UNCLASSIFIED

AD NUMBER

AD921755

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited. Document partially illegible.

FROM:

Distribution authorized to U.S. Gov't. agencies only; Test and Evaluation; 04 MAR 1973. Other requests shall be referred to Air Force Space and Missile Systems Organization, Attn: YEEU, Norton AFB, CA. Document partially illegible.

AUTHORITY

samso, usaf ltr, 17 jun 1977

THIS PAGE IS UNCLASSIFIED

THIS REPORT HAS BEEN DELIMITED
AND CLEARED FOR PUBLIC RELEASE
UNDER DOD DIRECTIVE 5200.20 AND
NO RESTRICTIONS ARE IMPOSED UPON
ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

AD NO.
DDC FILE COPY

WDL-TR5291
28 February 1974

SAMSO TR 74-183

AD921755

GLOBAL POSITIONING SYSTEM (GPS) FINAL REPORT

[redacted]

PART II

VOLUME A
System Trades and Analyses

Contract F04701-73-C-0296

Submitted to:
DEPARTMENT OF THE AIR FORCE
HEADQUARTERS SPACE AND MISSILE SYSTEMS ORGANIZATION (AFSC)
P.O. Box 92960, Worldway Postal Center
Los Angeles, California 90009



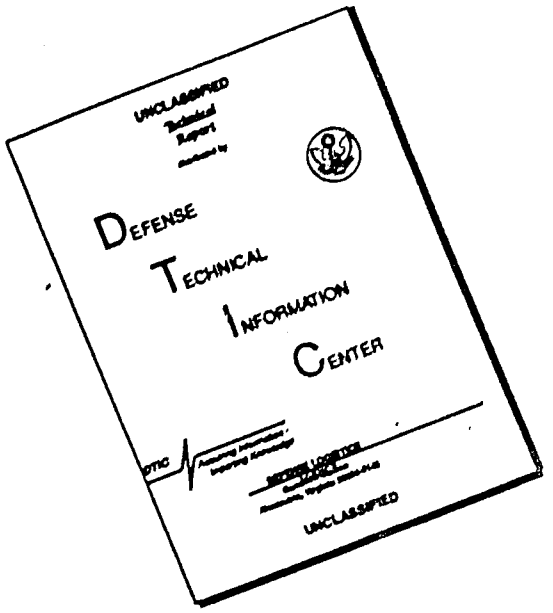
TRW SYSTEMS GROUP

PHILCO Ford
Philco-Ford Corporation
Western Development Laboratories Division
Palo Alto, California 94303

CONTROL DATA CORPORATION

[redacted]

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.


ACCESSION for
NTIS
DDC
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY
Dist.
B

Publication approval of this document does not constitute Department of Air Force approval of the presented findings or conclusions unless so designated by other authorized documents. It is published only for the exchange and stimulation of ideas.

SAMSO TR 74-183

WDL Technical Report 5291
28 February 1974

GLOBAL POSITIONING SYSTEM (GPS)
FINAL REPORT

PART II - VOLUME A
SYSTEM TRADES AND ANALYSES

Contract F04701-73-C-0296

D D C
RECEIVED
AUG 22
DETROIT

Prepared for

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS SPACE AND MISSILE SYSTEMS ORGANIZATION (AFSC)
Los Angeles, California 90009

ATTENTION: YEEM

DISTRIBUTION STATEMENT B

Distribution limited to U.S. Government agencies only; Test and Evaluation; 4 March 1973. Other requests for this document must be referred to SAMSO (YEEU).

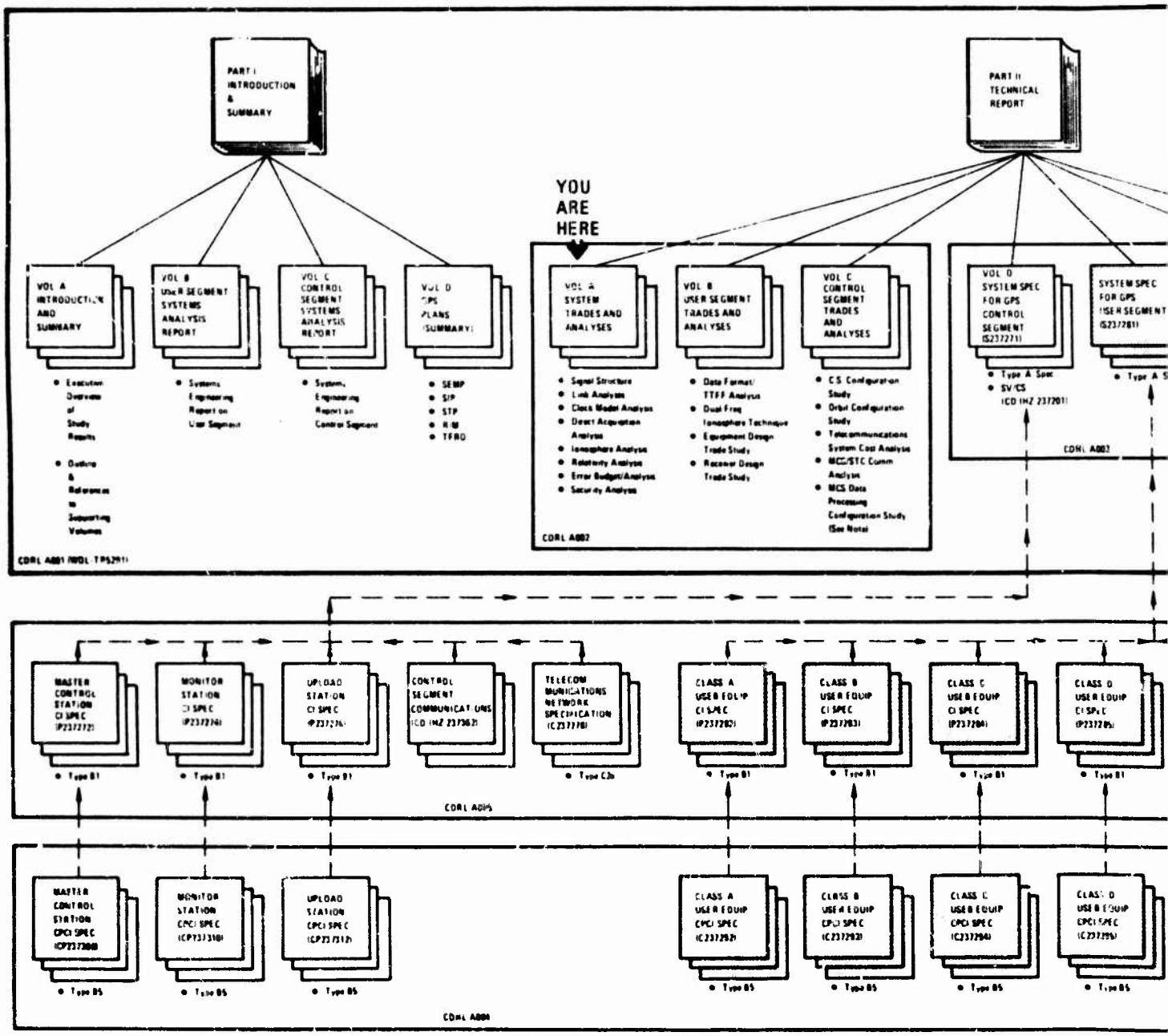
Western Development Laboratories Division
Palo Alto, California 94303

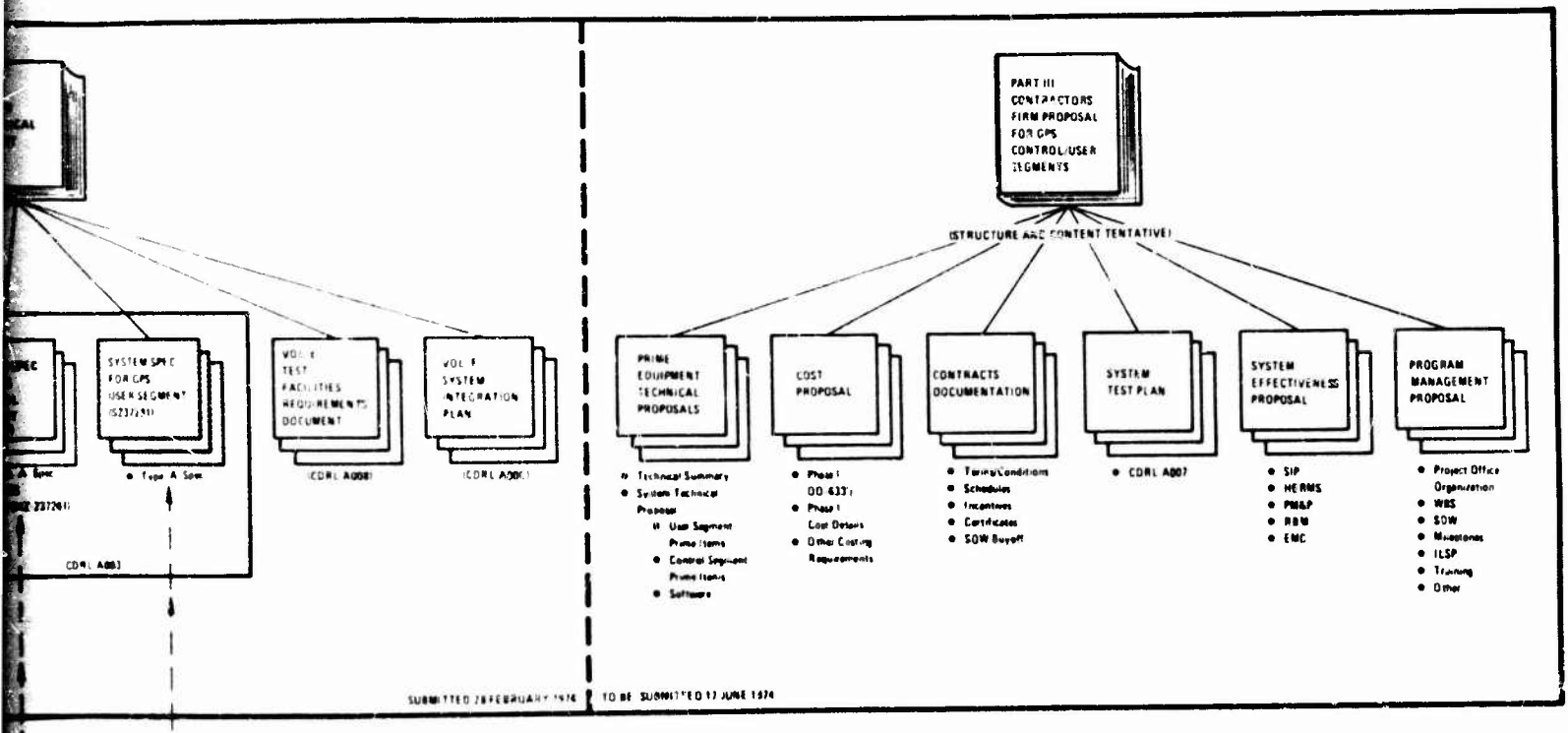
PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

