0.4
6	< −0.1					−0.1
7	< −0.1					

Clearly, if $K \geqq 7$, while $M = 15$, no significant degradation will result. In the implementation described in Section F, K was taken as 7.

The function of the DAC is to reconstruct an analog form of the loop error signal No studies were actually conducted to size the DAC Since $K = 7$ was the final choice for the ADC size, and $2N = 16$, the DAC size was simply chosen to be 11 bits plus sign.

F. Complete Mechanization

The digital portions of the data-aided loop have been implemented according to the block diagram shown in Fig. 4-17. As can be seen, two ADCs were used, one in each data stream, rather than a single ADC multiplexed between the two low-pass phase detector outputs. This allows a lower master clock rate for a given symbol rate, thus lowering the duty cycle of the ADCs. The ADCs selected for use in the data-aided loop are 8-bit (7-bit magnitude plus sign) devices with a throughput rate of 1.25 MHz, providing for a symbol rate as high as 78 kilosymbols per second (ksps)

The master clock frequency is 160 times the symbol rate This frequency is passed through a series of dividers and counters, yielding two pulse trains. The pulse trains are present in different phases of a frequency equal to 1/10 the master clock rate, and hence, cycle 16 times per symbol period.

One of the phases is passed through a 16-bit counter, and the occurrence of a 0 or 8 count is detected By logically ANDing these detected counts with the generated pulse trains, different phases and subfrequencies of the master clock are generated. These are necessary to implement the digital integrate and dump circuits. Three integrate and dump, or accumulator, circuits are used. They are the mid-phase and in-phase for the matched filter and the one-symbol delay. Each of the accumulator circuits consists of a 12-bit binary adder and a buffer The accumulators sum 16 ADC samples per symbol period and are cleared by the appropriate clock pulses The mid-phase accumulator is cleared by the 8-count detected bit, and both the in-phase and 1-bit delay accumulators are cleared by the 0-count detected bit.

The sign bit of the in-phase accumulator is the data estimate At each occurrence of the 0-count bit, the new data estimate is compared with the previous value. If a change is detected, the mid-phase accumulator is multiplied by the data estimate, and the result is added to the clock bump accumulator After every fourth change is detected, the clock is incremented or decremented 1/16 of the symbol period, depending upon the sign of the clock bump accumulator, which is subsequently cleared Thus, symbol synchronization is maintained.

The output of the one-symbol delay accumulator is multiplied (complemented) by the data estimate, with the result being input to a 12-bit (11 bits magnitude plus sign) DAC. The output of this DAC is the loop error signal which is input to the loop filter.

The digital logic of the data-aided loop was built entirely of 14- and 16-pin dip integrated circuits (Series 54, 74) which were assembled on three (64-socket) cards, as shown in Fig. 4-18. The power supplies for the logic and converters are external.

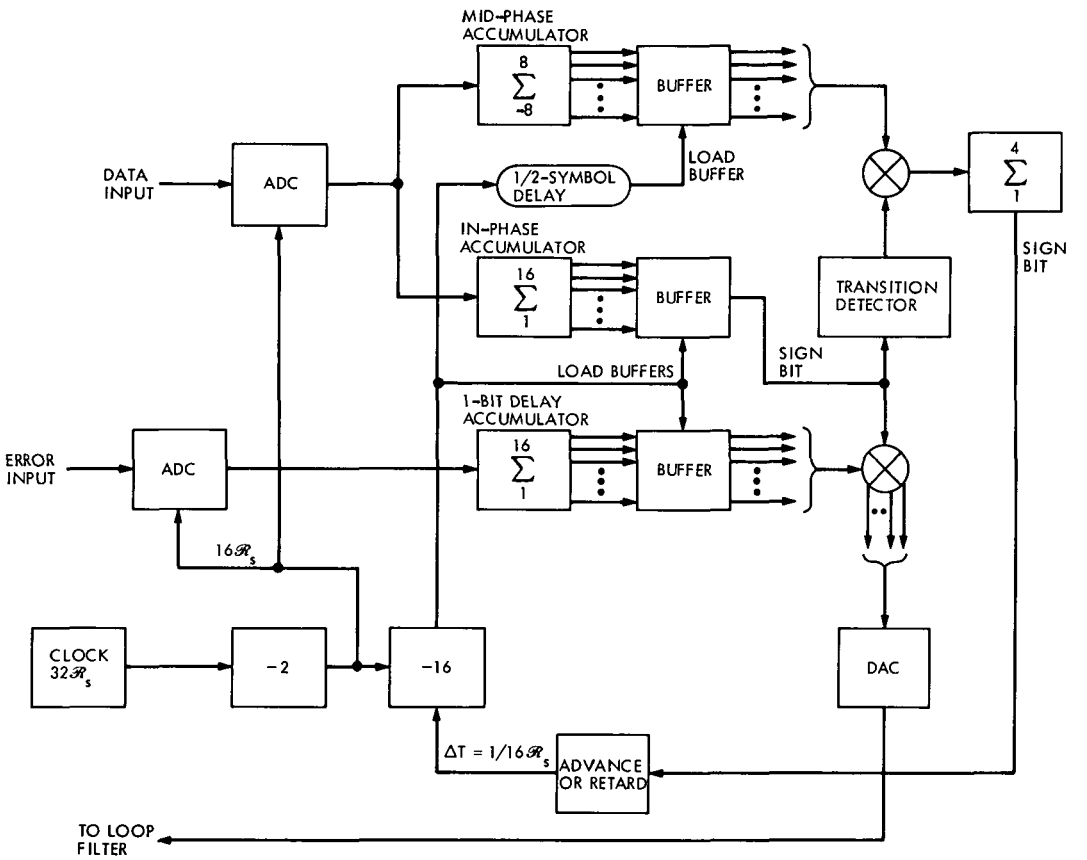

Fig. 4-17. Complete DAR logic mechanization

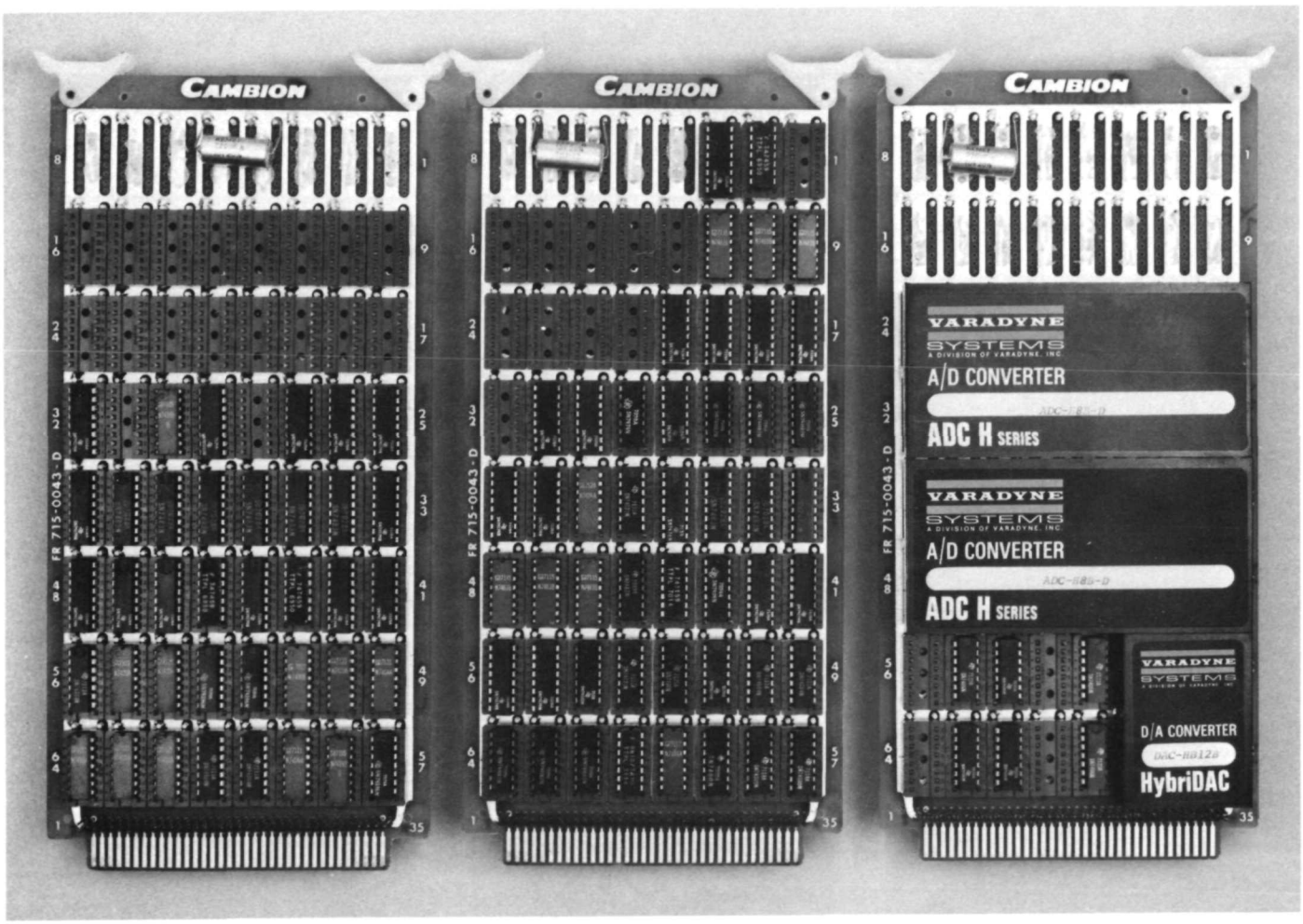

Fig. 4-18. DAR logic cards

G. Performance Test Results

In the preceding sections, we have discussed details of the DAR implementation, and presented experimental results as they pertained to the immediate subject under study. We present here the overall test results as obtained from measuring the symbol probability of error The experimental data are shown in Fig. 4-19. The tests were run at a symbol rate $\mathcal{R}_s = 1$ ksps using a pseudo-random, bi-orthogonal data stream The DAL tracking bandwidth was chosen to be 20 Hz.

The experimental data points were obtained for three sets of conditions. The first set involves "hardwired" RF and symbol sync references, and the corresponding data points (indicated by squares) are therefore a measure of the LPF and ADC losses As can be seen, the average loss attributable to this condition is about -0.3 dB.

The second set of experimental points (indicated by hexagons) was obtained for hardwired RF references, but with the symbol synchronizer operating ($M = 4$) and the symbol rate clocks offset by 0 01%. The data points average about 0.8 dB offset from the ideal PSK performance curve, indicating that the performance degradation introduced by the noisy symbol clock is -0.5 dB. This, of course, could be decreased by 0 2 to 0.3 dB by increasing M (Section B).

The third set of data points (indicated by circles) is taken for the complete operating DAR, for which the RF references are also noisy. As can be seen, the additional loss is between -0 1 and -0 2 dB, as the measured rms phase error ranged between 4 4 and 8.5 deg.

The worst-case signal-to-noise ratio in the loop tracking bandwidth occurs for $R_s = 4$, and is 20 8 dB. All of the data points were obtained for zero detuning, a loop gain $K = 50{,}000$, and a loop damping factor of $\sqrt{2}/2$. No limiter was used, as per Fig. 3-14. Additional operating conditions are shown in Fig 4-19.

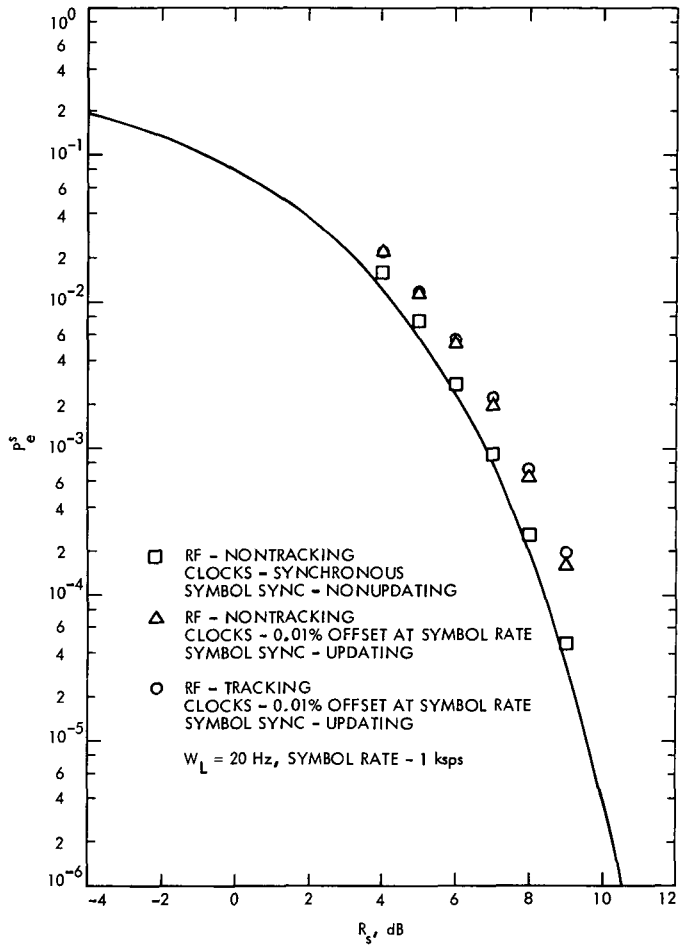

Fig. 4-19. DAR performance with bi-orthogonal data

References

4-1. Simon, M. K., "An Analysis of the Steady-State Phase Noise Performance of a Digital Data-Transition Tracking Loop," in *Supporting Research and Advanced Development*, Space Programs Summary 37-55, Vol. III, Feb. 28, 1969.

4-2 Simon, M. K., "The Steady-State Performance of a Data-Transition Type of First-Order Digital Phase-Locked Loop," in *Supporting Research and Advanced Development*, Space Programs Summary 37-66, Vol III, Dec. 31, 1970.

4-3 Simon, M. K., "On the Selection of a Sampling Filter Bandwidth for a Digital Data Detector," *IEEE Trans Communications*, Vol. COM-20, No. 3, June 1972, pp. 438–441.

4-4. Deutsch, R., *Nonlinear Transformations of Random Processes*, Prentice-Hall, Inc, Englewood Cliffs, N.J., 1962.

4-5. Tegnelia, C., and Meahl, M., *Final Report, Digital Command Detector Development*, JPL Report 900-547, Jet Propulsion Laboratory, Pasadena, Calif, April 28, 1972.

V. Attendant Functions

Two important functions usually associated with coherent receivers are (1) automatic gain control (AGC) and (2) in-lock/out-of-lock detection In the following, we discuss the implementation and performance of these functions with respect to the data-aided receiver

A. Automatic Gain Control

In discrete carrier receivers, the AGC voltage is derived from an in-phase detector, which in effect gives a voltage proportional to the carrier level (see Refs 5-1 and 5-2). The phase reference for this detector is shifted 90 deg relative to that of the carrier-tracking loop. A simplified block diagram of an AGC control loop is shown in Fig. 5-1, where we have assumed a perfect carrier reference.

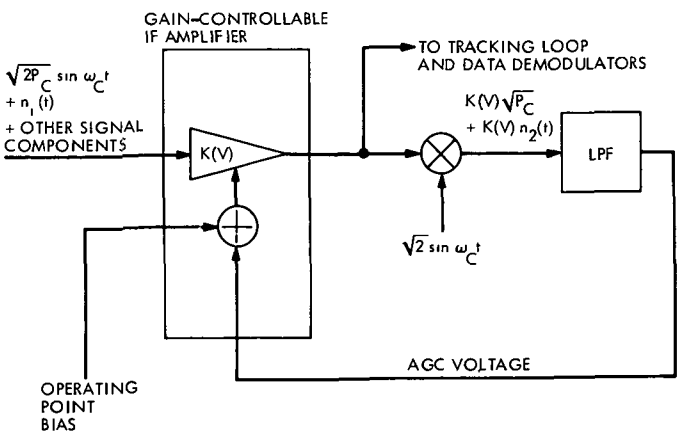

Fig. 5-1. Simple AGC loop for discrete carrier receivers

The purpose of the AGC loop is to maintain a certain degree of linearity through portions of the receiver, especially those that handle the data components of the received signal Since the received signal level may vary 100 dB or more throughout the course of using the receiver, the AGC loop prevents the high-gain RF and IF amplifiers from saturating (clipping) when the received signals are very strong.

We see from Fig. 5-1 that for large R_s, the open-loop AGC voltage derived from a discrete carrier is (in the absence of a noisy phase reference) a nearly linear function of the carrier voltage For the suppressed carrier signal, however, there is no carrier voltage. The output of the in-phase detector for the data-aided receiver (DAR) is given in Eq (3-8) and is proportional to $\sqrt{P_D}\,d(t)\cos\phi(t) + N_1[t, \phi(t)]$ If $d(t)$ is random (as assumed), with an equal number of $+1$ and -1 symbols, then the average value of the in-phase detector output is zero. Thus, no AGC voltage is generated at this point.

To circumvent this problem, we deal with the absolute value of the phase detector output. Because of the digital mechanization discussed in Section IV-B, it is more convenient to deal with the absolute value of the data symbol accumulator (matched filter) output, which is obtained by ignoring the sign bit. Thus, the basic AGC voltage signal V will be defined as the absolute value of the integrated data symbols, whose mean is defined by

$$\mu_V = E\{V\} = E\left|\int_{\hat{\epsilon}}^{T_s+\hat{\epsilon}} \{\sqrt{P_D}\,d(t)\cos\phi(t) + N_1[t,\phi(t)]\}\,dt\right| \tag{5-1}$$

For simplicity, we deal here with continuous rather than discrete variables and also assume the first transmitted symbol.

For the case of large loop signal-to-noise ratio (SNR), we can assume perfect carrier and symbol synchronization, i e , $\epsilon = 0$, $\cos\phi(t) = 1$. Thus, we have

$$\mu_V \cong E\left|\int_0^{T_s} \{\sqrt{P_D}\,d(t) + n_2(t)\}\,dt\right| \tag{5-2}$$

where $n_2(t)$ is low-pass white gaussian noise with single-sided spectral density, N_0 W/Hz. Now, the integral is a gaussian random variable, with mean $\sqrt{P_D}\,T_s d(t)$, where $d(t)$ is either $+1$ or -1, and variance $N_0T_s/2$. Reference 5-3 shows that

$$\mu_V = \sqrt{\frac{N_0T_s}{\pi}}\exp(-R_s) + \sqrt{P_D}\,T_s\,\mathrm{erf}(\sqrt{R_s}) \tag{5-3}$$

where erf ( ) is given as per Eq. (3-13).

We see that μ_V is a function of both the signal and noise voltages, and that its minimum value, i e , $\sqrt{N_0T_s/\pi}$,

occurs for $R_s = 0$ (no signal, noise only). Thus, it is more convenient to examine the normalized mean, viz.,

$$\mu_V^N \triangleq \frac{\mu_V}{\sqrt{N_0 T_s/\pi}} = \exp(-R_s) + \sqrt{\pi R_s}\,\mathrm{erf}(\sqrt{R_s}) \tag{5-4}$$

A plot of Eq. (5-4) is shown in Fig. 5-2. As can be seen, the AGC voltage is a nonlinear function of R_s. The curve becomes quasi-linear for $R_s > 6$, where we can write

$$\mu_V^N \cong \sqrt{\pi R_s} + \frac{\exp(-R_s)}{2R_s} \tag{5-5}$$

and for $R_s > 10$,

$$\mu_V^N \cong \sqrt{\pi R_s} \tag{5-6}$$

Thus, under very strong signal conditions, the AGC loop for the DAR operates no differently than its discrete carrier counterpart. For very weak signals, the AGC voltage becomes limited by the receiver noise. This latter condition is actually an advantageous one compared to the threshold behavior of the discrete carrier AGC loop, where the mean AGC voltage is a strong, nonlinear function of the tracking loop phase noise For the DAR AGC case, where $\hat{\epsilon}$ and $\cos\phi(t)$ become significant factors because $P_D << \sigma_{n2}^2$, the dependence of μ_V on these factors is very small, and thus we still have AGC control even though the tracking loop is out of lock.

It is likely that more filtering of the AGC voltage than that provided by the matched filter integrator in the DAR is desirable. This can be easily mechanized by adding an additional accumulator to the magnitude output of the in-phase accumulator buffer of Fig. 4-17 The output of this accumulator may then be converted by a digital-to-analog converter (DAC) to provide the gain control voltage. If very large gain variations are desired, the most significant bits (MSBs) of the AGC accumulator may be employed to switch in or out fixed attenuator pads in the RF and IF amplifiers, while the less significant bits (LSBs) provide, through a DAC, proportional control in the usual manner. Figure 5-3 shows a functional diagram of the DAR AGC loop

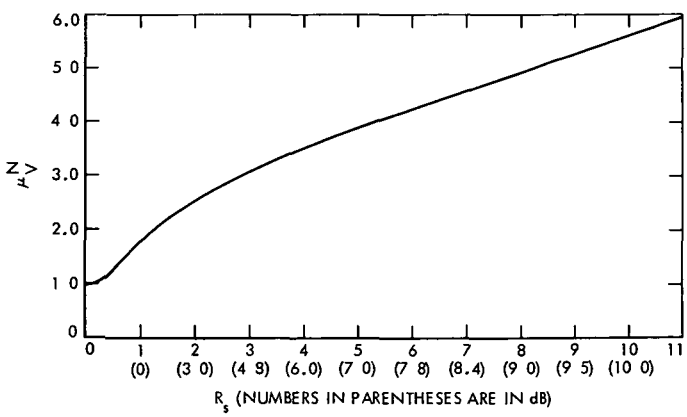

Fig. 5-2. Normalized DAR AGC voltage as a function of R_s

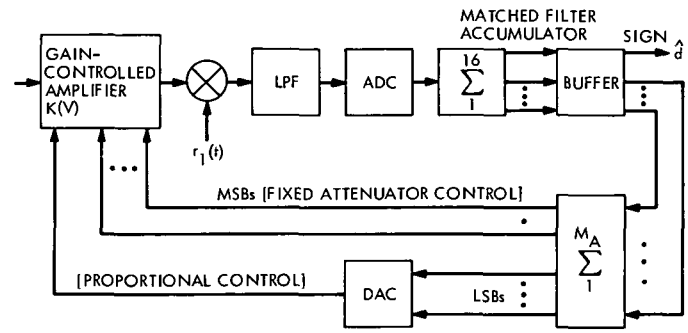

Fig. 5-3. DAR AGC loop

B. Lock Detection

To design a lock detector that indicates when the DAL is in or out of a locked condition, one must derive from the loop a voltage which on the average is large when the loop is *in lock* and commensurately small when the loop is *out of lock* Then, by comparing this voltage to a preset threshold (which may be set based upon the probability of detecting the in-lock or out-of-lock state), one may detect the presence or absence of lock.

One convenient method for deriving the lock indication voltage is to make use of the AGC voltage V, whose mean is defined in Eq (5-1), and its quadrature counterpart V_Q, which may be derived from the symbol delay accumulator (Fig 4-17) and whose mean is defined by

$$\mu_{V_Q} \triangleq E\{V_Q\} = E\left|\int_{\hat{\epsilon}}^{T_s+\hat{\epsilon}} \{\sqrt{P_D}\,d(t)\sin\phi(t) + N_2[t,\phi(t)]\}\,dt\right| \tag{5-7}$$

By accumulating V and V_Q over a large number of symbol intervals, and then taking the difference of these two quantities, one obtains a voltage which satisfies the above requirements on a lock indicator. A lock detector designed on this principle is illustrated in Fig. 5-4.

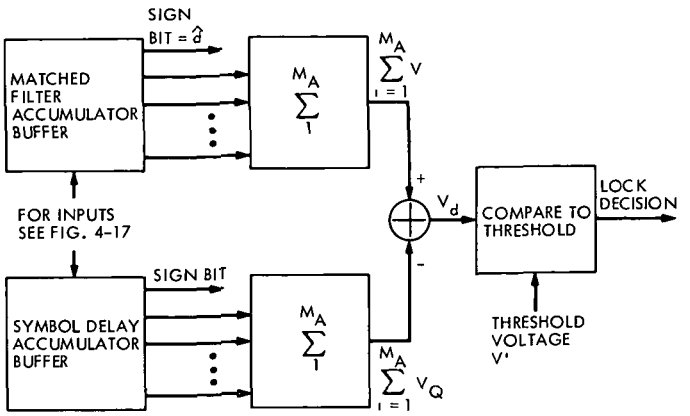

Fig. 5-4. Lock detector

The performance of this configuration is analyzed as follows If we assume that when the DAL is out of lock, the phase error varies such that it can be assumed to be constant over a symbol time, then, from Eqs (5-1), (5-3), and (5-7), we have

$$\mu_V = \frac{1}{2}\sqrt{\frac{N_0 T_s}{\pi}}\left\{\exp(-R_s\cos^2\phi) + \exp\left[-R_s\cos^2\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right]\right\} + \frac{1}{2}\sqrt{P_D}\,T_s\cos\phi\left\{\operatorname{erf}(\sqrt{R_s}\cos\phi) + \left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\operatorname{erf}\left[\sqrt{R_s}\cos\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right]\right\} \tag{5-8}$$

$$\mu_{V_Q} = \frac{1}{2}\sqrt{\frac{N_0 T_s}{\pi}}\left\{\exp(-R_s\sin^2\phi) + \exp\left[-R_s\sin^2\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right]\right\} + \frac{1}{2}\sqrt{P_D}\,T_s\sin\phi\left\{\operatorname{erf}(\sqrt{R_s}\sin\phi) + \left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\operatorname{erf}\left[\sqrt{R_s}\sin\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right]\right\}$$

Accumulating V and V_Q over M_A symbol intervals, and defining the normalized difference variable

$$V_d^N \triangleq \frac{\sum_{i=1}^{M_A} V - \sum_{i=1}^{M_A} V_Q}{\sqrt{N_0 T_s/\pi}} \tag{5-9}$$

the mean of V_d^N is

$$\mu_{V_d^N} = \frac{1}{2}\left\{\exp(-R_s\cos^2\phi) + \exp\left[-R_s\cos^2\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right] - \exp(-R_s\sin^2\phi) - \exp\left[-R_s\sin^2\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right]\right\} + \frac{1}{2}\sqrt{\pi R_s}\left\{\cos\phi\operatorname{erf}(\sqrt{R_s}\cos\phi) + \cos\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\operatorname{erf}\left[\sqrt{R_s}\cos\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right] - \sin\phi\operatorname{erf}(\sqrt{R_s}\sin\phi) - \sin\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\operatorname{erf}\left[\sqrt{R_s}\sin\phi\left(1-\frac{2|\hat{\epsilon}|}{T_s}\right)\right]\right\} \tag{5-10}$$

When the loop is out of lock, we may assume that the phase error ϕ is uniformly distributed, thus, from (5-10),

$$\mu_{V_d^N} = 0 \tag{5-11}$$

Now, when the loop is locked, we may assume, as was done in the AGC case, that for large loop SNR, $\epsilon = \phi = 0$. Thus,

$$\mu_{V_d^N} = M_A\left[\exp(-R_s) + \sqrt{\pi R_s}\operatorname{erf}(\sqrt{R_s})\right] \tag{5-12}$$