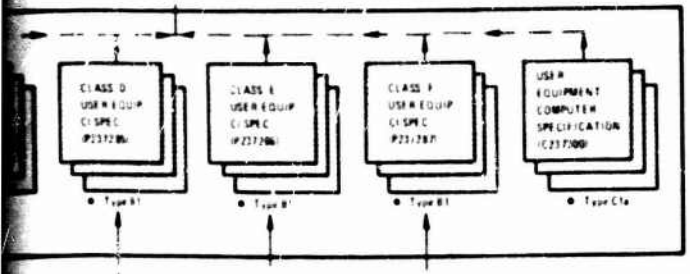
FOREWORD

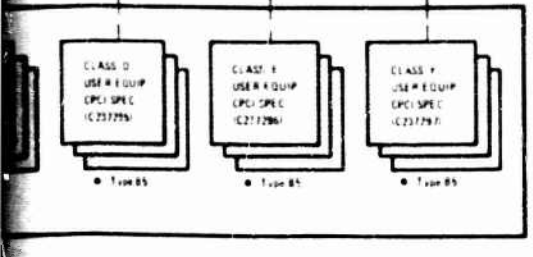
This is Part II, Volume A, of the GPS Definition Study Final Report, submitted by Philco-Ford, in accordance with Sequence Number A002 of Exhibit A to Contract F04701-73-C-0296. The period of performance for the report submitted here is from 28 June 1973 to 28 February 1974. The following figure identifies the structure of the Final Report and the relationship of this volume to the other volumes in this submittal.

WDL-TR5291
Part II
Volume A











INDICATES 30 JAN SUBMITTAL
(NOT INCLUDED IN THIS SUBMITTAL)

NOTE

ADD THESE ITEMS TO LIST

- WS Data Processing Configuration Study
- Reference Ephemeris Data Processing Cost Analysis
- Ephemeris Determination Analysis
- Signal Power Monitoring Techniques
- Ephemeris and Clocks Processing Simulations

Structure of Global Positioning System Definition Study Final Report

CONTENTS

REPORT NO.

1	GPS Signal Structure Analysis
2	GPS Link Analysis
3	Modeling Stable Clocks in the GPS
4	Direct Acquisition Analysis
5	Ionospheric Model Analysis
6	Relativistic Clock Errors
7	Security of Navigation Signal
8	Error Budget Analysis

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

Report No. 1

GPS SIGNAL STRUCTURE ANALYSIS

CONTENTS

SECTION		PAGE
1.0	Introduction	1-1
		1-1
2.0	Navigation Signal Performance	1-1
	2.1 Acquisition of Signal	1-1
	2.1.1 C/A Signal Acquisition	1-2
	2.1.2 Time-To-First Fix for Four Channel Receiver	1-8
	2.1.3 TTFF for the Sequential Receiver	1-19
3.0	C/A Signal Multiple Access Performance	1-26
	3.1 Properties of Gold Codes	1-27
	3.2 Multiple Access Performance	1-37
	3.2.1 Basic Correlation Receiver Model	1-37
	3.2.2 Acquisition Performance	1-40
	3.2.3 Noncoherent DLL Tracking Performance	1-44
	3.2.4 Coherent DLL Tracking Performance	1-47
	3.2.5 Coherent Data Detection	1-51
	3.2.6 Summary of Multiple-Access Performance	1-54
4.0	C-Signal Jamming Performance	1-57
	4.1 Sinewave Jammer	1-57
	4.2 Wideband Jammer	1-59
5.0	P-Signal Jamming Performance	1-62
	References	1-66

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	1023 Gold Code Spectrum	1-28
2	Sideband Cross-Correlation Levels for 511 Gold Codes	1-30
3	Sideband Cross-Correlation Levels for 1023 Gold Codes	1-31
4	Sideband Cross-Correlation Levels for 2047 Gold Codes	1-32
5	Cumulative Probability of Interference Level for 511 Gold Codes	1-33
6	Cumulative Probability of Interference Level for Gold Codes	1-34
7	Cumulative Probability of Interference Level for 2047 Gold Codes	1-35
8	First and Second Sideband Cross-Correlation for 1023 Gold Codes	1-36
9	Relative Dwell Time Vs Signal-To-Multiple Access-Interference Ratio	1-45
10	Multiple-Access Performance of Narrowband Noncoherent DLL	1-48
11	Multiple-Access Performance of Narrowband Coherent DLL	1-50
12	Multiple-Access Performance of Coherent Data Detector	1-53
13	Data Degradation Due to Sidewave Jammer	1-60
14	Regained Satellite Transmitter Power Vs Jamming Threat	1-64
15	Jamming Power	1-65

TABLES

TABLE		PAGE
1	Acquisition Frequency Prepositioning Errors	1-4
2	C-Code Acquisition Parameters	1-6
3	Start-up for C-Navigation	1-9
4	TTFF at Threshold, Four Channel Receiver, C-Navigation	1-10
5	Start-up for P-Navigation	1-17
6	TTFF at Threshold, Four Channel Receiver, L_1/P-Navigation	1-18
7	TTFF at Threshold, Four Channel Receiver, L_2/P-Navigation	1-18
8	Initial Steps for Sequential C-Navigation	1-20
9	TTFF at Threshold, Sequential C-Navigation	1-21
10	Initial Steps for Sequential L_1/P-Navigation	1-22
11	TTFF at Threshold, Sequential L_1/P-Navigation	1-23
12	Pseudo-Range, Range-Rate Tracking Errors	1-24
13	Data Demodulation Requirements	1-25
14	Cross-Correlation Properties of Gold Codes	1-27
15	S/I for 90%, 95% and 99% Probabilities	1-55
16	S/I_T for 90%, 95% and 99% Probabilities	1-56
17	Performance Comparison for 95% Probability	1-57

GPS SIGNAL STRUCTURE ANALYSIS

1.0 INTRODUCTION

In this section the recommended signal structure described in Appendix II of the "System Specification for the GPS" USAF/SAMSO SS-GPS-101A, 29 January, 1974 and in the GPS SV/MON ICD (Vol. E) is analyzed. The performance of alternate C/A code structures of different lengths are also analyzed.

Section 2 analyzes the acquisition performance of the C/A signal, and summaries the range and range-rate accuracy and data reception performance of both C/A and P signals.

Section 3 presents the analysis of the multiple access performance of the C/A signal.

Section 4 presents the predicted performance of the system in a jamming environment.

2.0 NAVIGATION SIGNAL PERFORMANCE

In this section the expected performance of the proposed C/A - and P-signal structures is presented. Section 2.1 presents the steps required to acquire both the C/A - and P-signals and the resulting time-to-first fix (TTFF). In section 2.2 the expected range and range-rate accuracies are summarized.* The data performance requirements are established in section 2.3.

2.1 Acquisition of Navigation Signal

The normal mode of acquiring the P-signal is by first acquiring the C/A-signal and then switching to the P-signal. Direct P-signal acquisition or reacquisition is not discussed in this study (See Direct Acquisition Trade Study, Report #4 of this volume.

* Detailed analysis of range and range rate accuracies is found in Report No. 7.

2.1.1 C/A-Signal Acquisition

The acquisition procedure for the C/A-code signal consists of the following steps at the user:

a. C/A-Code Selection - The C/A-code generator in the receiver is configured to generate the Gold code corresponding to the particular satellite signal chosen to be acquired.

b. Estimate of Frequency of Received Signal - The computer estimates the carrier frequency of the satellite signal to be acquired.

c. Prepositioning of the Receiver VCO Frequency - From the frequency estimate in Step b and the user oscillator frequency estimate, the receiver VCO frequency is positioned so that the incoming signal will fall within the acquisition bandwidth of the receiver when the C/A-code phase is found. Thus, a frequency search for the signal is not necessary.

d. Search All Possible Code Chip Positions - The Gold code generator phase is stepped in one-half chip steps looking for the presence of a carrier signal in the acquisition bandwidth.

e. Stop Search, and Allow Tracking Loops to Lock to Code and Carrier Phase When Signal is Detected in Acquisition Bandwidth - This completes acquisition of the C/A-code.

The acquisition time is the total time required to carry out the above acquisition steps. Step d is by far the largest contributor to the acquisition time as the total time for the other steps is less than one sec. The following paragraphs discuss the above steps in more detail, and compute the code search time in Step d as a function of the carrier-to-noise density ratio C/N_o.

The first step in initial operations is constellation selection by the user's computer. For this the classical orbital elements of each satellite in the system (stored in non-volatile memory) and the inputs from current horizontal position and the current system time are used. The accuracies of these inputs need only be $\pm$ 60 nm and $\pm$ 1 min. for position and system time, respectively.

The user computer will then select the best constellation and consequently satellite to be acquired. "Best" will be determined by certain logical criteria which will probably vary from user class to user class.

The second step is the estimation by the user's computer of the frequencies of the received signals from each of the selected satellites. These estimates will be based on:

a. The nominal L_1 transmitter frequency, provided that the GPS is designed such that no satellite transmitter frequency is allowed to vary from this system nominal by more than 16 Hz ($1:10^8$). Alternatively, the estimate will be based on a value for the transmitter frequency of the particular satellite, which is determined by the Control Segment and included with the classical orbital elements for that satellite. In this latter case, the GPS must be designed such that this value by more than 16 Hz during the period between updates of the orbital elements. The selection of which technique should be used will be based on a Satellite/Control Segment tradeoff, since it has insignificant impact on the User Segment.

b. The doppler shift induced by the motion of the satellite relative to the user's location on the rotating Earth, (that is, the total doppler shift if the user were stationary on Earth's surface). This is computed as a by-product of the calculations for constellation selection, with a maximum possible error of 40 Hz.

c. The doppler shift induced by the motion of the user. For User Classes A and B, the course and speed will be automatically, and hence very frequently, inputted to the computer from other (non-GPS) navigation sensors. The accuracy of these data will be such that the component of user velocity in the direction of the satellite will, throughout the start-up process, be known to within 30 m/sec, corresponding to a doppler shift at L_1 of 158 Hz. For User Class C, the operator will have to input course and speed at the beginning of the start-up procedure, and then fly straight and level until acquisition has been achieved. Course will be inputted in degrees east of north to to a resolution of 1 degree, and speed in Km/hr, relative to the ground, to a resolution of 1 Km/hr. The flying constraint will ensure that the actual course and speed will differ from the inputted values by no more than 5 degrees and 30 Km/hr. Additionally, we are assuming that, for the Class C user, the maximum velocity constraints at start-up will be 300 m/sec horizontally and 6 m/sec vertically. With these constraints and inputs, the user-induced doppler shift at L_1 will, for the Class C user, be estimable to within 141 Hz, for any satellite above the 5 degree horizon.