When the number of accumulation symbol intervals M_A is large—i.e, greater than 10 (as would be the case in lock detection for a carrier regeneration loop)—the normalized voltage V_d^N is approximately gaussian distributed with the mean given by (5-10) and variance

$$\sigma^2_{V_d^N} = \frac{M_A\pi}{N_0 T_s}\left[\sigma_V^2 + \sigma_{V_Q}^2\right] = \frac{M_A\pi}{N_0 T_s}\left\{E\{V^2\} + E\{V_Q^2\} - [\mu_V]^2 - [\mu_{V_Q}]^2\right\} \tag{5-13}$$

where

$$E\{V^2\} = \frac{P_D T_s^2 \cos^2\phi}{2}\left[1 + \left(1 - \frac{2|\hat{\epsilon}|}{T_s}\right)^2\right] + \frac{N_0 T_s}{2}$$

$$E\{V_Q^2\} = \frac{P_D T_s^2 \sin^2\phi}{2}\left[1 + \left(1 - \frac{2|\hat{\epsilon}|}{T_s}\right)^2\right] + \frac{N_0 T_s}{2} \tag{5-14}$$

Again, when the loop is locked and the loop signal-to-noise ratio is large, we may assume that $\phi = \hat{\epsilon} = 0$, in which case (5-13) reduces to

$$\sigma^2_{V_d^N} = M_A\{\pi + \pi R_s - [\exp(-R_s) + \sqrt{\pi R_s}\,\mathrm{erf}\sqrt{R_s}]^2 - 1\} \tag{5-15}$$

For very large R_s, Eqs (5-12) and (5-15), respectively, simplify to

$$\mu_{V_d^N} = M_A\sqrt{\pi R_s} \tag{5-16}$$

and

$$\sigma^2_{V_d^N} = M_A(\pi - 1) \tag{5-17}$$

If the normalized difference variable V_d^N is compared to a threshold V' and we decide that the loop is in lock when $V_d^N > V'$ and out of lock when $V_d^N < V'$, then *the probability of correctly detecting the loop's state* is given by

$$P_c = \text{prob}\left\{V_d^N > V' \,\middle|_{\text{loop is in lock}}\right\} \times \text{prob}\left\{\begin{matrix}\text{loop is}\\ \text{in lock}\end{matrix}\right\} + \text{prob}\left\{V_d^N < V' \,\middle|_{\text{loop is out of lock}}\right\} \times \text{prob}\left\{\begin{matrix}\text{loop is}\\ \text{out of lock}\end{matrix}\right\} \tag{5-18}$$

If the threshold is chosen between (perhaps midway) the values of $\mu_{V_d^N}$ when the loop is in lock and when it is out of lock, as given by Eqs (5-12) and (5-11), respectively, then the two conditional probabilities of (5-18) will be of the same order of magnitude Since the probability that the loop is in lock is much much greater than the probability that it is out of lock, Eq. (5-18) can be approximated by

$$P_c = \text{prob}\left\{V_d^N > V' \,\middle|_{\text{loop is in lock}}\right\} = \frac{1}{\sqrt{2\pi\sigma^2_{V_d^N}}}\int_{V'}^{\infty} \exp\left[-\frac{(x - \mu_{V_d^N})^2}{2\sigma^2_{V_d^N}}\right] dx$$

or

$$\boxed{P_c = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[\frac{\mu_{V_d^N} - V'}{\sqrt{2}\,\sigma_{V_d^N}}\right]} \tag{5-19}$$

where $\mu_{V_d^N}$ and $\sigma^2_{V_d^N}$ are given by (5-12) and (5-15), respectively.

For large R_s, (5-19) simplifies to

$$\boxed{P_c = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[\frac{M_A\sqrt{\pi R_s} - V'}{\sqrt{2M_A(\pi - 1)}}\right]} \tag{5-20}$$

Thus, for given values of R_s and P_c, and a specification on the time to detect the in- or out-of-lock state of the loop (equal to $M_A T_s$), one can specify the parameters M_A and V' For example, consider as suggested that $V' = \mu_{V_d^N}/2$ and $P_e^s = 1 \times 10^{-2}$. Then, from (5-19), it is calculated that for $M_A = 10$, $P_c = 0\,999989$

In conclusion, it should be noted that P_c is not a function of the absolute signal and noise levels, this is typical of a lock detector of the type illustrated in Fig. 5-4. The performance is only a function of the choice of V' relative to the value of $\mu_{V_d^N}$ corresponding to the worst-case value of R_s, and the accumulator size M_A Increasing M_A clearly improves the probability of correctly determining the loop's lock state at the expense of reaction time in the determination of either state

References

5-1 Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers*, Vol. I, Technical Report 32-819, Jet Propulsion Laboratory, Pasadena, Calif., Feb 15, 1966.

5-2 Clarke, K. K., and Hess, D. T., *Communication Circuits: Analysis and Design*, Addison-Wesley Publishing Co., Cambridge, Mass., 1971.

5-3. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw Hill Book Co., New York, N.Y., 1965.

VI. Complete System Configurations

In the following sections, we present outline discussions on how the data-aided loop (DAL) can be adapted to serve a number of specific tracking and demodulation functions. The intent here is only to define how the loop may be configured and to point out some of the salient features. The reader may then build upon these basic uses to obtain systems that satisfy a particular need.

A. Modifications of Existing Discrete Carrier Tracking Receivers to Achieve DAL Operation

Existing transponder and ground discrete carrier-tracking receivers may be easily modified so that they become suppressed carrier-tracking receivers of the DAL type. To illustrate this, a typical spacecraft transponder receiver is shown in Fig. 6-1, where the additions are drawn as dashed lines and existing blocks that may have to be modified are shaded. A ground receiver would be modified in a similar fashion.

The DAL implementation block embodies the functions shown in Figs. 4-1, 4-17, and 5-3. One existing receiver block that definitely would have to be changed is the IF bandpass filter (BPF). In most receivers this filter is quite narrow, and would have to be replaced by one whose bandwidth is sufficiently wide to pass the data modulation sidebands. It may further be necessary to modify or replace the IF amplifier following the filter, as this may also have undesirable narrowband characteristics. The automatic gain control (AGC) filter may also have to be changed, so that the AGC loop discussed in Section V-A will function properly. Finally, if a "tight" AGC loop is designed, it may be possible to eliminate the limiter entirely. If the limiter is retained, however, it should also be modified to eliminate any narrowband properties.

For an uplink DAL, an interesting possibility exists of resolving the carrier phase ambiguity (mentioned in Sections II-C and III-C) during initial acquisition Assuming that a lock detector of the type described in Section V-B is implemented, it can be used, when the loop is out of lock (e.g., during acquisition), to disable the data symbol estimate from the formation of the loop error function by forcing $\hat{d}$ equal to one (Fig. 3-2). Under these conditions, the loop is a normal discrete carrier-tracking loop, with the exception of the delay element T_s To acquire the uplink, then, an unmodulated carrier is transmitted, to which the loop locks at one unique point in the carrier cycle Now, when the lock detector indicates the in-lock condition, the data symbol estimate is enabled, and the loop functions as a DAL, with the transmitted data symbols being a series of ones The loop switches its effective S-curve from $\sin\phi$ to $\mathcal{S}(\phi)$, as given by Eq. (3-28), without perturbing the lock point. Later, when uplink information is to be transmitted, the carrier reference is

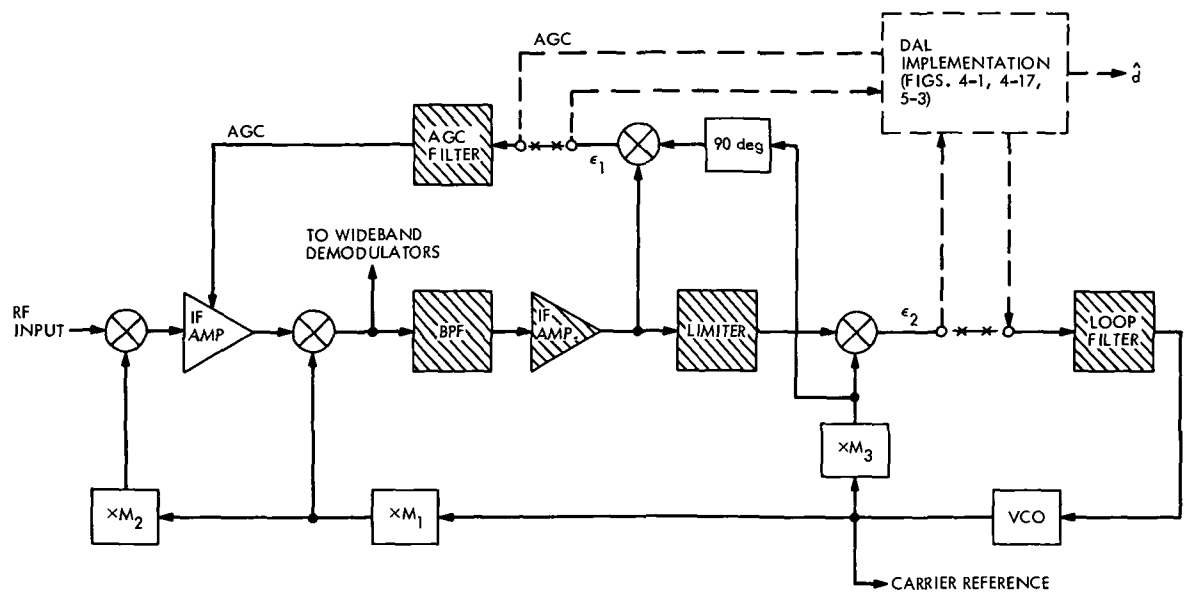

Fig. 6-1. Transponder receiver modifications necessary to implement the DAL

at the proper phase to demodulate the symbols without ambiguity.

B. Subcarrier Tracking

The use of the DAL is, of course, not constrained to carrier frequencies. It may also be used to track suppressed subcarriers that are encountered in discrete carrier systems, and suppressed carrier systems having more than one data channel (Section C). Assuming that a discrete carrier is modulated by a subcarrier which is, in turn, biphase modulated by the data, the configuration shown in Fig. 6-2 can be used to track and demodulate a sinusoidal subcarrier. Figure 6-3 indicates how a square-wave subcarrier might be handled. The only real difference between these two configurations is that the type 2

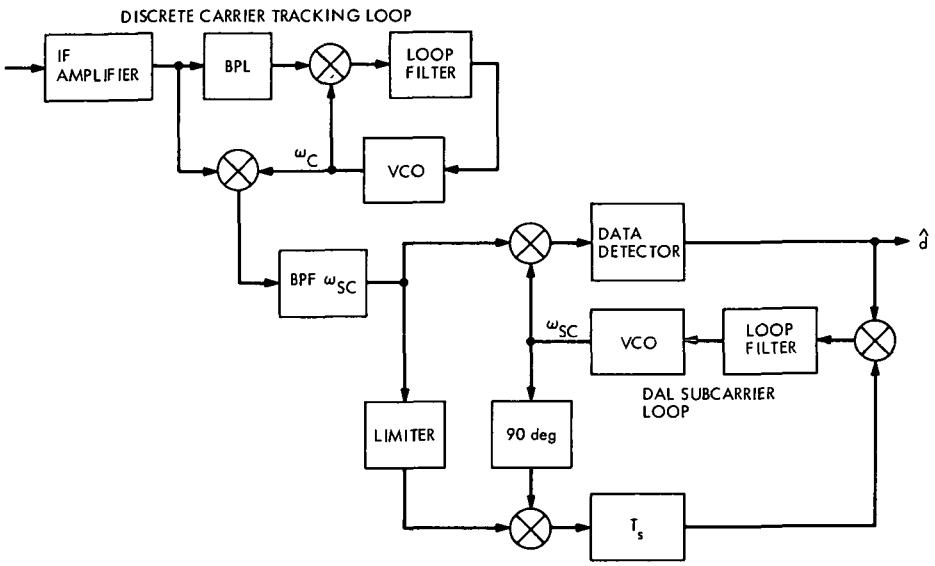

Fig. 6-2. Type 1 subcarrier DAL

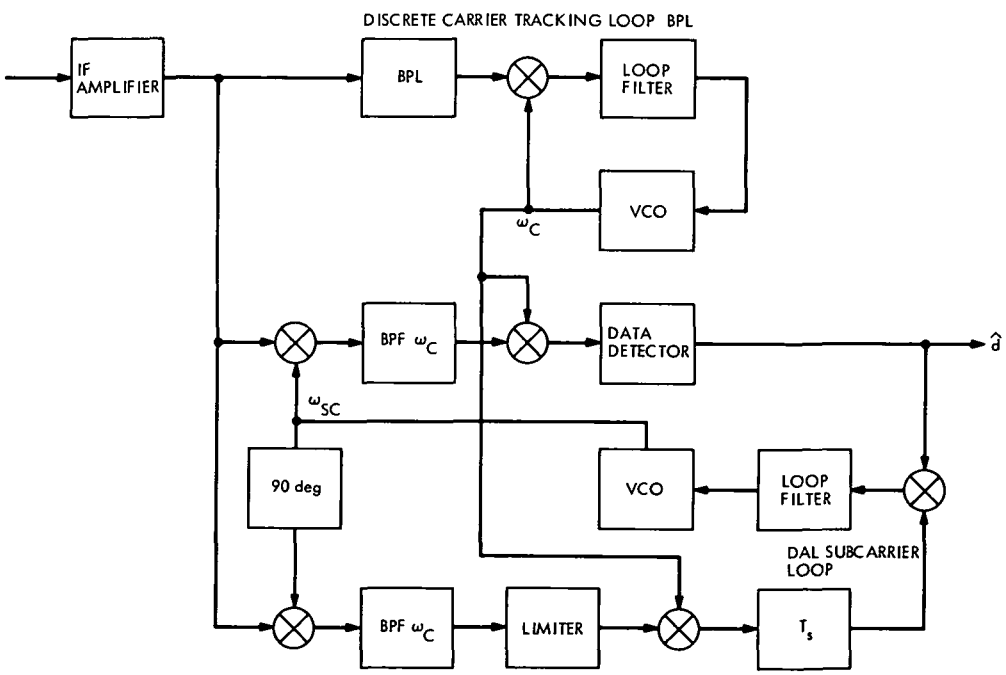

Fig. 6-3. Type 2 subcarrier DAL

subcarrier DAL (Fig. 6-3) is easily able to recover the harmonic power of the square-wave subcarrier without undue specifications on the BPFs, while the type 1 subcarrier DAL, if used with square-wave subcarriers, either filters out all of the harmonic power or, if the subcarrier BPF is removed, has to operate with extreme limiter suppression factors