For User Classes D, E, and F, course and speed will not be inputted, and the received signal frequency estimate will assume that the user is stationary. However, the constraints on maximum velocity for these classes will be 30 m/sec horizontal and 6 m/sec vertical, so the maximum user-induced doppler shift at L_1 cannot exceed 160 Hz.

From the above, it can be seen that the received signal frequencies will be estimated with a maximum possible error of 216 Hz or less (all user classes).

The third step is the calculation, by the user's computer, of the frequency to which the second IF VCO should be prepositioned. This is based on the previously described received signal frequency estimate and the nominal value of the user oscillator frequency. The user oscillator must be specified such that, after warm-up, its frequency deviation from nominal should not exceed $1:10^7$. Thus, the maximum error in VCO prepositioning will not exceed (216 + 158), or 374 Hz.

The receivers are being configured to have an IF acquisition bandwidth of $\pm$500 Hz, so that the signal will always lie within the passband, even under the simultaneous occurrence of all adverse errors. From the statistical point of view, we can assume that each error source is independently and uniformly distributed between its maximum values, which leads to a standard deviation of frequency error of 132 Hz, as shown in Table 1.

TABLE 1

ACQUISITION FREQUENCY PREPOSITIONING ERRORS

Error Source	Maximum Frequency Error	S. D. of Frequency Error
Knowledge of transmitter frequency	$\pm$ 16 Hz	9.2 Hz
Satellite-induced doppler shift	$\pm$ 40 Hz	23.1 Hz
User-induced doppler shift	$\pm$160 Hz	92.4 Hz
User oscillator drift	$\pm$158 Hz	91.2 Hz
All	$\pm$374 Hz	132.2 Hz

In the fourth step, the receiver is prepositioned in accordance with the above described estimates. Positioning the VCO is accomplished within msec. An acquisition search is then carried out for the C/A-code. The C/A-code generator is clocked from the VCO output (after suitable translation and multiplication). A noncoherent detection processor is used, detecting the presence of reconstructed carrier, after correlation of the received signal with the replica C/A-code (with the code tracking modulation off tau-dither). The acquisition detector status is checked every t_d sec (t_d being known as the dwell time per search cell). If the check shows that the signal is not present, the Gold code generator is stepped one-half chip; if signal presence is indicated, the tau-dither modulation is activated, and the code and carrier tracking loops are enabled. At a period kt_d later, the lock status of the (coherent) carrier tracking loop is sampled. If this shows that the loop is not in lock, it is assumed that the original acquisition indication was false, and the noncoherent search is resumed.

A considerable amount of analysis and simulation of the acquisition process has been performed. The values of t_d, k, and also of the post-detector bandwidth, must be selected to minimize the acquisition time at signal threshold. The numerical results are shown in Table 2.

In Table 2, we show the optimized design parameters (post-detection bandwidth, B_{pd}, dwell time per search cell, t_d, and false alarm penalty factor k) for various values of the threshold design point. The design point is shown in the second column of Table 2 as the required carrier-to-noise density ratio of the L_1-C signal at the acquisition detector input. As may be seen, it is 3.8 dB lower than the (C/N_o) ratio at the receiver input, to allow for

a. Out of Sync Demodulation Loss - The replica and received codes can differ in phase, during dwell on the correct search cell, by as much as 1/4 chip. This would cause, at most, a 2.5 dB correlation loss.

TABLE 2

C-CODE ACQUISITION PARAMETERS

(C/N_o) Received	(C/N_o) at Det.	B_{pd}	t_d	k	P_d	P_{fa}	$\bar{T}$	T_{SD}	T(90)
dBHz	dBHz	Hz	msec				sec	sec	sec
30.8	27	12.5	24	5	.676	.097	71.4	53.8	140.3
31.8	28	10.7	28	4	.780	.051	54.0	40.7	106.1
32.8	29	13.0	23	4	.820	.055	41.3	31.1	81.2
33.8	30	14.3	20	4	.865	.056	32.9	24.8	64.7
34.8	31	-	-	-	-	-	26.7	20.1	52.5
35.8	32	18.8	17	3	.913	.021	22.0	16.6	43.2
36.8	33	-	-	-	-	-	18.3	13.8	36.0
37.8	34	25.0	12	3	.908	.020	15.7	11.8	30.9

- These values were not computed; however, the acquisition times were derived by interpolating between the adjacent values.

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

b. Frequency Offset of the Predetection Bandpass Filter - The 3 dB bandwidth of this filter is 500 Hz (one sided), and the standard deviation of the frequency offset will be (as was earlier shown) 132 Hz. The Manchester code bit rate is 200 Hz, and hence there will be some power loss in the sidebands. This loss has been estimated to be 0.3 dB.

c. Multiple Access Loss - The effect of multiple access interference on the C/A-code acquisition has been analyzed in Section 3.0. The results show that, for a 1023 bit code and an interference level which is not exceeded 95% of the time, the degradation in performance is about 1.0 dB.

The values of B_{pd}, t_d, and k are those of the optimum set, with the predetection bandwidth set at 500 Hz. P_d and P_{fa} are the resulting performance factors at threshold (the probabilities of true detection and false detection, respectively). T, T_{SD}m abd T (90) are further performance factors at threshold, namely the mean and standard deviations of the acquisition time, and the acquisition time not exceeded 90% of the time.

The optimum design parameters were not computed for (C/N_o) values at the detector of 31 and 33 dB Hz, but the acquisition times were interpolated for completeness in the table. The values of $\bar{T}$ were computed with the analytic expressions derived in Reference 1. Based on the Monte Carlo simulations reported in Reference 1 the standard deviation was found from the relationship.

$$T_{SD} = .754\ \bar{T}$$

Also, the 90% probability value is obtained from

$$T(90) = \bar{T} + 1.28\ T_{SD}$$

From Table 2, the required C/N_o at the receiver input can be found which results in a 90% probability of acquiring the C/A-code within 50 sec. By interpolation from the table, the required value for C/N_o is 35.1 dB Hz. This value is based on a single acquisition circuit searching the entire 1023 bit code in one-hald chip increments.

Acquisition time is very nearly directly proportional to the number of possible search cells or half-chip positions. Therefore, for the same signal levels, bandwidths, dwell times, etc., a receiver using two acquisition circuits working in parallel, each searching one-half of the code, will acquire the code in half the time needed by a single acquisition circuit receiver. Equivalently, the signal levels, bandwidths, etc., which correspond to a single circuit receiver acquiring the code in 100 sec are the same as those which correspond to a 2-circuit receiver acquiring in 50 sec. Thus, again by interpolating in Table 2, we see that the required C/N_o at the receiver input, for 90% probability of acquisition within 50 sec for a 2-circuit receiver, is reduced to 32.0 dB Hz.

Based on a worst case receiver noise density of -199.6 dBw/Hz, the required signal levels at the PN receiver input for 50 sec acquisition time are:

-164.5 dBw (single acquisition circuit)
-167.6 dBw (two parallel acquisition circuits)

2.1.2 Time-To-First Fix (TTFF) for Four Channel Receivers

An important performance parameter of the GPS user receiver is the time required from "power-on" to the first determination of user position and velocity. This is called the "time-to-first fix" (TTFF).

The time-to-first fix (TTFF) includes, in addition to the C/A-code acquisition time, discussed in the preceding section, the steps discussed in the following paragraphs. The discussion assumes four channel receivers in all cases. Sequential receivers are discussed in the next section.

When the C/A-code search is successful, the carrier and code loops are enabled and track the signal, and the Manchester code and bit synchronizer lock-up. The time for loop pull-in and settling, and synchronizer lock-up, is less than 1 sec, and is an insignificant contributor to the TTFF.

The following applies to only L_1-C navigation. When the tracking loops, bit synchronizers, etc., have locked up, measurements of pseudorange and pseudorange rate are made and sent to the computer. However, they are not used there until the satellite ephemeris and clock data are received too. For this, the receiver performs a real time search for main frame sync. Since the frame duration is 30 sec, the wait time for frame sync is uniformly distributed between 0 and 30 sec. The mean value is 15 sec and the standard deviation is 8.7 sec.

We assume that all the normal navigation data will be contained within the first 3/5 of the frame, that is, within the first 720 bits. Thus, 18 seconds after the beginning of the main frame sync is received, the first fix can be made. The computation takes an insignificant amount of time.

To recapitulate, the significantly time-consuming steps in the start-up process are as shown in Table 3.

TABLE 3

START-UP FOR C-NAVIGATION

Step	Mean Time (sec)	S. D. of Time (sec)
C/A-code search	See Table 2	
Wait for frame sync	15.0	8.7
Collect navigation data	18.0	0

From any tabulation of the normal error function, it can be seen that

$$T_{90} = \text{mean} + 1.94 \text{ (standard deviation)}$$

$$T_{90} = (m + 33) + 1.94(\sigma^2 + 75)^{\frac{1}{2}}$$

Based on the values in Tables 2 and 6, and using the above expression, we can find the mean and 90% confidence values of the TTFF, at threshold, as a function of threshold, where threshold is given in terms of the received carrier-to-noise density ratio. These values are shown in Table 4.

TABLE 4

TTFF AT THRESHOLD, FOUR CHANNEL RECEIVER, C-NAVIGATION

Threshold (C/N_o) -Received dBHz	TTFF (sec)	
	Mean	90% Confidence (T_{90})
30.8	104.4	210.1
31.8	87.0	167.7
32.8	74.3	136.9
33.8	65.9	116.9
34.8	59.7	102.2
35.8	55.0	91.3
36.8	51.3	82.9
37.8	48.7	77.1

Let the mean and standard deviation of the C/A-code search be m and σ. Then by invoking the Central Limit Theorem, we can say that the time for any one channel to complete all its initial operations is normally distributed with a mean value of (m + 15 + 18), that is (m + 33), and with a standard deviation of $(\sigma^2 + 8.7^2)^{\frac{1}{2}}$. These same statistics apply to each of the 4 channels, but the variates are independent.

The time to first fix is the time for the last channel to complete its initial operations. Hence, the mean TTFF is the mean for any one channel, or (m + 33). Of greater interest than the mean, however, is the TTFF not exceeded on some stated percentage of start-ups. We here adopt 90% as this confidence level, and call the time, T_{90}. Then,

$$\text{Prob (all channels complete by } T_{90}) = 0.9$$

$$\text{or} \quad [\text{Prob (any one channel complete by } T_{90})]^4 = 0.9$$

$$\text{i.e., Prob (any one channel complete by } T_{90}) = 0.974$$

The following applies only to four channel L_1-P navigation. There are, conceptually, three possible techniques which follow acquisition of the C-code, as follows:

a. Wait for C-data main frame sync, collect the navigation data from C, then handover to P.

b. Wait for C-data main frame sync, handover to P, wait for P-data main frame sync, then collect the navigation data from P.

c. Wait for next C-data sub-frame sync, handover to P, wait for P-data main frame sync, then collect the navigation data from P.

d. Collect data while waiting for frame sync.

The requirements on the signal structure for each of these techniques are:

a. One HOW per frame, located immediately after the navigation data. Also C- and P-data may be transmitted in phase.

b. One HOW per frame, located immediately after the main frame sync. Also P-data delayed slightly behind C-data.

c. Multiple HOWs in each frame, each preceded by a sync pattern. Also, P-data delayed slightly behind C-data.

It can be shown that the TTFF will be the same in each of these techniques. Accordingly, the best technique should be selected on other grounds. Clearly, technique (c), with multiple HOWs is least desirable, requiring additional complexity in both spacecraft and user equipment; indeed, the existence of multiple handover words can, itself force an extension to the frame length, which will increase the TTFF. Technique (b) is slightly more complex, in the spacecraft, than technique (c), because it requires a phase delay between the two data signals.

To demonstrate the equivalance of the TTFFs, consider the following:

Let frame duration = T

duration from beginning of frame to end of navigation data = T_n

duration from beginning of sync pattern preceding a HOW to accomplishment of the handover = T_h

Then, for technique (a), we have

- Wait for C-data main frame sync: uniformly distributed between zero and T

  mean $\frac{T}{2}$

  S. D. $\frac{T}{\sqrt{12}}$

- Collect navigation data

  mean T_n

  S. D. 0

- Handover to P: (HOW follows immediately behind navigation data)

  mean T_h

  S. D. 0

- Time for all above operations

$$\text{mean} = \frac{T}{2} + T_n + T_h$$

$$\text{S. D.} = T/\sqrt{12}$$

For technique (b), we have

- Wait for C-data main frame sync

$$\text{mean} = T/2$$

$$\text{S. D.} = T/\sqrt{12}$$

- Handover to P: (HOW follows immediately behind main frame sync)

$$\text{mean} = T_h$$

$$\text{S. D.} = 0$$

- Wait for P-data main frame sync: (P-data delay by T_h)

$$\text{mean} = 0$$

$$\text{S. D.} = 0$$

- Collect navigation data from P

$$\text{mean} = T_n$$

$$\text{S. D.} = 0$$

- Time for all operations

$$\text{mean} = T/2 + T_n + T_h$$

$$\text{S. D.} = T/\sqrt{12}$$

For technique (c), we have

- Wait for C-data subframe sync: Uniformly distributed between zero and T/N

$$\text{mean} = T/(2N)$$

$$\text{S. D.} = T/(N\sqrt{12})$$

- Handover to P

$$\text{mean} = T_h$$

$$\text{S. D.} = 0$$

- Wait for P-data main frame sync: (P-data are delayed by T_h behind C-data)

If handover was made on the i-th HOW, then this wait is

$$T - (T/N)(i - 1) \quad \text{for } i > 1$$

or $$0 \quad \text{for } i = 1$$

Since it is equally likely that the handover is made on any HOW, the mean and S. D. of the wait for P-data main frame sync will be

$$\text{mean} = \frac{1}{N}\sum_{i=2}^{N}\left[T - \frac{T}{N}(i-1)\right]$$

$$= \frac{T(N-1)}{2N}$$

$$(\text{S. D.})^2 = \frac{1}{N}\left\{\sum_{i=2}^{N}\left[T - \frac{T}{N}(i-1) - \frac{T(N-1)}{2N}\right]^2 + \left[\frac{T(N-1)}{2N}\right]^2\right\}$$

$$\text{or} \quad \text{S. D.} = \frac{T}{N}\sqrt{\frac{N^2-1}{12}}$$

- Collect navigation data from P

$$\text{mean} = T_n$$

$$\text{S. D.} = 0$$

- Time for all above operations

$$\text{mean} = \frac{T}{2N} + T_h + \frac{T(N-1)}{2N} + T_n$$

$$= \frac{T}{2} + T_n + T_h$$

$$(\text{S. D.})^2 = \left(\frac{T}{N\sqrt{12}}\right)^2 + \left(\frac{T}{N}\sqrt{\frac{N^2-1}{12}}\right)^2$$

$$\text{or} \quad \text{S. D.} = \frac{T}{\sqrt{12}}$$

Thus, the TTFF cannot be reduced by employing more than one HOW per frame. Recapitulating the significant time-consuming steps in the start-up process, and assuming technique (a), we have the following table:

TABLE 5

START-UP FOR P-NAVIGATION

	Mean Time (sec)	S. D. of Time (sec)
C/A-code search	See Table 2	
Wait for frame sync	15.0	8.7
Collect navigation data	18.0	0
Handover to P	1.3	0

In Table 5, we have assumed, as before, a frame duration of 30 sec (T = 30) and all the navigation data in the first 3/5 of the frame (T_n = 18). Additionally, we have assumed 1.3 sec for handover time (T_h = 1.3), made up of 1 sec for reception of the HOW itself and 0.3 sec for loop pull-in and settling.

Based on the values in Tables 2 and 5, and the expressions previously derived, we can find the mean and 90% confidence values of the TTFF at threshold, as a function of threshold, where threshold is given in terms of the received carrier-to-noise density ratio. These values are shown in Table 6.

The following applies to L_2-P navigation. L_2 measurements would be made every 10 sec, with a 10 point moving arc average over a 90 sec span. Thus, in the case of L_2-P navigation, the first L_2 correction is made 10 sec after the L_1-P TTFF. This is not, in general the most accurate fix, but, of course, this comment applies to all TTFFs. This leads to the results shown in Table 7.

TABLE 6

TTFF AT THRESHOLD, FOUR CHANNEL RECEIVER, L_1/P-NAVIGATION

Threshold (C/N_o) -Received dBHz	TTFF (sec)	
	Mean	90% Confidence
30.8	105.7	211.4
31.8	88.3	169.0
32.8	75.6	138.3
33.8	67.2	118.2
34.8	61.0	103.5
35.8	56.3	92.6
36.8	52.6	84.2
37.8	50.0	78.3

TABLE 7
TTFF AT THRESHOLD, FOUR CHANNEL RECEIVER, L_2/P-NAVIGATION

Threshold (C/N_o) -Received dBHz	TTFF (sec)	
	Mean	90% Confidence
30.8	115.7	221.4
31.8	98.3	179.0
32.8	85.6	148.3
33.8	77.2	128.2
34.8	71.0	113.5
35.8	66.3	102.6
36.8	62.6	94.2
37.8	60.0	88.3

By interpolation of the data shown in Table 2, it can be seen that the C/A-code can be acquired with the probability of 0.9 in less than 50 sec with a received carrier-to-noise density ratio of 35.1 dBHz.

With such a threshold signal level, the TTFF at threshold, for the L_1/P navigator is, from Table 6.

$$\text{mean TTFF} = 59.6 \text{ sec}$$
$$90\% \text{ TTFF} = 100.2 \text{ sec}$$

Thus, the expected TTFF will be 59.6 sec at the same threshold which gives a 90% acquisition time for the C/A-code of 50 sec.

Clearly, if the restriction of acquiring the C/A-code within 50 sec is removed, the expected TTFF requirements of 3 or 5 min can be met with much lower signal levels. We do not have computer calculations for each low thresholds, but as may be seen from Table 6, the expected TTFF is 105.7 sec with (C/N_o) of 30.8 dBHz, even though, at this signal level, acquisition time of the C-code is 71.4 sec (mean value), or 140.3 sec (90% probability value).

2.1.3 TTFF For the Sequential Receiver

In this section the TTFF with sequential tracking receivers is computed. Also, the 2-channel sequential receiver is considered, in which one channel (known as Channel B), is used for search and acquisition of the C-code and collection of data, performing these functions in an uninterrupted manner, and the other channel (Channel A) is used to sequentially track the four signals (either C or P) and make the range and range rate measurement. Channel A is initially prepositioned by Channel B, and then routinely reacquires the signals using the flywheel method.

The initial operation steps for a sequential C-navigator which enter into the calculations for the TTFF are shown in Table 8. The TTFF is given by

$$T = \sum_{i=1}^{4} \left(t_{1i} + t_{2i} + t_{3i} \right)$$

TABLE 8

INITIAL STEPS FOR SEQUENTIAL C-NAVIGATION

TTFF

Channel A	Channel B	Time
	Search and acquire C1	t_{11}
	Preposition A to C1	negl.
Sequentially track C1	Wait for C1 data frame sync	t_{21}
	Collect C1 navigation data	t_{31}
	Search and acquire C2	t_{12}
	Preposition A to C2	negl.
Sequentially track C1 and C2	Wait for C2 data frame sync	t_{22}
	Collect C2 navigation data	t_{32}
	Search and acquire C3	t_{13}
	Preposition A to C3	negl.
Sequentially track C1, C2, and C3	Wait for C3 data frame sync	t_{23}
	Collect C3 navigation data	t_{33}
	Search and acquire C4	t_{14}
	Preposition A to C4	negl.
Sequentially track C1, C2, C3, and C4	Wait for C4 data frame sync	t_{24}
	Collect C4 navigation data	t_{34}

Now, the four variates t_{1i} (i = 1,2,3,4) are independent; their mean and standard deviation values (call them m and) are shown in Table 2.

Similarly, the four variates t_{2i} are independent; each is uniformly distributed between zero and 30 sec, and hence has a mean of 15 and a standard deviation of 8.7.

The four variates t_{3i} are all fixed and equal to 18 sec.

Hence, it can easily be shown that

$$T_{mean} = 4m + 132$$

and

$$T_{SD} = (4\sigma^2 + 300)^{\frac{1}{2}}$$

Also, the 90% confidence value is

$$T_{90} = T_{mean} + 1.28\, T_{SD}$$

$$= 4m + 132 + 1.28\, (4\sigma^2 + 300)^{\frac{1}{2}}$$

Using the values of m and σ from Table 2, we have the TTFFs for Sequential C-Navigation, as shown in Table 9.