C. Two-Channel Configurations

Equation (2-8) gives the general expression for two-channel suppressed carrier modulation. The situation for quadriphase, with two equal power, equal data-rate carrier orthogonal modulations, is a special case and is covered in Section D We are concerned here with the two-channel cases in which there is a significant difference between the individual modulating data rates

A useful signal form for keeping the two data channels spectrally separate is

$$s(t) = \sqrt{2P_{D1}}\, d_1(t) \cos \omega_C t + \sqrt{2P_{D2}}\, d_2(t)\, \mathrm{S}(t) \sin \omega_C t \tag{6-1}$$

which may be obtained from (2-8) by letting

$$\mathrm{S}_1(t) = 1$$

and

$$\mathrm{S}_2(t) \triangleq \mathrm{S}(t)$$

In (6-1), the data modulation $d_1(t)$ has symbol rate $\mathscr{R}_{s1}$, $d_2(t)$ has rate $\mathscr{R}_{s2}$, and we assume that $\mathscr{R}_{s1} >> \mathscr{R}_{s2}$, implying that $P_{D1} >> P_{D2}$. The square-wave subcarrier $\mathrm{S}(t)$, onto which $d_2(t)$ is biphase modulated, is selected so that there is as little mutual interference between the two data signals as possible within the overall bandwidth constraint placed on the system A typical spectrum for $s(t)$ appears in Fig. 6-4

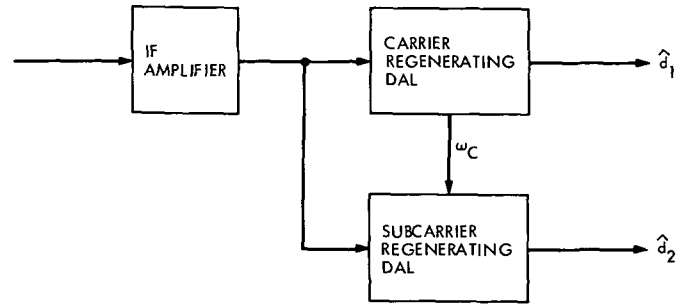

Fig. 6-5. Two-channel suppressed carrier functional data-aided receiver

With this signal format, the receiver system is essentially the same as that shown in Fig. 6-3, with the exception that the discrete carrier-tracking loop is replaced with a DAL which tracks the high-data-rate signal component. A functional diagram is shown in Fig. 6-5. The combination of bandpass and lowpass filters in each loop configuration effectively eliminates mutual interference.

There are situations in which it may be advantageous always to have the low-power, low-data-rate signal directly modulate the carrier in the two-channel mode. Consider as an example an uplink to the spacecraft. At times, this link is operated single-channel, i.e., command modulation only, while on other occasions it is operated in a dual-channel mode, i e., command plus ranging modulation

Typically, the command component has considerably less power than the ranging component. Thus, the superior data-aided receiver (DAR) operating conditions would be obtained by tracking the ranging signal. This would mean that the ranging signal would directly modulate the carrier, while the command data would have to be modulated onto an appropriate subcarrier. This, however, presents certain receiver reconfiguration problems when

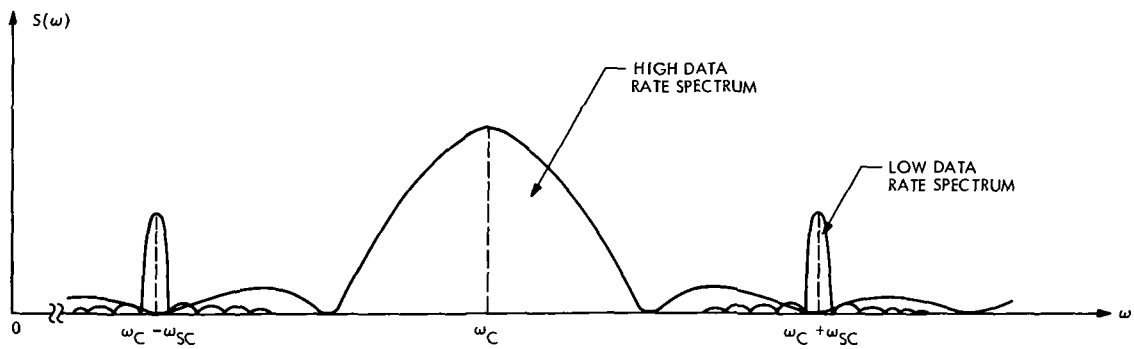

Fig. 6-4. Typical two-channel suppressed carrier spectrum

changing from single-channel (command only) to dual-channel (command plus ranging) operation, and vice-versa To change from single- to dual-channel operation, the receiver would have to recognize that the signals had changed, and then switch the carrier regenerating DAL to the ranging symbol rate, reacquire, and allow the subcarrier regenerating DAL to take up the command demodulation/detection function. Not only is this a complex switching situation to both recognize and implement in the receiver, but it also means that uplink carrier lock has to be broken for a period of time, thus disrupting two-way tracking

The alternate, and perhaps more realistic, solution is to continue transmitting the command data directly on the carrier, and place the ranging on the subcarrier. Thus, even though the DAL signal-to-noise ratio (SNR) is then reduced by the amount of transmitter power allocated to the ranging, the loop does not require reconfiguration, and does not lose lock. If, in the two-channel case, the DAL SNR falls below some desired minimum, it is possible to form a hybrid DAL wherein the ranging subcarrier DAL may be switched into the command DAL by the subcarrier DAL lock detector to aid and thus increase the SNR of the carrier regenerating loop. This concept is discussed in Section III-A.

D. Quadriphase and Higher-Order Tracking Loops

The extension of biphase modulation of a carrier to more than two phases is referred to as *polyphase* or *N-phase modulation*, where N is the total number of transmitted phases. In particular, during a given transmission interval of T seconds, the transmitted signal can be expressed in the form

$$s(t) = \sqrt{2P_D}\sin\left[\omega_C t + \frac{(2k+1)\pi}{N}\right];$$

$$k = 0, 1, 2, \quad , N-1 \qquad (6\text{-}2)$$

where ω_C is again the carrier radian frequency. For $N = 2$, the above signaling format represents phase-shift-keying, as discussed in Section II, while for $N = 4$, it corresponds to quadriphase-shift-keying Actually, for any value of $N > 2$, we refer to the signaling technique as multiple-phase-shift-keying (MPSK). In the quadriphase case, the transmitted signal in (6-2) assumes the form

$$s(t) = \sqrt{P_D}\,[d_1(t)\sin\omega_C t + d_2(t)\cos\omega_C t] \qquad (6\text{-}3)$$

where $d_1(t)$ and $d_2(t)$ are ± 1 digital waveforms whose transitions occur at T-second intervals For quadriphase signaling, the above equation suggests the modulator depicted in Fig. 6-6a Setting $N = 8$ (i.e., octaphase modulation), it is easy to show that the circuit in Fig. 6-6b generates an octaphase signal. In this figure, $d_3(t)$ and $d_4(t)$ also correspond to data sequences of plus or minus ones. The generalization of the transmitter modulator to N phases is straightforward.

If one assumes that the channel adds white gaussian noise $n(t)$ of single-sided spectral density N_0 W/Hz and a phase shift to the signal $s(t)$, then the received signal can be characterized by

$$x(t) \triangleq \sqrt{2P_D}\sin\left[\Phi(t) + \frac{(2k+1)\pi}{N}\right] + n(t) \qquad (6\text{-}4)$$

where $\Phi(t) \triangleq \omega_C t + \theta_0$, with θ_0 a uniformly distributed random phase.

The reconstruction of a carrier from $x(t)$ can be accomplished in many ways. For example, the squaring and

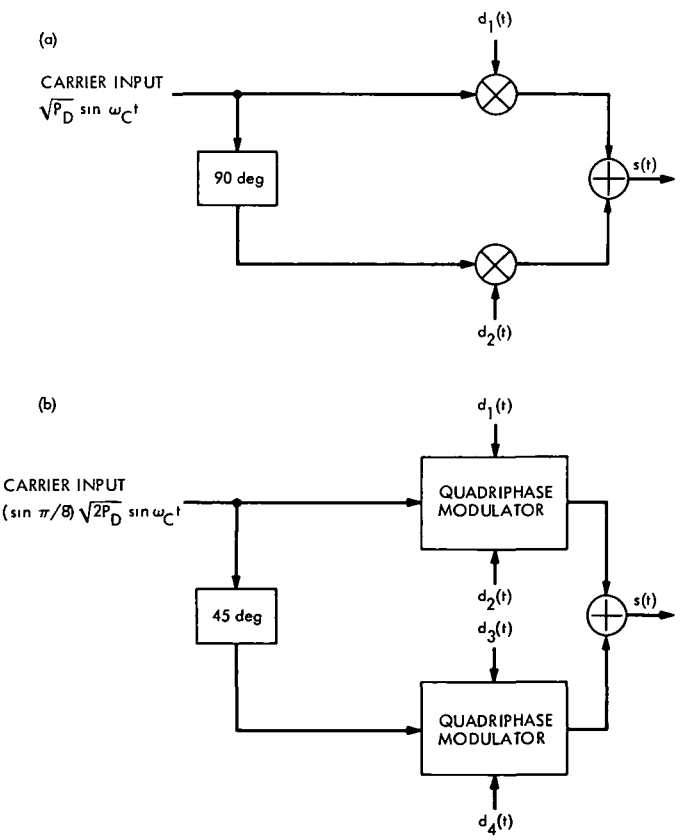

Fig. 6-6. Modulator mechanization: a. quadriphase, b. octaphase

Costas loops used for tracking a biphase-modulated carrier can be generalized to Nth power loops and N-phase Costas loops, respectively (Ref 6-1). We are interested here, however, in the generalization of the biphase suppressed carrier DAL to an N-phase configuration. In what follows, we shall demonstrate one configuration which is directly applicable to quadriphase signaling and then a more general configuration that is useful for all values of N

The biphase suppressed carrier DAL can easily be modified to accommodate a quadriphase input signal Since, from (6-3), the quadriphase signal input can be modeled as the sum of two orthogonal, equal power suppressed carrier signals, one might anticipate that the necessary extension of the above loop is the addition of a quadrature loop to the existing configuration. Indeed, this is the case, as is illustrated in the quadriphase decision-feedback loop of Fig. 6-7. Letting $\hat{\Phi}(t)$ represent the loop's estimate of $\Phi(t)$ and

$$r_u(t) = \sqrt{2}\,K_1 \cos\hat{\Phi}(t)$$

$$r_l(t) = -\sqrt{2}\,K_1 \sin\hat{\Phi}(t) \tag{6-5}$$

respectively, denote the upper and lower loop reference signals, then the upper and lower dynamic reference signals $\epsilon_u(t)$ and $\epsilon_l(t)$ are given by

$$\begin{aligned}\epsilon_u(t) = \frac{K_1K_{1m}}{\sqrt{2}}\{&\sqrt{P_D}\,d_1(t)\sin\phi(t)\\ &+\sqrt{P_D}\,d_2(t)\cos\phi(t)\\ &+\sqrt{2}\,N_u[t,\phi(t)]\}\\ \epsilon_l(t) = \frac{K_1K_{2m}}{\sqrt{2}}\{&-\sqrt{P_D}\,d_1(t)\cos\phi(t)\\ &+\sqrt{P_D}\,d_2(t)\sin\phi(t)\\ &+\sqrt{2}\,N_l[t,\phi(t)]\}\end{aligned} \tag{6-6}$$

where $\phi(t) \triangleq \Phi(t) - \hat{\Phi}(t)$ and $N_u[t,\phi(t)]$ and $N_l[t,\phi(t)]$ are uncorrelated noise processes, which are modeled as

$$\begin{aligned}N_u[t,\phi(t)] &= N_c(t)\cos\phi(t) - N_s(t)\sin\phi(t)\\ N_l[t,\phi(t)] &= N_c(t)\sin\phi(t) + N_s(t)\cos\phi(t)\end{aligned} \tag{6-7}$$

Proceeding with an analysis analogous to that presented for the suppressed carrier DAL, it can be shown that the signals $z_u(t)$ and $z_l(t)$ are given by

$$\begin{aligned}z_u(t) = \frac{K_1K_{1m}}{\sqrt{2}}F_1(p)\exp(-pT)&\\ \times\{&\sqrt{P_D}\,d_1(t)\hat{d}_1(t)\sin\phi(t)\\ &+\sqrt{P_D}\,d_2(t)\hat{d}_1(t)\cos\phi(t)\\ &+\sqrt{2}\,\hat{d}_1(t)N_u[t,\phi(t)]\}\\ z_l(t) = \frac{K_1K_{2m}}{\sqrt{2}}F_2(p)\exp(-pT)&\\ \times\{&-\sqrt{P_D}\,d_1(t)\hat{d}_2(t)\cos\phi(t)\\ &+\sqrt{P_D}\,d_2(t)\hat{d}_2(t)\sin\phi(t)\\ &+\sqrt{2}\,\hat{d}_2(t)N_l[t,\phi(t)]\}\end{aligned} \tag{6-8}$$

Following the same averaging argument for $E\{d_i(t)\hat{d}_i(t)\}$ as previously given for the biphase DAL and the fact that $E\{d_i(t)\hat{d}_j(t)\} = 0$ for $i \neq j$, (6-8) reduces to

$$\begin{aligned}z_u(t) = \frac{K_1K_{1m}}{\sqrt{2}}F_1(p)\exp(-pT)&\\ \times\sqrt{P_D}[1-2P_{E1}[\phi(t)]]\sin\phi(t)&\\ +\sqrt{2}\,\hat{d}_1(t)N_u[t,\phi(t)]\}&\\ z_l(t) = \frac{K_1K_{2m}}{\sqrt{2}}F_2(p)\exp(-pT)&\\ \times\{\sqrt{P_D}[1-2P_{E2}[\phi(t)]]\sin\phi(t)&\\ +\sqrt{2}\,\hat{d}_2(t)N_l[t,\phi(t)]\}&\end{aligned} \tag{6-9}$$

It is clear that, in general, the transfer function factor $\exp(-pT)$, with $p = j\omega$, affects loop stability and reduces the signal acquisition or pull-in range. These phenomena have been discussed relative to the binary DAL, and thus, we shall not pursue them further here. Instead, we make the usual assumption that $W_LT << 1$, whereupon $\exp(-j\omega T)$ is approximately unity for all ω within the equivalent bandwidth of the loop. Thus, from a steady-state performance standpoint, this factor has negligible effect.