TABLE 9
TTFF AT THRESHOLD, SEQUENTIAL C-NAVIGATION

Threshold (C/N_o) -Received dBHz	TTFF (min)	
	Mean	90% Confidence
30.8	6.96	9.29
31.8	5.80	7.58
32.8	4.95	6.33
33.8	4.39	5.52
34.8	3.98	4.91
35.8	3.67	4.46
36.8	3.42	4.11
37.8	3.25	3.88

The operations for the sequential L_1/P navigator are shown in Table 10. The TTFF is given by

$$T = \sum_{i=1}^{4} \left(t_{1i} + t_{2i} + t_{3i} + t_{4i} \right)$$

TABLE 10
INITIAL STEPS FOR SEQUENTIAL L_1/P-NAVIGATION

	Channel A	Channel B	Tim
TTFF		Search and acquire C1	t_{11}
		Wait for C1 data frame sync	t_{21}
		Collect C1 navigation data	t_{31}
		Handover C1 → P1, B → A	t_{41}
	Sequentially track P1	Search and acquire C2	
		Wait for C2 data frame sync	
		Collect C2 navigation data	
		Handover C2 → P2, B → A	
	Sequentially track P1 and P2	Search and acquire C3	t_{13}
		Wait for C3 data frame sync	t_{23}
		Collect C3 navigation data	t_{33}
		Handover C3 → P3, B → A	t_{43}
	Sequentially track P1, P2, and P3	Search and acquire C4	t_{14}
		Wait for C4 data frame sync	t_{24}
		Collect C4 navigation data	t_{34}
		Handover C4 → P4, B → A	t_{44}
	Sequentially track P1, P2, P3, and P4		

The first three sets of variates are the same as those described before. The four variates, t_{4i}, are taken as fixed values, each equal to 1.3 sec.

Hence, it can easily be shown that

$$T_{mean} = 4m + 137.2$$

$$T_{SD} = (4\sigma^2 + 300)^{\frac{1}{2}}$$

and

$$T_{90} = T_{mean} + 1.28\, T_{SD}$$

Using the values of m and σ from Table 2, we have the TTFFs for Sequential L_1/P-Navigation, as shown in Table 11.

TABLE 11

TTFF AT THRESHOLD, SEQUENTIAL L_1/P-NAVIGATION

Threshold (C/N_o) - Received dBHz	TTFF (min)	
	Mean	90% Confidence
30.8	7.05	9.38
31.8	5.89	7.67
32.8	5.04	6.42
33.8	4.48	5.61
34.8	4.07	5.00
35.8	3.76	4.55
36.8	3.51	4.20
37.8	3.34	3.97

2.2 Navigation Performance

Report No. 7 of this volume contains an analysis of the pseudorange and range-rate tracking errors. Table 12 summarizes the results for a non-accelerating user:

TABLE 12

PSEUDO-RANGE, RANGE-RATE TRACKING ERRORS

C/N_o - dB-Hz	Range Error (σ) ft		Range-Rate Error (σ) ft/sec
	C/A	P	P and C/A
24	165	8.9	0.17
26	109	6.1	0.13
28	72	4.4	0.10
30	49	3.5	0.08
32	34	2.5	0.063
34	23	1.9	0.05
36	17	1.5	0.04
38	12	1.2	0.034
40	8.5	1.0	0.026

2.3 Data Performance

Both the P-signal and C/A-signal transmit data at a 50 bps rate. The data is differentially encoded in order to automatically resolve a possible phase ambiguity in the phase reference. Table 13 presents the data signal-to-noise density ratio required to achieve a bit error rate of 10^{-5}.

TABLE 13
DATA DEMODULATION REQUIREMENTS

ITEM	VALUE	
	C/A-Data Signal	P-Data Signal
Data Rate (50 bps)	17.0 dB	17.0 dB
E_b/N_o (1)	9.8 dB	9.8 dB
Demodulation Loss (2)	2.0 dB	2.0 dB
Multiple Access Loss (3)	1.0 dB	0.0 dB
Required Data Signal-to Noise Density Ratio	29.8 dB-Hz	28.8 dB-Hz

(1) Required energy-to-noise density ratio for 10^{-5} bit error rate assuming coherent PSK detection with differential decoding.

(2) Includes τ-dither loss (1.3 dB) plus losses due to carrier and bit sync jitter and filter mismatch.

(3) The loss due to multiple access interference for 1023 chip code with an interference level which is not exceeded 95% of the time.

Based on a worst case receiver noise density of -199.6 dBW/Hz the required signal levels at the PN receiver input for an error rate of 10^{-5} at 50 bps are:

-169.8 dBW for the C/A-signal
-170.8 dBW for the P-signal

3.0 C/A-SIGNAL MULTIPLE ACCESS PERFORMANCE

In this section, the performance of the C-signal in the expected multiple-access environment is analyzed.

Multiple-access interference is caused by all the C-signals from satellites which are in view, except the desired C-signal. Since only the actual code is different for each satellite, the level of interference depends upon the cross-correlation properties of the desired code with all the other codes. In addition, it is estimated that an undesired C-signal may be 11.2 dB above the desired signal. In order to determine the performance of the C-signal, the probability that a given interference level will not be exceeded will be determined. This relationship is affected by the cross-correlation properties of the Gold codes and the geometric relationship between the satellites.

Section 3.1 presents the cross-correlation properties for Gold-code lengths of 511, 1023, and 2047 chips.

Section 3.2 analyzes the effects of multiple-access interference on code-acquisition time, ranging performance with a coherent and non-coherent delay-lock loop (DLL), and coherent data detection. Expressions are developed relating the above performance parameters to the ratio of C-signal power-to-multiple-access interference power.

3.1 Properties of Gold Codes

All of the Gold codes of a family can be generated from a pair of selected PN sequences. The number or codes in a family, all of which are the same length, is $2^n + 1$ where $2^n - 1$ is the code length. A Gold code is generated by mod-2 addition of one PN sequence with one possible shift of the other PN sequence. Taking all possible shifts, this results in $2^n - 1$ code. The other two Gold codes are the two PN sequences themselves.

The property which makes the Gold codes desirable for the C-signal is that all the levels of the cross correlation function between any two Gold codes of a family is known to be relatively low. The levels of the cross correlation function are determined by integrating the cross correlation waveform over a single code length or integer multiples of a code length. It has been shown by R. Gold (Reference 2) that the dc component of the cross correlation function has only three levels. This result also assumes equal code clock frequencies and that the code clocks are to be exactly in phase. Table 14 summarizes the cross correlation properties of any two Gold codes for the above assumptions.

TABLE 14

CROSS-CORRELATION PROPERTIES OF GOLD CODES

Register Length	Code Length	Normalized Cross Correlation Level	Probability of Level
n Odd	$L = 2^n - 1$	$K = -(2^{\frac{(n+1)}{2}} + 1)/L$	~ 0.25
		$-1/L$	~ 0.50
		$(2^{\frac{(n+1)}{2}} - 1)/L$	~ 0.25
n Even	$L = 2^n - 1$	$K = -(2^{\frac{(n+2)}{2}} + 1)/L$	~ 0.125
		$-1/L$	~ 0.75
		$(2^{\frac{(n+2)}{2}} - 1)/L$	~ 0.125

For example, if n = 9 ($2^n - 1 = 511$), the possible cross correlation levels for the dc term are

$$K = \frac{-33}{511} = -0.06457 \text{ (-23.8 dB) with a probability of 0.25}$$

$$-\frac{1}{511} = -0.00195 \text{ (-54.2 dB) with a probability of 0.50}$$

$$\frac{31}{511} = 0.06066 \text{ (-24.4 dB) with a probability of 0.25}$$

As mentioned above, Gold only considered the dc levels of the cross correlation function. It will be shown in Section 3.2.1 that not only is the dc level of the cross-correlation function important but higher frequency terms may also be important. Thus, we would like to know the frequency spectrum of all possible cross-correlation waveforms. It will be shown in Section 3.2.1 that the cross- correlation waveform is actually a two-level waveform which is periodic with a period equal to the code length. Therefore, the spectrum of the cross-correlation waveform is a line spectrum with frequency components at dc and an integer multiple of the code epoch rate. For example, if the chip rate is 1.023 Mbps and code length 1023, the cross-correlations waveform has values at dc, 1.0 KHz, 2.0 KHz, 3.0 KHz, ---. The level of these components depends upon the relative phase of the two Gold codes being considered. A typical spectrum is illustrated in Figure 1 for a 1023 Gold code.

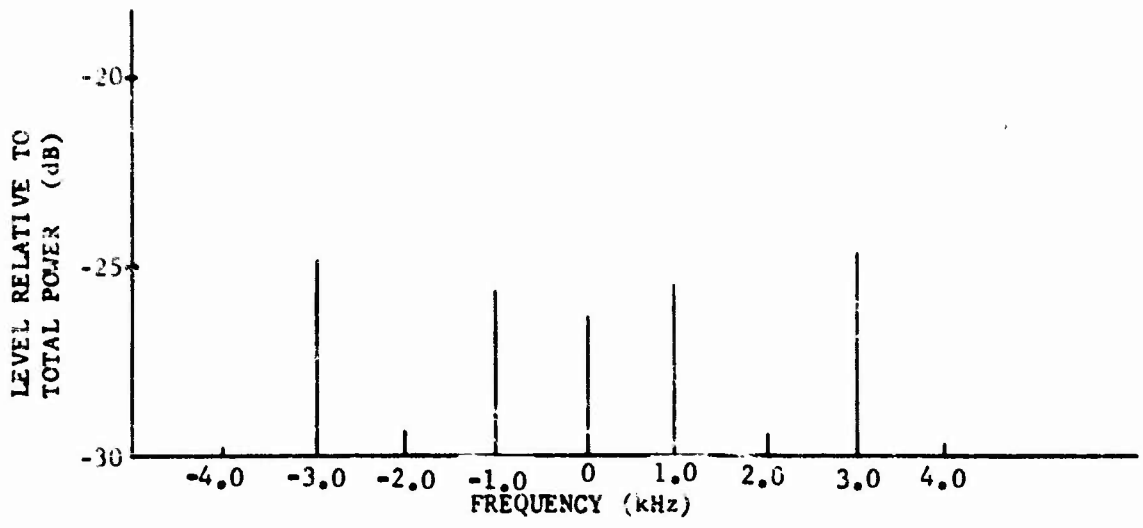


Figure 1 1023 Gold Code Spectrum

The cross-correlation waveform between any two Gold codes of a family is itself one of the Gold codes of the same family. Thus, by determining the frequency spectrum of all the Gold codes in a family, all possible frequency spectra of the cross-correlation waveform are determined. The frequency spectrum of the Gold codes have been obtained in two ways. The first approach used actual Gold-code generators and measured the levels as a function of frequency. Figure 2 presents a typical output. All possible second sideband levels of the cross-correlation function for the Gold codes are presented in Figure 2. Note that a peak of approximately -18.6 dB occurs only once. Figures 3 and 4 present the levels of the first sideband for 1023 and 2047 Gold codes, respectively. Note that the peak level decreases approximately 3 dB each time the code length is doubled.