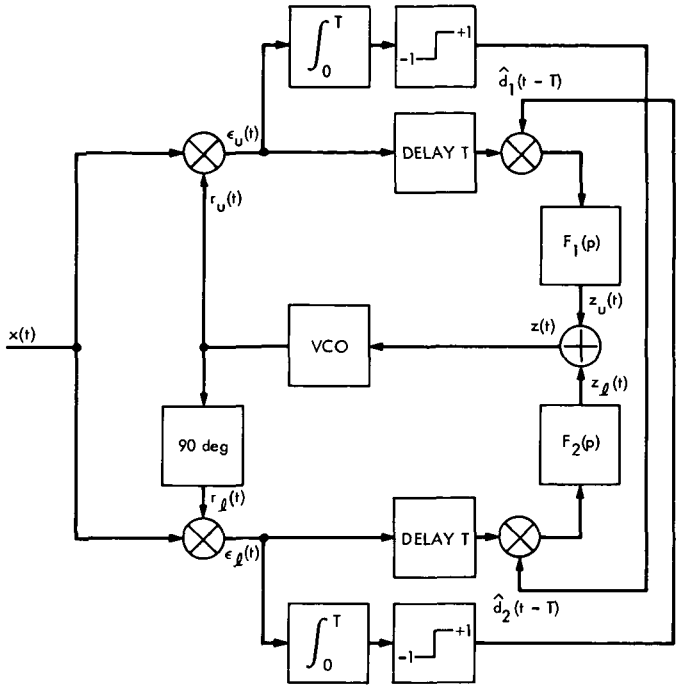

Fig. 6-7. A quadriphase decision-feedback loop

Letting $z(t) = z_u(t) + z_l(t)$, then the stochastic integro-differential equation of the quadriphase decision-feedback loop becomes

$$\dot{\phi}(t) = \dot{\theta}(t)$$

$$- \frac{K_u}{\sqrt{2}} F_1(p) \{\sqrt{P_D}[1 - 2P_{E1}[\phi(t)]] \sin\phi(t)$$

$$+ \sqrt{2}\, \hat{d}_1(t) N_u[t, \phi(t)]\}$$

$$- \frac{K_l}{\sqrt{2}} F_2(p) \{\sqrt{P_D}[1 - 2P_{E2}[\phi(t)]] \sin\phi(t)$$

$$+ \sqrt{2}\, \hat{d}_2(t) N_l[t, \phi(t)]\} \qquad (6\text{-}10)$$

where $K_u \triangleq K_1 K_{1m} K_V$ and $K_l \triangleq K_1 K_{2m} K_V$.

The quadriphase decision-feedback loop of Fig. 6-7 can in principle be extended to N phases To see this, we first note that the polyphase signal in (6-2) can also be written in the binary representation

$$s(t) = \left(\sin\frac{\pi}{N}\right)\sqrt{2P_D}\sum_{k=0}^{N/2-1} d_k(t) \sin\left(\omega_C t + \frac{k\pi}{N}\right) \qquad (6\text{-}11)$$

where the set $\{d_k(t); k = 0, 1, 2, \quad, N/2 - 1\}$ are again ± 1 digital waveforms whose transitions occur at T-second intervals. Equation (6-11) represents a generalization of the quadriphase signal given in (6-3) The N-phase decision-feedback loop then would have $N/2$ arms corresponding to the $N/2$ decisions $\hat{d}_k(t)$; $k = 0, 1, 2, \quad, N/2 - 1$ on the binary signals $d_k(t)$, $k = 0, 1, 2, \cdot \quad, N/2 - 1$. The output of the VCO would be phase-shifted by $k\pi/N$; $k = 0, 1, 2, \quad, N/2 - 1$ to provide $N/2$ coherent reference signals. As a result, the $N/2$ equivalent additive low-pass noise processes at the input of the summing junction are no longer mutually uncorrelated, and hence, for $N > 4$, the loop becomes less efficient.

Another N-phase tracking loop which makes use of the principle of decision feedback and is equally efficient for all N is illustrated in Fig 6-8. Here, in each T-second interval, the decision $\hat{\theta}_k$ on the transmitted phase symbol $\theta_k = (2k + 1)\pi/N$ is used to produce the decision-feedback signals. Under the same assumptions given for the previous decision-feedback loop, the dynamic error signal becomes

$$\epsilon(t) = K_1 K_m \{\sqrt{P_D}\cos(\theta_k - \hat{\theta}_k)\sin\phi(t)$$

$$+ \sqrt{P_D}\sin(\theta_k - \hat{\theta}_k)\cos\phi(t)$$

$$+ \cos\hat{\theta}_k N_u[t, \phi(t)] + \sin\hat{\theta}_k N_l[t, \phi(t)]\} \qquad (6\text{-}12)$$

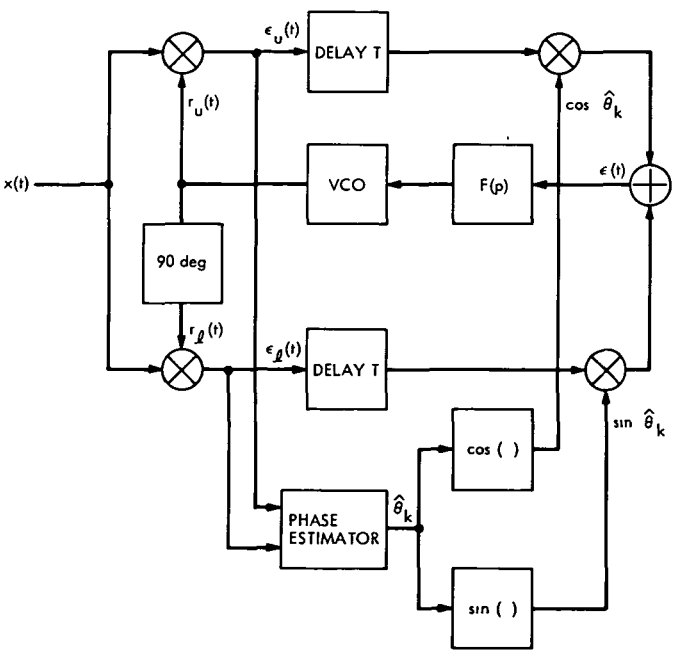

Fig. 6-8. An N-phase decision-feedback loop

where $N_u[t, \phi(t)]$ and $N_l[t, \phi(t)]$ are given by (6-7). Based upon the same averaging argument used previously, the output of the loop filter in the tracking mode can be expressed in terms of the circular moments of $\theta_k - \hat{\theta}_k$, viz.,

$$z(t) = K_1 K_m F(p) \{\sqrt{P_D}\,\overline{\cos(\theta_k - \hat{\theta}_k)} \sin\phi(t) + \sqrt{P_D}\,\overline{\sin(\theta_k - \hat{\theta}_k)} \cos\phi(t) + \cos\hat{\theta}_k N_u[t, \phi(t)] + \sin\hat{\theta}_k N_l[t, \phi(t)]\} \tag{6-13}$$

This discrete random variable, $\theta_k - \hat{\theta}_k$, ranges over the set of allowable values $2j\pi/N$, $j = 0, \pm 1, \pm 2, \ldots, \pm N/2 - 1, N/2$ with probabilities (Ref 6-1)

$$P_j(\phi) \triangleq \text{prob}\left\{\theta_k - \hat{\theta}_k = \frac{2j\pi}{N}\right\} = \frac{1}{\pi}\int_0^\infty \left[\exp\{-(u - \sqrt{R_D}\cos\phi)^2\} \times \int_{u\tan[(2j+1)\pi/N]}^{u\tan[(2j-1)\pi/N]} \exp\{-(v - \sqrt{R_D}\sin\phi)^2\}\, dv\right] du \tag{6-14}$$

where we have assumed that the loop phase error $\phi(t)$ is essentially constant over several signaling intervals. Also, from the law of total probability,

$$\sum_{j=-N/2+1}^{N/2} P_j(\phi) = 1 \tag{6-15}$$

Thus, from (6-14) and (6-15), the circular moments of $\theta_k - \hat{\theta}_k$ can be expressed as

$$\overline{\sin(\theta_k - \hat{\theta}_k)} \triangleq \sum_{j=-N/2+1}^{N/2} P_j(\phi) \sin\frac{2j\pi}{N} = \sum_{j=-N/2+1}^{N/2}{}' P_j(\phi) \sin\frac{2j\pi}{N}$$

$$\overline{\cos(\theta_k - \hat{\theta}_k)} \triangleq \sum_{j=-N/2+1}^{N/2} P_j(\phi) \cos\frac{2j\pi}{N} = P_0(\phi) + \sum_{j=-N/2+1}^{N/2}{}' P_j(\phi) \cos\frac{2j\pi}{N} \tag{6-16}$$

where the prime on the summation denotes omission of the $j = 0$ term.

We note that $P_0(\phi)$ is the conditional probability that the decision on θ_k is correct given the phase error ϕ. Thus, letting

$$P_E(\phi) \triangleq 1 - P_0(\phi) = \sum_{j=-N/2+1}^{N/2}{}' P_j(\phi) \tag{6-17}$$

Eq. (6-16) may be expressed in the equivalent form

$$\overline{\sin(\theta_k - \hat{\theta}_k)} = \sum_{j=-N/2+1}^{N/2}{}' P_j(\phi) \sin\frac{2j\pi}{N}$$

$$\overline{\cos(\theta_k - \hat{\theta}_k)} = 1 - 2P_E(\phi) + \sum_{j=-N/2+1}^{N/2}{}' P_j(\phi)\left[1 + \cos\frac{2j\pi}{N}\right] \tag{6-18}$$

Substituting (6-18) into (6-13) and recalling that

$$\phi(t) = \theta(t) - K_V \frac{z(t)}{p} \tag{6-19}$$

the stochastic integro-differential equation of operation for the N-phase decision-feedback loop of Fig. 6-8 becomes

$$\dot{\phi}(t) = \dot{\theta}(t) - KF(p)\left\{\sqrt{P_D}\,[1 - 2P_E(\phi)] \sin\phi + \sqrt{P_D}\sum_{j=-N/2+1}^{N/2}{}' P_j(\phi) \sin\phi\left(1 + \cos\frac{2j\pi}{N}\right) + \sqrt{P_D}\sum_{j=-N/2+1}^{N/2}{}' P_j(\phi) \cos\phi \sin\frac{2j\pi}{N} + \cos\hat{\theta}_k N_u(t, \phi) + \sin\hat{\theta}_k N_l(t, \phi)\right\} \tag{6-20}$$

where $K \triangleq K_1 K_m K_V$. Recognizing from (6-14) that $P_j(\phi) = P_{-j}(-\phi)$, the second and third terms of (6-20) are odd functions of ϕ and as such contribute to the overall tracking error characteristic. When $N = 2$, both of these terms are zero and (6-20) reduces to the stochastic integro-differential equation of operation for the biphase suppressed carrier DAL as given by (3-15).

In conclusion, we note that the theory presented in this section and the results given in Ref. 6-2 can be used to study the cycle slipping phenomenon of these two circuits. This, however, shall not be done here Also, a discussion of the data detection and noisy reference problem associated with the use of polyphase signaling is presented in Ref. 6-1.

E. Single-Sideband Receivers

The data-aided receiver may be used to track and demodulate single-sideband suppressed carrier (SSB-SC) signals The mathematical form of such a signal is given by

$$s(t) = \sqrt{P_D}\, d(t) \cos \omega_C t - \sqrt{P_D}\, \tilde{d}(t) \sin \omega_C t \quad (6\text{-}21)$$

where $\tilde{d}(t)$ is the Hilbert transform of $d(t)$ (See Ref 6-3 for a general discussion of the SSB-SC signal.)