The second approach was to simulate the Gold codes on a digital computer. This approach allows the statistics of the various levels to be determined very quickly. Figure 5 presents the cumulative probability of the dc and sideband levels for the 511 Gold codes. It was found that the cumulative probability was essentially the same for all the sidebands. The inphase and random-phase curves correspond to when the code clocks are exactly inphase and when the code-clock phases are random, respectively. Figures 6 and 7 present the same information for the 1023 and 2047 Gold codes, respectively. Note that the peak level of the dc term is equal to -23.8 dB for both the 511 and the 1023 Gold codes, but is -30 dB for the 2047 Gold code. However, the level of each sideband is 3 dB and 6 dB lower for the 1023 and 2047 Gold codes, respectively. On the other hand, there are twice as many sideband components with the 1023 Gold codes and four times as many components with the 2047 Gold codes. The importance of the number of sideband components is discussed below.

A noncoherent delay-lock-loop (DLL) receiver with an IF bandwidth of 1 kHz would only see one component from each satellite with a 511 or 1023 code but would see two components with a 2047 code. Figure 8 presents the level of the first plus second sidebands code. A comparison of the two-sideband level of the 1023 code with a single-sideband level indicates that the average levels are both

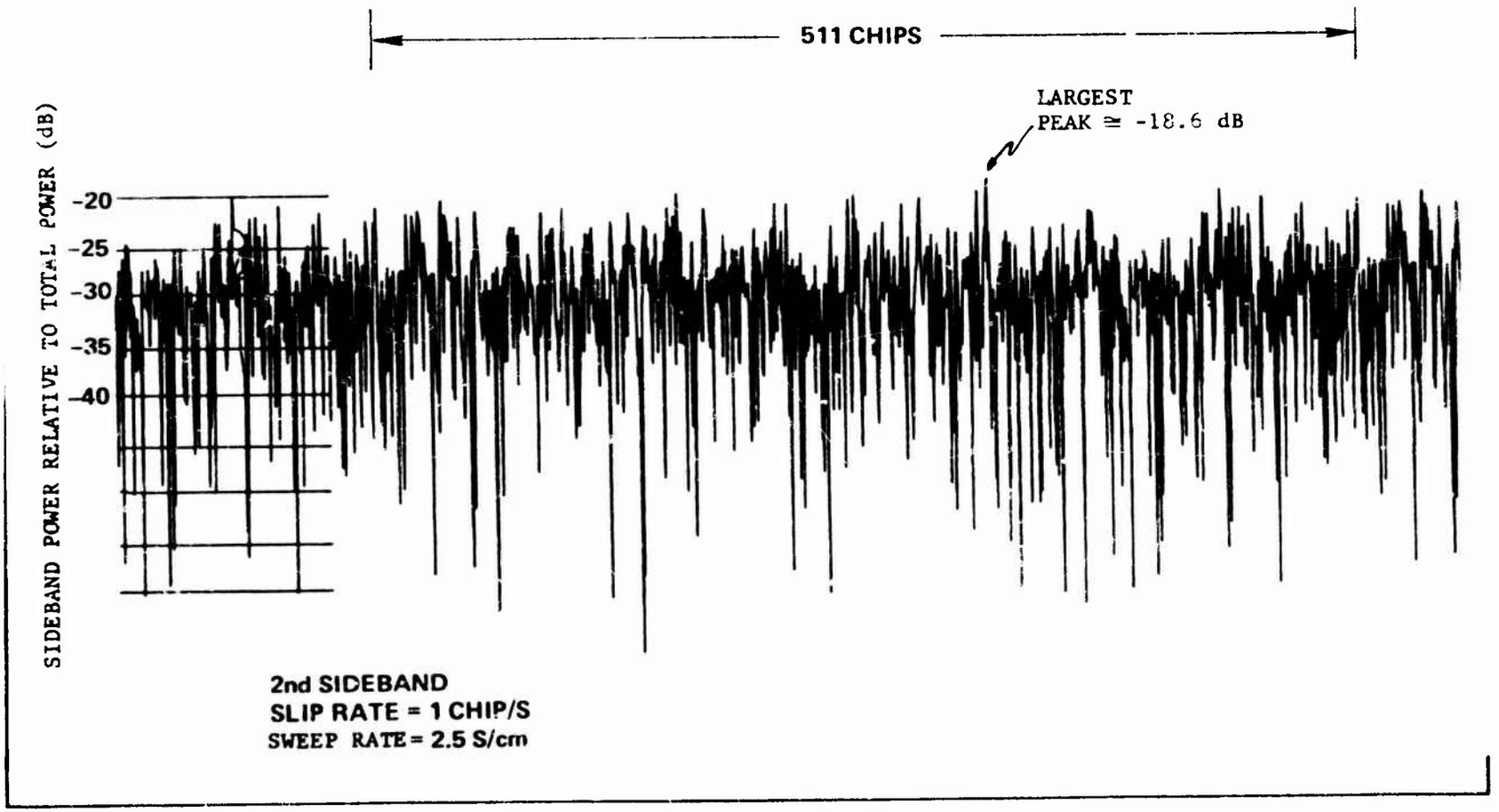


Figure 2 Sideband Cross-Correlation Levels for 511 Gold Codes

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

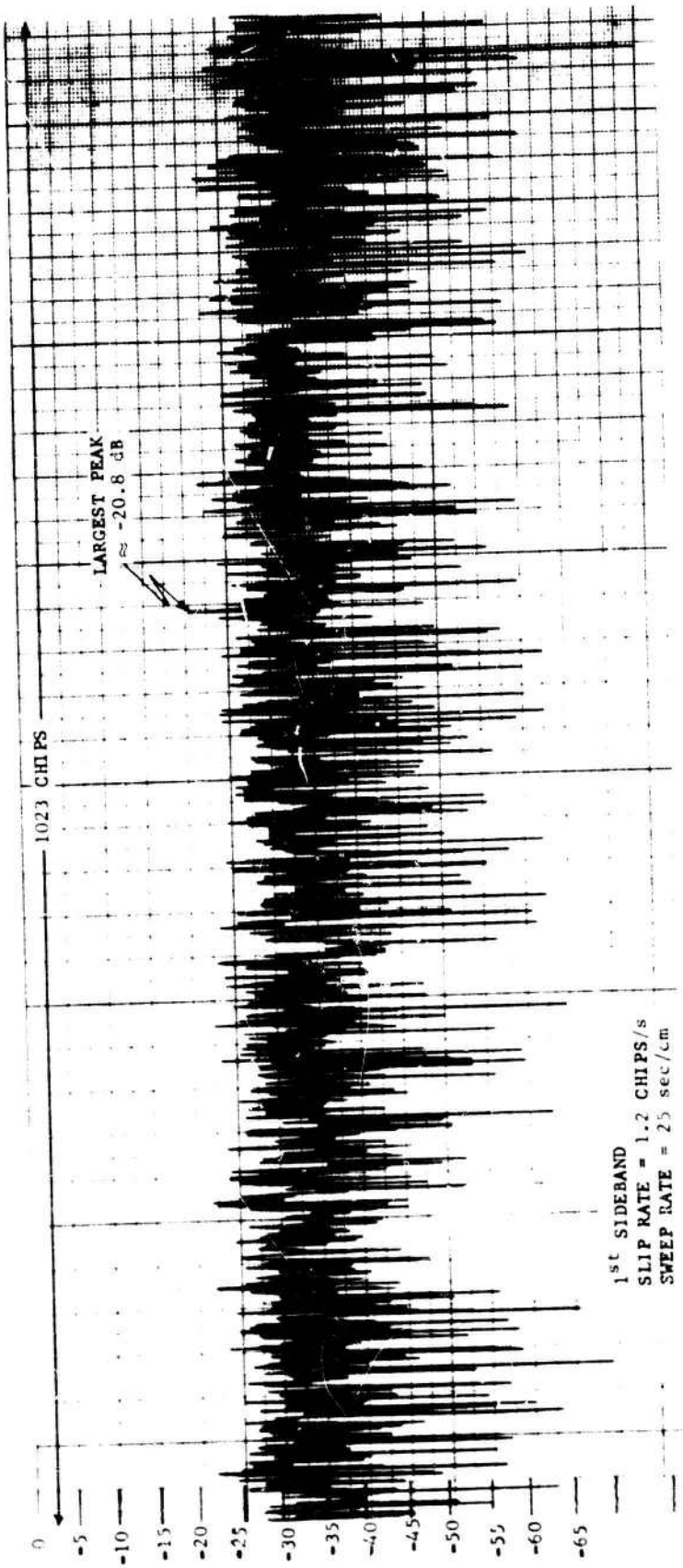


Figure 3 Sideband Cross-Correlation Levels for 1023 Gold Codes

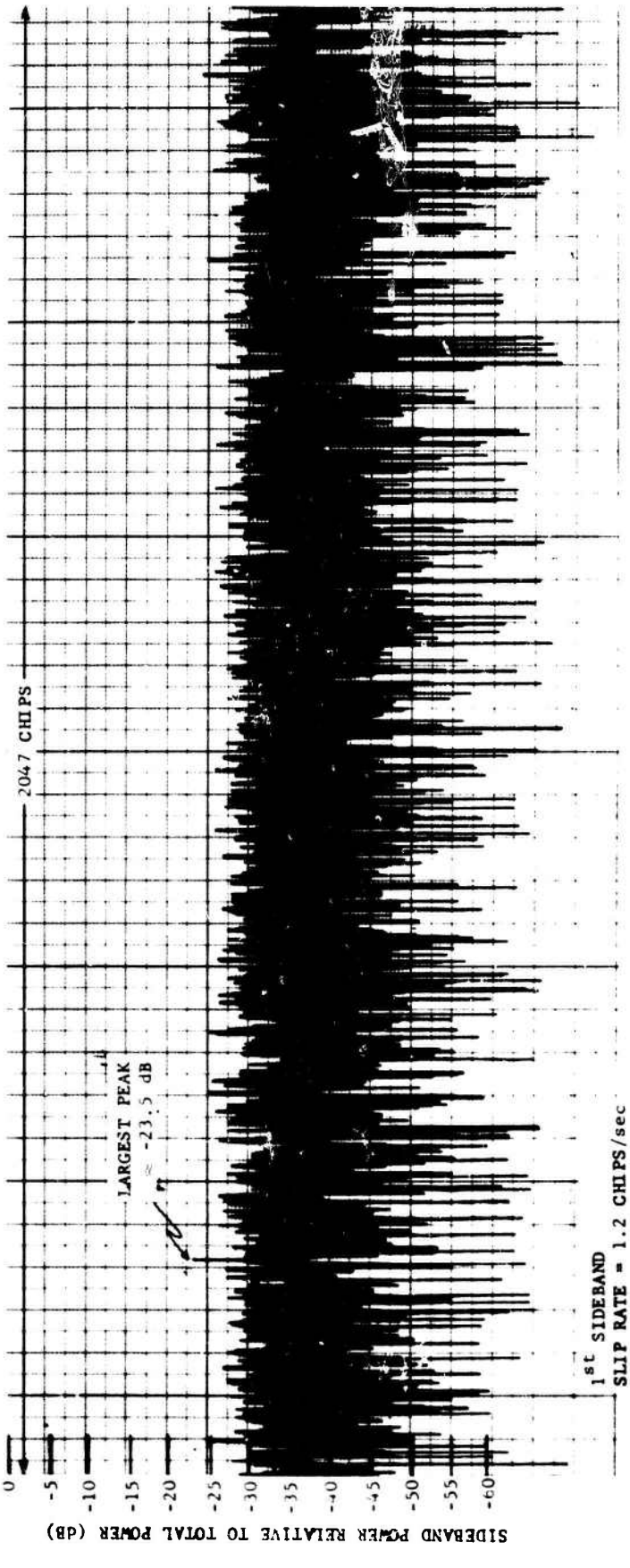


Figure 4 Sideband Cross-Correlation Levels for 2047 Gold Codes

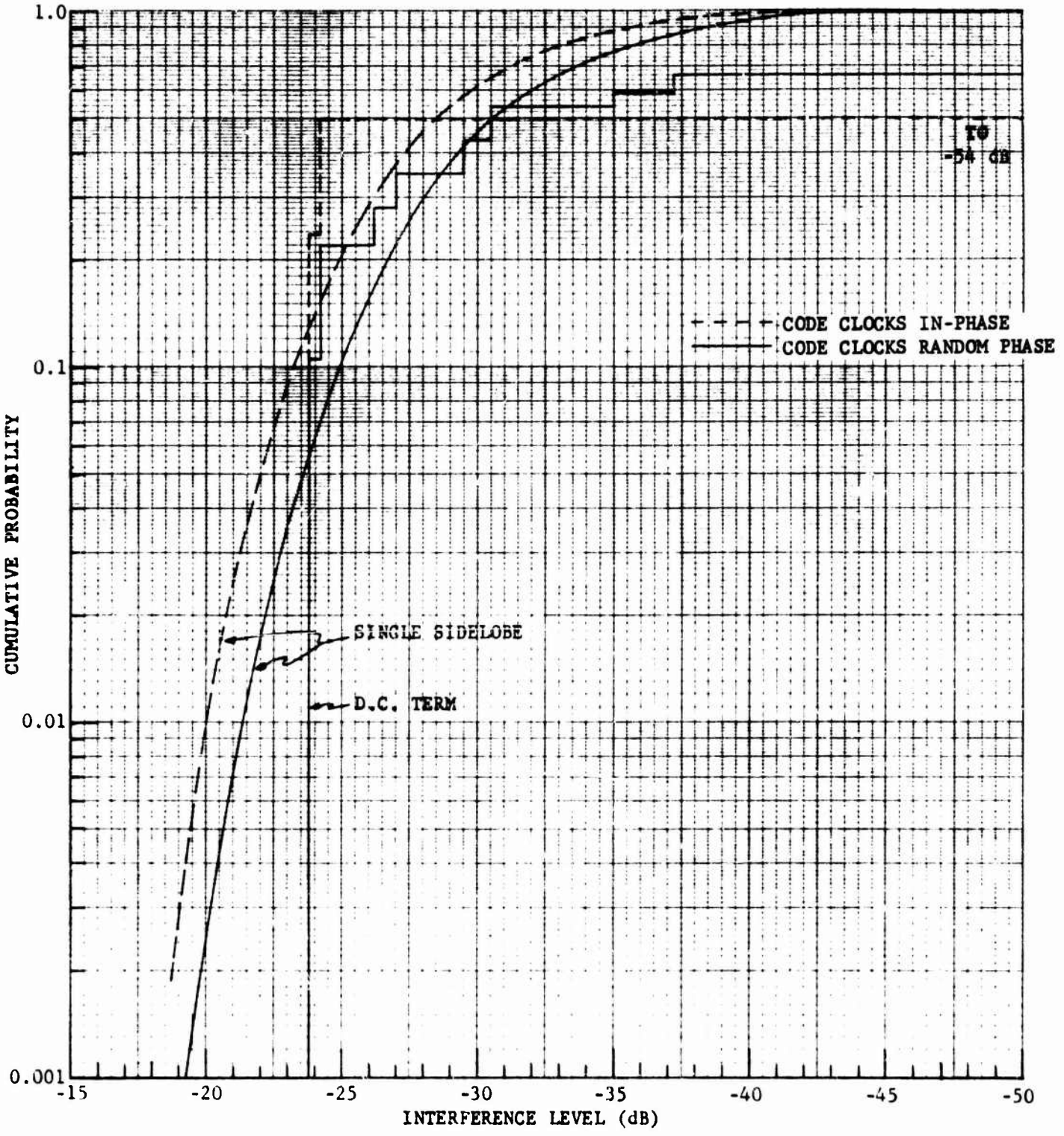


Figure 5 Cumulative Probability of Interference Level for 511 Gold Codes

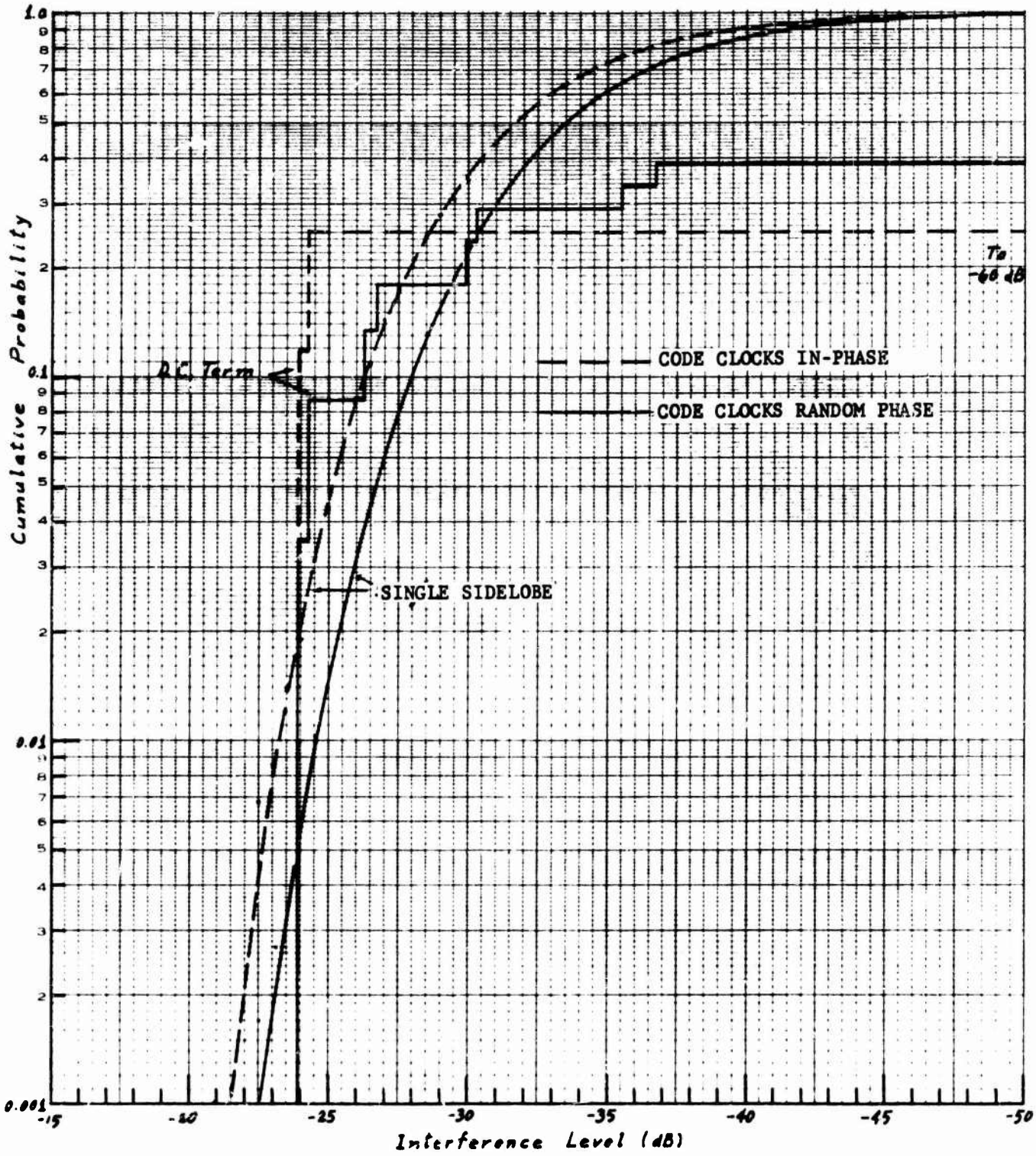


Figure 6 Cumulative Probability of Interference Level for 1023 Gold Codes

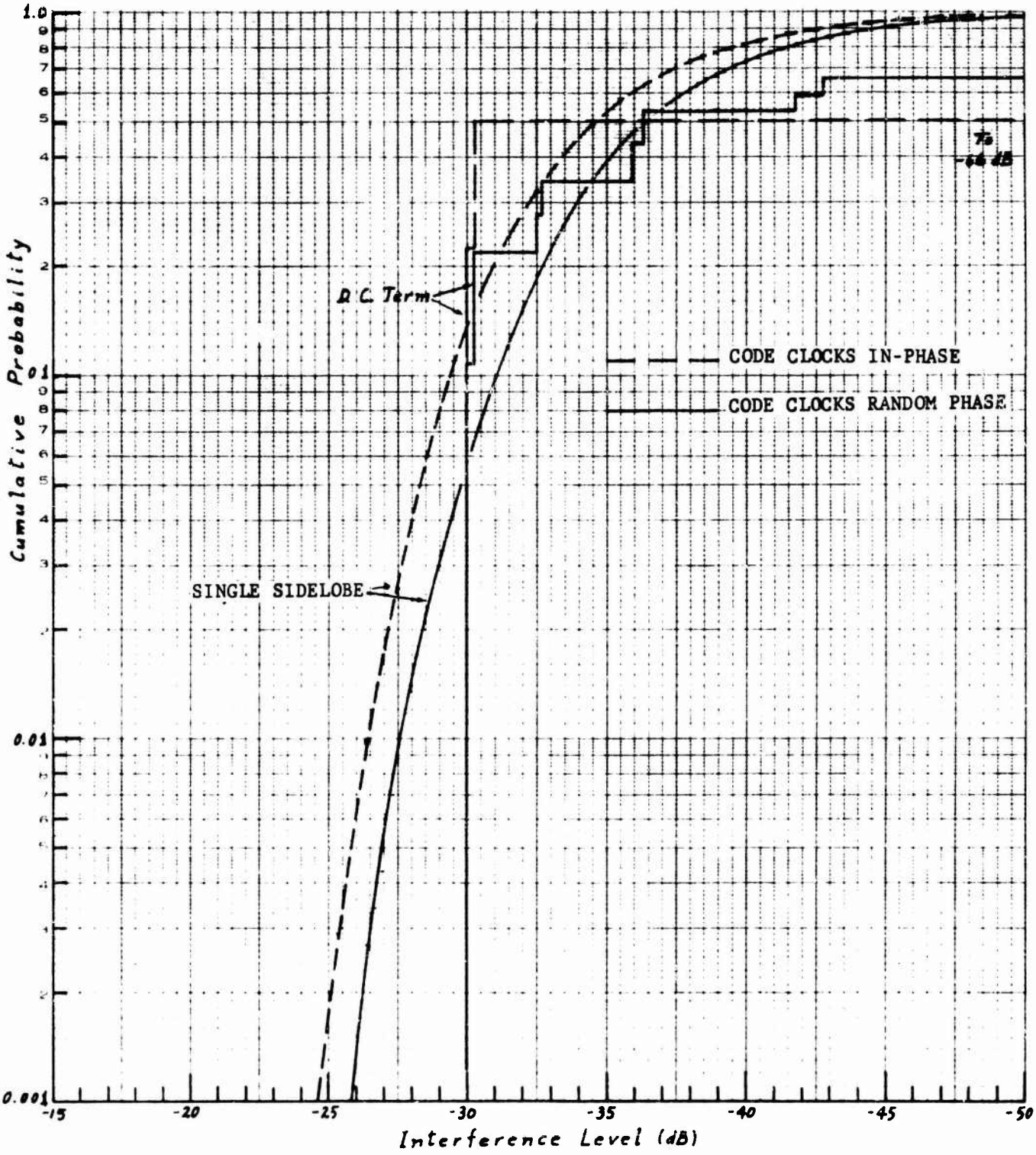


Figure 7 Cumulative Probability of Interference Level for 2047 Gold Codes

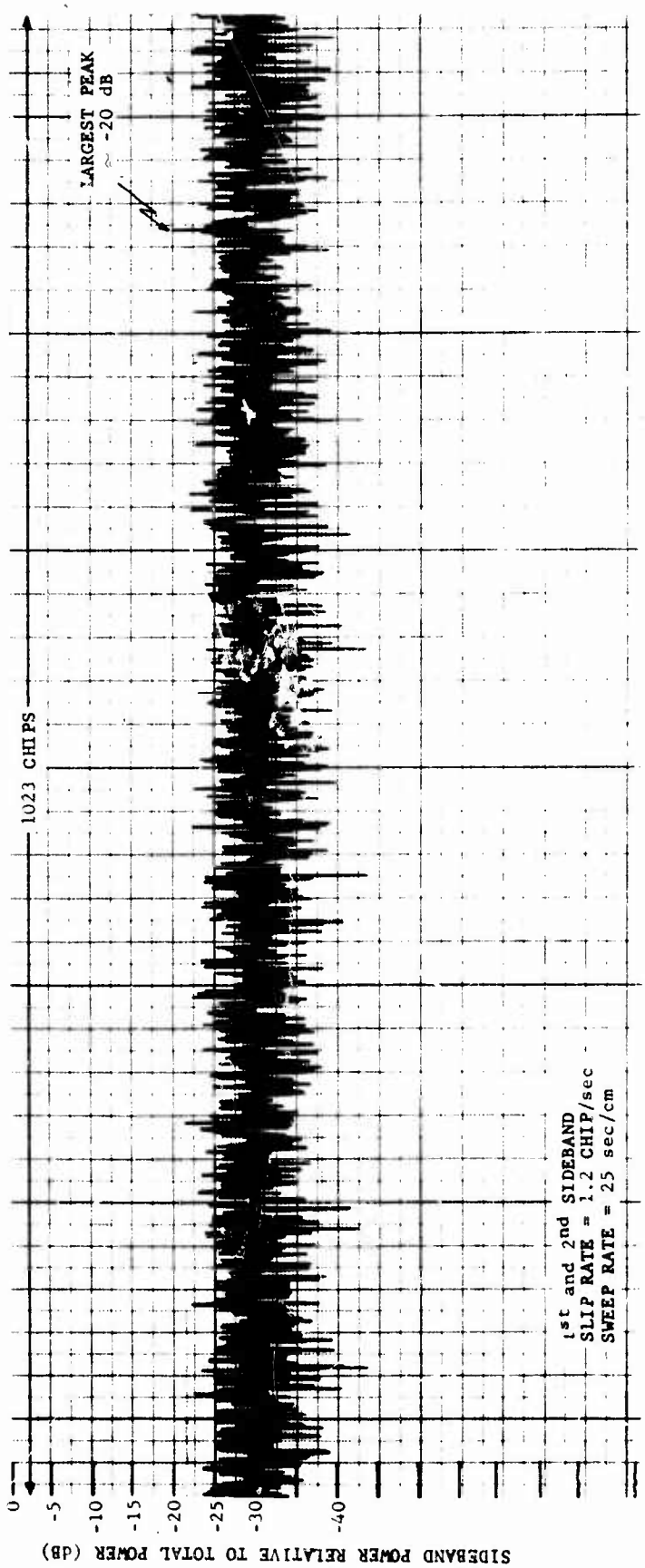


Figure 8 1st and 2nd Sideband Cross-correlation Levels for 1023 Gold Codes