There are two ways by which the DAR can be realized The first involves choosing the input BPF (see Fig. 2-5) so that it is essentially centered on the sideband If an ideal rectangular filter is postulated, then the outputs of the loop phase detectors (Fig 3-2) can be written as

$$\epsilon_1 = \frac{\sqrt{2}}{2} \{\sqrt{P_D}\, d(t) \cos \phi - \sqrt{P_D}\, \tilde{d}(t) \sin \phi + n_{1s}(t)\} \quad (6\text{-}22)$$

$$\epsilon_2 = \frac{\sqrt{2}}{2} \{\sqrt{P_D}\, d(t) \sin \phi - \sqrt{P_D}\, \tilde{d}(t) \cos \phi - n_{2s}(t)\} \quad (6\text{-}23)$$

where $n_{1s}(t)$ and $n_{2s}(t)$ are the low-pass equivalents of the single-sideband noise formed by the input BPF operating on $n_i(t)$ (see Eq. 3-2), and are independent and gaussian with single-sided spectral densities $N_0/2$ W/Hz.

When $\phi = 0$, Eq (6-22) is essentially the same for the basic DAL discussed in Section III We see, however, that ϵ_2 now contains an extra term, $-\sqrt{P_D}\, \tilde{d}(t)$. The presence of this term has no direct effect on the loop error function since $E\{\tilde{d}(t)\} = 0$ In fact, $\tilde{d}(t)$ is a series of identical—except for sign—odd functions, each of which is centered on the respective symbols comprising $\tilde{d}(t)$ Even when $\phi \neq 0$, the presence of $\tilde{d}$ in either ϵ_1 or ϵ_2 will have no effect since d, when averaged over one symbol period (excepting large symbol timing errors), is zero. Thus, the DAL will operate as if it were tracking a double-sideband suppressed carrier signal.

What counts in terms of the receiver performance is the exact variance of n_{1s} and n_{2s} Ideally, as already stated, these noise components will have noise spectral density $N_0/2$. However, with practical bandpass filters, it is impossible to filter out all of the noise on the side of the carrier frequency for which the sideband has been suppressed Thus, the variance of n_{1s} and n_{2s} will be somewhat larger than ideally possible. For two-pole bandpass filters, the SNR degradation due to this effect has been measured to be about 0.7 dB The P_e^s-related performance is shown in Fig. 6-9, together with the double-sideband performance taken from Fig. 4-19

The second method by which the SSB-SC DAL receiver may be implemented is shown in Fig 6-10 The blocks labeled HT perform the Hilbert transform operation. For this configuration, there are no special requirements on the input BPF. Now, when $\phi(t)$ is slowly varying with respect to $d(t)$, it can be shown that

$$\epsilon_1 + \tilde{\epsilon}_2 = \sqrt{2P_D}\, d(t) \cos \phi + n_1(t) - \tilde{n}_2(t) \quad (6\text{-}24)$$

$$\epsilon_2 + \tilde{\epsilon}_1 = \sqrt{2P_D}\, d(t) \sin \phi + \tilde{n}_1(t) - n_2(t) \quad (6\text{-}25)$$

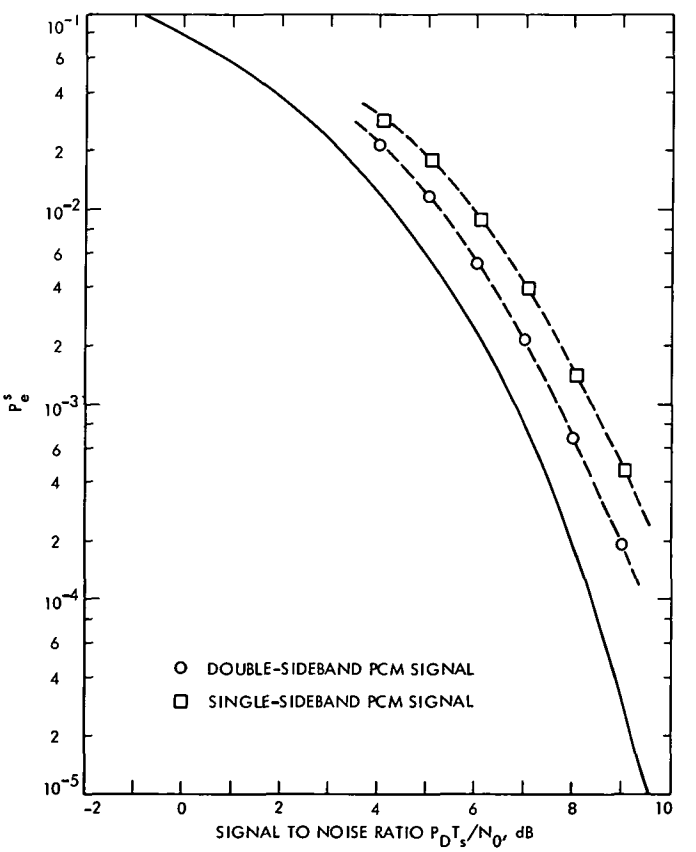

Fig. 6-9. Single-sideband suppressed carrier DAR performance

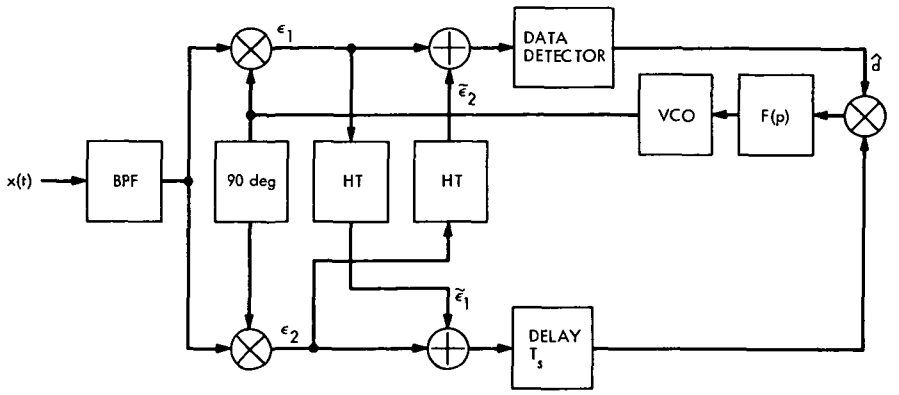

Fig. 6-10. An alternative SSB-SC data-aided receiver

where $n_1 - \tilde{n}_2$ and $\tilde{n}_1 - n_2$ are uncorrelated, gaussian, and have single-sided noise spectral densities $2N_0$ W/Hz. This receiver is ideally optimum, and in practice will perform according to how well the Hilbert transform operations can be implemented.

F. Subsymbol Integration and Decision for Very-Low-Data-Rate Applications

In Section III-D, it was shown that when the symbol rate $\mathcal{R}_s$ approaches the loop noise bandwidth W_L, the DAL will become unstable. From Eq. (3-60), we require that $\mathcal{R}_s > 4.3\,W_L$ if the loop is to be stable. Thus, if W_L is specified by, say, the need to track signal doppler rates, then there is a lower bound set on the symbol rate. For example, if $W_L = 18$ Hz and $\mathcal{R}_s = \mathcal{R}_b = 16$ bps (an uncoded bit rate), clearly the loop would be unstable.

For the situation in which the symbol (or bit) error probability is nominally small, there is a solution to the problem. For the bandwidth and bit-rate numbers already cited, suppose that, in addition, $P_e^b = 10^{-5}$. These numbers are typical of a command situation. Consider now that each bit is subdivided into n equal-length sub-bits, and that the DAL is implemented so that its $\hat{d}$ decisions are made on the basis of the sub-bits rather than the whole bit (Fig 6-11). Then, as far as the loop is concerned, it is operating at an effective symbol rate of $n\mathcal{R}_b$, so that the stability criterion of (3-60) becomes

$$n\mathcal{R}_b > 4.3\,W_L \tag{6-26}$$

The cost is that the error probability which determines the efficiency loss L_R (see Sections II and III) is much larger. However, if P_e^b is very small to begin with, this is a tolerable situation. Let us continue the example with the above command system numbers. Solving for n using Eq. (6-26) gives $n > 4.8$. Let us select $n = 8$, since this then gives a considerable stability margin. The effective symbol error probability for the loop then becomes 6×10^{-2} If we use the $L_R = 1 - 2P_e^s$ relationship, then $L_R = 0.6$ dB and the loop SNR, ρ, is 8.8 dB, an acceptable operating point. The bit decision itself is, of course, still based on a whole bit, so that the bit error probability remains 10^{-5}.

G. Two-Way Systems

The application of DALs to *two-way* communication systems necessitates consideration of three possible modulation–demodulation alternatives

(1) Suppressed carrier uplink and discrete carrier downlink—the spacecraft receiver contains a DAL and the ground receiver contains a discrete carrier-tracking loop.

(2) Discrete carrier uplink and suppressed carrier downlink—the spacecraft receiver contains a discrete carrier-tracking loop and the ground receiver contains a DAL

(3) Suppressed carrier uplink and suppressed carrier downlink—the spacecraft and ground receivers both contain DALs.

Since Fig. 6-1 shows the necessary modifications of an existing spacecraft transponder receiver to accommodate a DAL, let us begin by considering the first of the above three alternatives Also, we shall assume two-channel configurations for both the up- and downlinks, wherein the transmitted signal is modeled as in Eq. (6-1).

A block diagram of a two-way communication system based upon the above considerations is given in Fig. 6-12.

The ground transmitter emits the phase-modulated waveform

$$s_1(t) = \sqrt{2P_{D_{11}}}\, d_{11}(t)\cos\omega_C t + \sqrt{2P_{D_{21}}}\, d_{21}(t)\mathsf{S}_{21}(t)\sin\omega_C t \qquad (6\text{-}27)$$

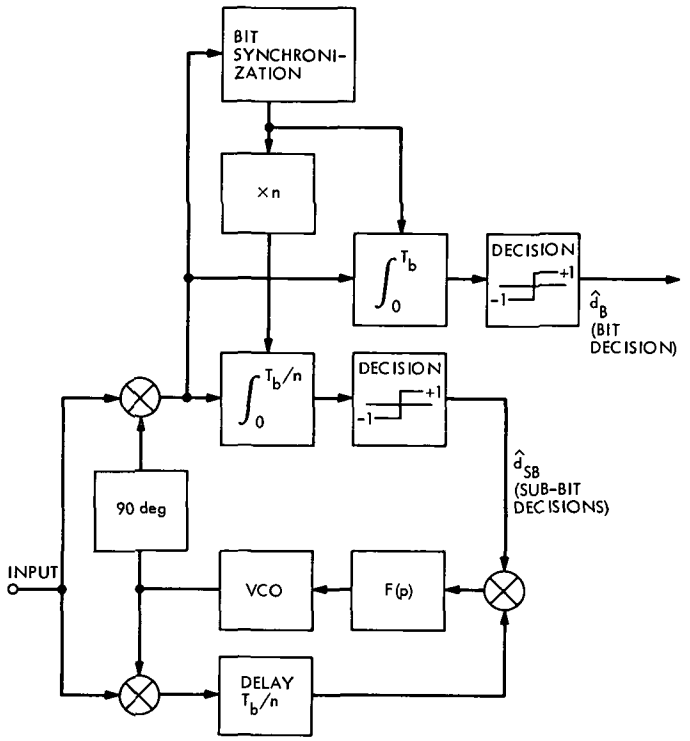

Fig. 6-11. Sub-bit decision DAR

We observe in the spacecraft the phase-shifted, noise-corrupted waveform

$$y_1(t) = \sqrt{2P_{D_{11}}}\, d_{11}(t)\cos(\omega_C t + \theta_1) + \sqrt{2P_{D_{21}}}\, d_{21}(t)\mathsf{S}_{21}(t)\sin(\omega_C t + \theta_1) + n_1(t) \qquad (6\text{-}28)$$

where $n_1(t)$ is white gaussian noise of single-sided spectral density N_{01} W/Hz Since a carrier component is not present in the received signal, the spacecraft employs a DAL to reconstruct a pair of reference signals $r_{2s}(t)$ and $r_{2c}(t)$ for demodulating $d_{11}(t)$ and $d_{21}(t)\mathsf{S}_{21}(t)$ off the carrier The carrier reference signal $r_{2s}(t)$ is also supplied to a subcarrier regenerating DAL (see Fig 6-5) for demodulating $d_{21}(t)$ off its subcarrier.