approximately -30 dB; however, the single 511 sideband has peaks which are greater. Thus, if a 1-kHz IF bandwidth is used with the longer codes, the multiple-access interference will be approximately the same as with a 511 or 1023 code (except for the interference levels which have very low probability of occurrance). The IF bandwidth would have to be reduced in order to realize the full 3-dB improvement possible with longer codes.

The code tracking performance of a coherent DLL receiver should only be affected by one sideband from each satellite at any one time since the loop bandwidth is in the order of 20 Hz. The same is true for coherent data detection since the data bandwidth is 50 Hz. Therefore, coherent ranging and data detection performance is improved by 3 dB and 6 dB for the 1023 and 2047 Gold codes, respectively. However, the probability of the carrier frequency being equal to one of the sideband frequencies is double and quadruple for the 1023 and 2047 codes, respectively. This is because the density of sidebands is doubled and quadrupled for the longer codes.

3.2 MULTIPLE-ACCESS PERFORMANCE

Multiple-access interference will affect the code-acquisition time, the tracking error, and the bit error rate. However, the magnitude of the effect will be different for each case. Therefore, the effect of multiple-access interference must be evaluated for each.

3.2.1 Basic Correlation Receiver Model

The received signal, r(t), consists of a desired component denoted by the subscript "o", a collection of N other similarly constructed signal components, and noise.

$$r(t) = A_o c_o(t) \cos(\omega_o t + \phi_o) + \sum_{i=1}^{N} A_i C_i(t) \cos(\omega_i t + \phi_i) + n(t)$$

The model is general in that each component may have a different amplitude, carrier frequency, and phase angle. The major distinction between the several signal components is the individual Gold codes, c_i (i = 0 to N), which occur at the same nominal chip rate and have the same degree of mutual cross-correlation.

Correlation detection consists of mixing the received signal with a reference signal, s(t).

$$s(t) = 2\, C_o(t) \cos \left[(\omega_o - \omega_d)\, t + \phi_d\right]$$

and narrowband filtering at the difference frequency, ω_d. The difference can be zero. The averaging time of the filter should be approximately equal to the code length or a multiple of the code length in order to avoid a reduction in signal level due to partial correlation. The mixer output at the difference frequency, ω_d, is given by