If we model $r_{2s}(t)$ and $r_{2c}(t)$ as

$$r_{2s}(t) = \sqrt{2}\sin[\omega_C t + \hat{\theta}_1(t)]$$
$$r_{2c}(t) = \sqrt{2}\cos[\omega_C t + \hat{\theta}_1(t)] \qquad (6\text{-}29)$$

where $\hat{\theta}_1(t)$ is the DAL's estimate of θ_1, then on the downlink, we transmit*

$$s_2(t) = \sqrt{2P_{T2}}\cos[G\omega_C t + \theta_{m_{12}} d_{12}(t)\mathsf{S}_{12}(t) + \theta_{m_{22}} d_{22}(t)\mathsf{S}_{22}(t) + G\hat{\theta}_1(t)] \qquad (6\text{-}30)$$

*Whether the standard or interplexed two-channel modulation format is used here is of no consequence

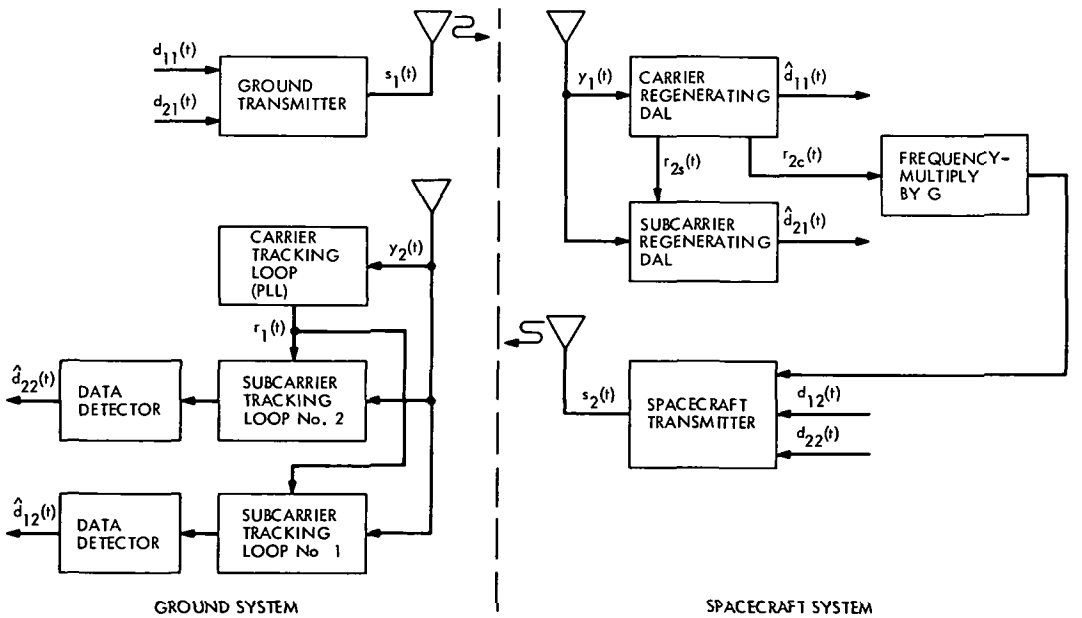

Fig. 6-12. Two-way communication system block diagram

where G is the transponder gain, whose value depends on the frequency band assignments of the uplink and downlink, e.g., S/S, S/X, X/X. At the ground station, we receive the phase-shifted, noise-corrupted waveform

$$y_2(t) = \sqrt{2P_{T2}} \cos [G\omega_C t + \theta_{m_{12}} d_{12}(t) \sqsubset_{12}(t) + \theta_{m_{22}} d_{22}(t) \sqsubset_{22}(t) + G\hat{\theta}_1(t) + \theta_2] + n_2(t) \quad (6\text{-}31)$$

where $n_2(t)$ is again white gaussian noise of single-sided spectral density N_{02} W/Hz and is independent of $n_1(t)$. Separating the signal component of $y_2(t)$ into its spectral components, we get

$$\begin{aligned} y_2(t) = {} & \sqrt{2P_{c2}} \cos [G\omega_C t + G\hat{\theta}_1(t) + \theta_2] \\ & - \sqrt{2P_{D_{12}}}\, d_{12}(t) \sqsubset_{12}(t) \sin [G\omega_C t + G\hat{\theta}_1(t) + \theta_2] \\ & - \sqrt{2P_{D_{22}}}\, d_{22}(t) \sqsubset_{22}(t) \sin [G\omega_C t + G\hat{\theta}_1(t) + \theta_2] \\ & - \sqrt{2P_{l_2}}\, d_{12}(t)\, d_{22}(t) \sqsubset_{12}(t) \sqsubset_{22}(t) \\ & \times \cos [G\omega_C t + G\hat{\theta}_1(t) + \theta_2] + n_2(t) \end{aligned} \quad (6\text{-}32)$$

The ground station receiver tracks $y_2(t)$ with a phase-locked loop (PLL) and produces the reference signal

$$r_1(t) = \sqrt{2} \sin [G\omega_C t + \hat{\theta}_2(t)] \quad (6\text{-}33)$$

where $\hat{\theta}_2(t)$ is the PLL's estimate of $G\hat{\theta}_1(t) + \theta_2$.

The phase error probability density functions (p d.f.'s) in the spacecraft and ground receivers' tracking loops are determined as follows. Let

$$\phi_1(t) \triangleq \theta_1(t) - \hat{\theta}_1(t)$$

and

$$\phi_2(t) \triangleq \theta_2(t) - \hat{\theta}_2(t)$$

be the loop phase errors in the spacecraft and ground receiver tracking loops, respectively, and let ϕ_1 and ϕ_2 be the modulo-2π reduced versions of these loop phase errors. Then, $p(\phi_1)$ is given by (3-16) together with (3-17), where P_D and R_s are now replaced by $P_{D_{11}}$ and $R_{s_{11}} = P_{D_{11}} T_{11}/N_{01}$, with T_{11} denoting the reciprocal of the data rate of $d_{11}(t)$. The p.d.f. $p(\phi_2)$ is given (assuming a first-order loop for simplicity) by the Tikhonov distribution

$$p(\phi_2) = \frac{\exp(\rho_2 \cos \phi_2)}{2\pi I_0(\rho_2)} \quad (6\text{-}34)$$

where

$$\rho_2 \triangleq \frac{2P_{c2}}{N_{02} W_{L2}}$$

with W_{L2} denoting the two-sided bandwidth of the ground receiver's PLL. The necessary modifications of (6-34) for a second-order loop follow those given in Section III-C5 for the DAL.

One problem which is becoming of increasing importance in the study of two-way systems (particularly in the case of S/X-band transmission, where the transponder gain G is on the order of four) is the uplink command modulation feedthrough, which causes suppression of the downlink carrier. This problem has been investigated in Refs. 6-4 and 6-5 with respect to its effect on Mariner transponders operating with two-channel command modulation such as on the Mariner Venus Mercury 1973 mission and on the Mariner Jupiter Saturn 1977 mission, which will use a single-channel type of command modulation. In both of these cases, the spacecraft receiver employs a PLL for tracking the carrier, thus, as shown in Ref. 6-4, knowledge of the closed-loop transfer function of the PLL is sufficient to compute the downlink carrier suppression caused by the modulation components which are tracked by the PLL.

In the two-way system proposed here, the spacecraft receiver employs a DAL, for which the calculation of the downlink carrier suppression due to command modulation feedthrough requires a different technique. As discussed in Section III-B, the input to the VCO is $z(t)$, as given by (3-4), and which can be rewritten as

$$\begin{aligned} z(t) = {} & K_2 K_{2m} F(p) \exp(-pT_s) \sqrt{P_{D_{11}}} \{[d(t)\hat{d}(t) \\ & - \operatorname{erf}(\sqrt{R_{s_{11}}} \cos \phi_1)] \sin \phi_1\} + K_2 K_{2m} F(p) \\ & \times \exp(-pT_s) \{\sqrt{P_{D_{11}}} \operatorname{erf}(\sqrt{R_{s_{11}}} \cos \phi_1) \sin \phi_1 \\ & + \hat{d}(t) N_2[t, \phi_1(t)]\} \end{aligned} \quad (6\text{-}35)$$

The second and third terms of (6-35) represent the signal and noise terms previously discussed in deriving the stochastic integro-differential equation of the DAL. The first term in (6-35) represents a *self-noise* due to the digital nature of the data detection process and the possibility of errors in the data caused by the additive noise and loop phase error. Heretofore, this self-noise term has been neglected. It behooves us here to investigate the effect of its presence in the phase of the reference signal produced by the VCO, since any modulation present here appears (multiplied by the factor G) on the downlink carrier.

Based upon arguments given in Section III-B, the self-noise process

$$n_s(t) \triangleq d(t)\hat{d}(t) - \operatorname{erf}(\sqrt{R_{s_{11}}} \cos \phi_1) \quad (6\text{-}36)$$

has zero mean and variance (conditioned on ϕ_1) given by

$$\sigma^2_{n_s}(\phi_1) \triangleq E_n\{n_s^2(t)\} = 1 - \text{erf}^2(\sqrt{R_{s_{11}}}\cos\phi_1) \quad (6\text{-}37)$$

(Note that as $R_s \to \infty$, $\sigma^2_{n_s} \to 0$.) Furthermore, since $n_s(t)$ is in essence a random pulse train, which in each symbol interval takes on value $1 - \text{erf}(\sqrt{R_{s_{11}}}\cos\phi)$ with probability

$$\frac{1 + \text{erf}(\sqrt{R_{s_{11}}}\cos\phi_1)}{2}$$

and value $-1 - \text{erf}(\sqrt{R_{s_{11}}}\cos\phi_1)$ with probability

$$\frac{1 - \text{erf}(\sqrt{R_{s_{11}}}\cos\phi_1)}{2}$$

its *power spectral density* can be calculated based upon the techniques presented in Ref. 6-6. Denoting this density (conditioned on ϕ_1) by $S_{n_s}(f;\phi_1)$, we obtain

$$\begin{aligned} S_{n_s}(f,\phi_1) = \frac{1}{T_s^2}\sum_{n=-\infty}^{\infty} &\left| pS_1\left(\frac{n}{T_s},\phi_1\right)\right. \\ &\left. + (1-p)S_2\left(\frac{n}{T_s},\phi_1\right)\right|^2 \delta\left(f - \frac{n}{T_s}\right) \\ &+ \frac{1}{T_s}p(1-p)\left|S_1(f;\phi_1) - S_2(f;\phi_1)\right|^2 \end{aligned} \quad (6\text{-}38)$$

where $\delta(x)$ is the Dirac delta function and

$$\begin{aligned} p &\triangleq \frac{1}{2}[1 + \text{erf}(\sqrt{R_{s_{11}}}\cos\phi_1)] \\ S_1(f,\phi_1) &= [1 - \text{erf}(\sqrt{R_{s_{11}}}\cos\phi_1)]\,T_s \\ &\quad \times \exp(-j\pi fT_s)\left[\frac{\sin\pi fT_s}{\pi fT_s}\right] \\ S_2(f,\phi_1) &= [-1 - \text{erf}(\sqrt{R_{s_{11}}}\cos\phi_1)]\,T_s \\ &\quad \times \exp(-j\pi fT_s)\left[\frac{\sin\pi fT_s}{\pi fT_s}\right] \end{aligned} \quad (6\text{-}39)$$

Substituting (6-39) into (6-38) gives, upon simplification,

$$S_{n_s}(f;\phi_1) = T_s[1 - \text{erf}^2(\sqrt{R_{s_{11}}}\cos\phi_1)]\left[\frac{\sin\pi fT_s}{\pi fT_s}\right]^2 \quad (6\text{-}40)$$

Note that

$$\int_{-\infty}^{\infty} S_{n_s}(f,\phi_1)\,df = 1 - \text{erf}^2(\sqrt{R_{s_{11}}}\cos\phi_1) \quad (6\text{-}41)$$

which agrees with (6-37) The self-noise process is passed through the loop filter, with the transfer function given by (3-46) Since the bandwidth of this filter is quite narrow relative to that of the self-noise power spectral density $S_{n_s}(f;\phi_1)$, we can model $F(p)$ as a rectangular low-pass filter (LPF) of unit height and width $1/\tau_1$ Then, at the input to the VCO, the self-noise appears as a smooth, continuous type of random modulation given by

$$n_s'(t) \triangleq K_2 K_{2m}\sqrt{P_{D_{11}}}\sin\phi_1\,[F(p)n_s(t - T_s)] \quad (6\text{-}42)$$

with approximate mean-squared value (conditioned on ϕ_1)

$$\begin{aligned} \sigma^2_{n_s'}(\phi_1) &\triangleq E\{n_s'^2(t)\} \\ &= (K_2 K_{2m}\sqrt{P_{D_{11}}}\sin\phi_1)^2\left[\frac{1 - \text{erf}^2(\sqrt{R_{s_{11}}}\cos\phi_1)}{\tau_1}\right] \end{aligned} \quad (6\text{-}43)$$

Recalling that the VCO is characterized by (3-14), then $n_s'(t)$ appears as a frequency modulation on the downlink carrier Since the rms value of $n_s'(t)$ is quite small, the frequency deviation of the downlink carrier is very small, and hence, the frequency spectrum of the modulated downlink carrier, due to the presence of the self-noise, is approximately given by

$$S_s(f,\phi_1) = \frac{\pi W(0)}{[\pi W(0)/2]^2 + (2\pi f - G\omega_C)^2} \quad (6\text{-}44)$$

where

$$W(0) \triangleq K_2^2 K_{2m}^2 K_V^2 P_{D_{11}} G^2 \sin^2\phi_1\,[1 - \text{erf}^2(\sqrt{R_{s_{11}}}\cos\phi_1)] \quad (6\text{-}45)$$

The spectrum in (6-44), which results whether the modulating self-noise $n_s'(t)$ is gaussian or not, has the bell-like shape of a Witch of Agnesi centered on the downlink carrier frequency $G\omega_C$, with a width $\pi W(0)$ between 3-dB points. Note from (6-44) that the spectrum of the downlink modulated carrier does not possess a discrete carrier component, but instead contains a continuum of values in a narrow band around the downlink carrier frequency. This phenomenon, namely, the absence of a discrete carrier component in the modulated carrier spectrum, is not unlike the effect on the downlink carrier produced by the spacecraft transmitter instability itself, i e. the downlink carrier is never a pure line spectrum. However, as long as the bandwidth of the modulated downlink carrier around $G\omega_C$ is small (relative to the bandwidth of the ground receiver's tracking loop W_{L2}), then the effective carrier suppression produced by the self-noise of the spacecraft tracking loop will be small In fact, we may define the

effective carrier suppression on the downlink due to the self-noise $n_s'(t)$ by

$$\eta_C(\phi_1) \triangleq \frac{\int_{-\infty}^{\infty} S_s(f,\phi)\,df - 2\int_{G\omega_C/2\pi+W_{L2}/2}^{\infty} S_s(f,\phi)\,df}{\int_{-\infty}^{\infty} S_s(f;\phi)\,df} \tag{6-46}$$

Substituting (6-44) into (6-46) gives

$$\boxed{\eta_C(\phi_1) = \frac{2}{\pi}\tan^{-1}\left(\frac{2W_{L2}}{W(0)}\right) = 1 - \frac{2}{\pi}\tan^{-1}\left(\frac{W(0)}{2W_{L2}}\right)} \tag{6-47}$$

The actual value of carrier suppression would be obtained by averaging (6-47) over the p.d.f of ϕ_1

For large loop SNR, we can approximate $\sin\phi_1$ by ϕ_1 and $\cos\phi_1$ by 1. Then, making use of (3-23) and (3-33), and noting that for small x, $\tan^{-1}x \cong x$, we get

$$\begin{aligned}\eta_C \triangleq \overline{\eta_C(\phi_1)} &\cong 1 - \frac{K^2G^2P_{D_{11}}}{\pi W_{L2}}\,\overline{\phi_1^2}\,[1 - \mathrm{erf}^2(\sqrt{R_{s_{11}}})] \\ &\cong 1 - \frac{K^2G^2P_{D_{11}}}{\pi\rho_1 W_{L2}}\,[1 - \mathrm{erf}^2(\sqrt{R_{s_{11}}})] \\ &= 1 - \frac{K^2G^2N_0}{2\pi}\left(\frac{W_{L1}}{W_{L2}}\right)\left[\frac{1 - \mathrm{erf}^2(\sqrt{R_{s_{11}}})}{\mathrm{erf}^2(\sqrt{R_{s_{11}}})}\right]\end{aligned} \tag{6-48}$$

where $K \triangleq K_2K_{2m}K_V$, and the overbar denotes statistical average with respect to the p.d.f. of ϕ_1. For small P_e^s, Eq. (6-48) can be further approximated by

$$\eta_C \cong 1 - \frac{2}{\pi}\left(\frac{W_{L1}}{W_{L2}}\right)K^2G^2N_0P_e^s \tag{6-49}$$

As an example of how small the carrier suppression is, consider the Mariner Jupiter Saturn example cited in Ref. 6-4, with the minimum operating conditions of $K \cong 4\times10^7$ and $P_e^s = 10^{-5}$, together with $N_0 \cong -167$ dBm/Hz, $G = 880/221$, and $W_{L2}/W_{L1} \cong 1$. Substituting these numerical values into (6-49) gives

$$\eta_C \cong 1 - 10^{-8} \tag{6-50}$$

In contrast to the worst-case condition for the discrete carrier system (carrier suppression = 0 83), which occurs for strong signal conditions, the worst-case condition for the DAL receiver occurs at system threshold, and, as can be seen, is totally negligible. For strong signal conditions with the DAL, $P_e^s \to 0$ and therefore, $\eta_C \to 1$. This supports our earlier claim that the DAR receiver does not exhibit any significant feedthrough problems

In conclusion, we remind the reader that we have considered only one of the three modulation–demodulation alternatives cited in the beginning of this subsection. Extension of the foregoing results to the other two alternatives is straightforward.

References

6-1 Lindsey, W C., and Simon, M. K., "Carrier Synchronization and Detection of Polyphase Signals," *IEEE Trans Comm. Technol.*, Vol. COM-20, No. 3. June 1972, pp 441–454

6-2 Lindsey, W. C., *Synchronization Systems in Communication and Control,* Prentice Hall, Inc , Englewood Cliffs, N.J., 1972.

6-3 Chadwick, H. D., "Binary Single-Sideband Phase Modulated Communication Systems," *IEEE Trans Information Theory,* Vol. IT-18, No. 1, Jan. 1972, pp. 214–215

6-4 Wood, G E., *Command Modulation Feedthru in Mariner Transponders,* Interoffice Memorandum 3396-72-361, October 16, 1972 (JPL internal document).

6-5 Wood, G. E , *Command Feedthru in the MJS 77 RFS,* Interoffice Memorandum 3396-73-047, Feb. 6, 1973 (JPL internal document).

6-6 Titsworth, R. C., and Welch, L. R., *Power Spectra of Signals Modulated by Random and Pseudorandom Sequences,* Technical Report 32-140, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 1961.

VII. Summary and Conclusions

In summary, we have presented an efficient means for coherently demodulating a suppressed carrier signal which has been modulated by digital data. The data may either be in the form of binary symbols (or bits), in which case they are transmitted as biphase modulation of the carrier, or they may be N-level information, in which case the carrier is polyphase-modulated. In either situation, the carrier-tracking loop, which reconstructs the demodulation reference signals from the received signal in additive white gaussian noise, employs the principle of decision-directed feedback; i e., the operation of the loop depends upon the data modulation detection decisions (see Table II-2). This dependence greatly enhances the carrier-tracking ability of a loop which is required to track a data-modulated suppressed carrier. Clearly, an increase in the efficiency of carrier tracking manifests itself as an improvement in the area of radio link-derived navigation, and in the performance of the data detection, since the reference signals supplied by the loop contain less noise

From the theoretical and experimental results presented in this report, we are able to conclude that, for a wide variety of applications, the concept of suppressed carrier transmission together with a data-aided tracking receiver is generally more efficient than its discrete carrier counterpart which employs a PLL Despite this, we must proceed with some caution since a suppressed carrier data-aided tracking receiver does possess certain anomalies relative to the tracking receivers which are employed in present communication system designs. Furthermore, the operational experience with discrete carrier systems cannot be discounted. We shall review these anomalies here, comment on their significance, and discuss how their potential detrimental effect may be reduced by the techniques presented in the text.

In the first case, suppressed carrier binary transmission (as opposed to discrete carrier binary transmission) is inherently accompanied by a 180-deg phase ambiguity in the true system lock point This fact is true irrespective of whether a DAL or any other type of suppressed carrier tracking loop is employed. The significance of this ambiguity is with respect to its effect on the data estimates emanating from the data detector output With probability 1/2, the data estimates may be sign inverted relative to their transmitted sense This problem can be resolved by transmitting a known synchronizing sequence (and recognizing it), or by differentially encoding the data (which results in a small penalty of about a factor of 2 in average symbol error probability performance).

Secondly, the data-aided tracking loop incorporates a delay element which can potentially cause instability problems and which reduces the acquisition range of the loop. This delay also increases the loop bandwidth and natural frequency of operation relative to that which would exist if the delay were zero. These problems are discussed in Sections III-D and III-E From the results given there, we conclude that the problems introduced by the presence of the delay element are significant only for symbol rates comparable in magnitude to the loop bandwidth Furthermore, for such low data-rate applications, and sufficiently small symbol error probability, one can employ subsymbol integration (Section VII-F), which, insofar as the tracking loop is concerned, increases the effective symbol rate by a factor of n, where n is the number of subsymbols per symbol. This minimizes the delay and therefore its related problems

Another seemingly potential problem is the effect of the modulation feedthrough in two-way links employing a DAL, caused by the loop's dependence on the data modulation As it turns out, the magnitude of this feedthrough is essentially negligible relative to the values characteristic of a discrete carrier system (Section VII-D). Thus, this heretofore unknown effect actually is an advantage in favor of the DAL.

We now summarize some of the salient advantages of data-aided receivers (DARs) To begin with, in the DAR, no subcarrier is required for single-channel modulation since the data may be modulated directly onto the carrier. This results in a considerable savings in receiver hardware because a subcarrier tracking loop is no longer necessary Furthermore, since the DAL requires symbol (or bit) synchronization, integration, and decision, this processing, which has heretofore been separate or apart from the carrier tracking function, now becomes an integral part of the receiver

The operating point of the DAL (as measured by loop SNR, etc.) does not depend on the structure of the data

symbols, and works equally well if long strings of ones or zeros, or a random mixture is transmitted. Nor does the DAL require any special acquisition prefix or sequence or comparative threshold decisions to achieve phase lock. (We have already noted the special problem of resolving the carrier phase ambiguity.)

Since the proposed mechanization of the DAR involves sampled-data digital processing, data rate switching is easily accommodated. Sections IV and V deal extensively with implementation, including the important topics of automatic gain control and lock detection

Finally, the section on complete system configurations (VI), gives considerable insight into how the basic DAL can be integrated into present receiver designs. In addition, a number of common applications are discussed, and some suggestions are made concerning how the DAL may be used for bandwidth compressive modulations such as polyphase and single-sideband

Nomenclature

A	first level of quantizer output
$d(t)$	data stream with ± 1 symbols
$\hat{d}(t)$	data detector's estimate of $d(t)$
$\tilde{d}(t)$	Hilbert transform of $d(t)$
$F(s)$	transfer function of loop filter
F_1	ratio of τ_2 to τ_1
G	transponder turnaround ratio
$G(\phi)$	difference between the loop S-curve and its statistical mean with respect to $p(\phi)$
G_D	ratio of η_{SC} to η_{DC} (data signal improvement of the suppressed carrier system to the discrete carrier system)
G_{TR}	ratio of ρ_{SC} to ρ_{DC} (tracking loop improvement of suppressed carrier-tracking loops to discrete carrier-tracking loops)
$H(s)$	closed-loop transfer function
K	number of bits of quantization in ADC; product of K_2, K_{2m}, and K_V (total open-loop gain)
$K(V)$	gain of controllable IF amplifier in AGC loop
K_L	a shape factor determined by transfer function of the bandpass filter in squaring and Costas loops
K_V	VCO gain
K_{1m}, K_{2m}	In-phase and quadrature phase detector gains
K_1^2, K_2^2	average power of the in-phase and quadrature reference signals provided by the DAL
L_{ADC}	quantizer loss function for ADC
L_R	efficiency loss associated with the carrier regeneration process in suppressed carrier systems
M_A	number of symbols of accumulation in forming lock detector signal
$n_1(t)$	additive input noise (narrowband gaussian process)
$n_s(t)$	self-noise process due to the digital nature of the data detection process
N_0	noise spectral density, W/Hz
$N_1[t, \phi(t)]$, $N_2[t, \phi(t)]$	low-pass noise processes at the in-phase and quadrature phase detector outputs of the DAL
p	Heaviside operator
$p(\phi)$	phase error probability density function
P_c	probability of correctly determining loop's lock condition
P_C	discrete carrier power, W
P_D	data signal power, W
P_T	total transmitter power, W
P_e^s	system average symbol error probability
r	loop damping factor (equal to four times the square of the true loop damping)
$r(t)$	carrier reference signal produced by carrier regeneration loop
R_s	data signal-to-noise ratio
$\mathcal{R}_b$	system bit rate, bits/s
$\mathcal{R}_s$	data symbol rate, symbols/s
$s(t)$	transmitted signal
$\bar{S}$	average rate of cycle slipping, slips/s
$S_{n_s}(f, \phi_1)$	power spectral density of $n_s(t)$ conditioned on ϕ_1
$\mathcal{S}(\phi)$	loop S-curve (effective loop nonlinearity)
$Sq(\phi)$	a unit power square wave in ϕ with period 2π
$S(t)$	unit power square-wave subcarrier of angular frequency ω_{SC}
T_d	arbitrary loop delay
T_s	reciprocal of $\mathcal{R}_s$ (data symbol period), s
$U_0(\phi)$	loop potential function
V	first threshold of quantizer input, also AGC voltage
V'	threshold for testing normalized lock detector signal V_d^N

Nomenclature (contd)

V_d	lock detector signal
V_Q	quadrature counterpart of AGC voltage used in lock detection
V_d^N	normalized lock detector signal
W_i	input IF filter bandwidth, Hz
W_L	two-sided loop bandwidth, Hz
$x(t)$	total received signal
y	ratio of the design point loop bandwidth to the input IF filter bandwidth
$z(t)$	input signal to VCO
α, β	parameters characterizing the phase error probability density function $p(\phi)$
$\widetilde{\alpha}_1$	bandpass limiter suppression factor
$\gamma(\omega_f)$	synchronization boundary of loop
Γ_p	limiter performance factor affecting loop signal-to-noise ratio
δ	reciprocal of the single-sided loop bandwidth–symbol time product
ϵ	epoch of transmitted data stream, s
$\hat{\epsilon}$	symbol synchronizer's estimate of ϵ
$\epsilon(t)$	phase detector output
η_c	critical value of normalized loop delay determining absolute stability region
η_C	average of $\eta_C(\phi_1)$ over the probability density function of ϕ_1
$\eta_C(\phi_1)$	effective carrier suppression on the downlink due to the self-noise process (conditioned on ϕ_1)
η_{DC}	efficiency of discrete carrier system
η_{SC}	efficiency of suppressed carrier system
$\theta(t)$	phase shift added to transmitted signal by the channel characteristic
θ_m	data signal modulation index
θ_0	random phase component of $\theta(t)$
μ	ratio of $\widetilde{\alpha}_1$ at the design point to its value at the operating point
μ_V	mean of AGC voltage
μ_V^N	normalized mean of AGC voltage
$\mu_{V_d^N}, \sigma^2_{V_d^N}$	mean and variance of V_d^N
ρ_{ADC}	input signal-to-noise ratio to the ADC
ρ_{DC}	loop signal-to-noise ratio for the discrete carrier-tracking loop
ρ_i	signal-to-noise ratio in the noise bandwidth of the input bandpass filter
ρ_{SC}	loop signal-to-noise ratio for the suppressed carrier-tracking loop
σ_ϕ^2	variance of the phase error
$\sigma^2_{n_s}(\phi_1)$	variance of $n_s(t)$ conditioned on ϕ_1
τ_1, τ_2	loop filter time constants
Υ_0	design point of loop preceded by a bandpass limiter
$\phi(t)$	loop phase error
$\Phi(t)$	total phase of received carrier
$\hat{\Phi}(t)$	carrier regeneration loop's estimate of $\Phi(t)$
ϕ_{ss}	steady-state phase error (static phase error)
ω_C	carrier angular frequency, rad/s
ω_f	fundamental beat note in a Fourier expansion of $\phi(t)$
ω_{SC}	subcarrier angular frequency, rad/s
Ω_0	doppler frequency offset, rad/s

INDEX

INDEX (contd)

NASA — JPL — Coml , L A , Calif