$$y(t) = r(t)\, s(t) \Big|_{\text{difference frequency}}$$

$$= A_o \cos(\omega_d t + \theta_d) + \sum_{i=1}^{N} A_i c_o(t)\, c_i(t) \cos\left[(\omega_d + \Omega_i)\, t + \theta_i\right] + n'(t)$$

where

$$\theta_i = \phi_i - \phi_d$$

$$\Omega_i = \omega_i - \omega_o$$

Thus, we have the desired component with steady amplitude, A_o, immersed in a collection of several signals with different frequency offsets, Ω_i, amplitudes, A_i, and modulated by the code products, $c_o(t)\, c_i(t)$. Let us focus on the code product term, $c_o(t)\, c_i(t)$, for a moment.

The chip rate of all the signals will be approximately the same; therefore, if the desired and the undesired code clocks are inphase, the product of $c_o(t)$ $c_i(t)$ is just another Gold code, $c_i(t)$. If the two code clocks are not inphase, the product is not a Gold code; however, its spectral properties are very similar. In both cases, the product of $c_o(t)\, c_i(t)$ is a two-level sequence with a length equal to the length of individual Gold codes. Therefore, the waveform, $c_o(t)\, c_i(t)$, can be represented using a Fourier series expansion. That is,

$$C_o(t)\, C_i(t) = K + \sum_{n=1}^{\infty} \alpha_n \cos\left(\frac{2\pi n t}{T} + \varphi_n\right) \qquad i = 1, 2, \ldots N$$

where $\alpha_n = \sqrt{A_n^2 + B_n^2}$

$$\varphi_n = \tan^{-1}(-B_n/A_n)$$

$$K = \frac{1}{T}\int_o^T C_o(t)\, C_i(t)\, dt$$

$$A_n = \frac{2}{T}\int_o^T C_o(t)\, C_i(t) \cos\left(\frac{2\pi nt}{T}\right) dt$$

$$B_n = \frac{2}{T}\int_o^T C_o(t)\, C_i(t) \sin\left(\frac{2\pi nt}{T}\right) dt$$

T = (GOLD CODE LENGTH) ÷ CHIP RATE

Thus, the mixer output can be written as

$$y(t) = A_o \cos(w_d t + \Theta_d) + \sum_{i=1}^{N} A_i \left[K_i + \sum_{i=1}^{\infty} \alpha_{n_i} \cos\left(\frac{2\pi nt}{T}\right) + \varphi_n\right] \cos\left[w_d + \Omega_i)\, t + \Theta_i\right]$$

Expanding a few of the terms and neglecting the sum terms,

$$y(t) = A_o \cos(w_d t + \Theta_d) + \sum_{i=1}^{N} A_i K_i \cos\left[(w_d + \Omega_i)\, t + \Theta_i\right]$$

$$+ \sum_{i=1}^{N} \tfrac{1}{2}\, \alpha_{1_i} \cos\left[(w_d + \Omega_i - \frac{2\pi}{T})t + \Theta_i - \varphi_1\right]$$

$$+ \sum_{i=1}^{N} \tfrac{1}{2}\, \alpha_{2_i} \cos\left[(w_d + \Omega_i - \frac{4\pi}{T})t + \Theta_i - \varphi_2\right]$$

$$+ \ldots .$$

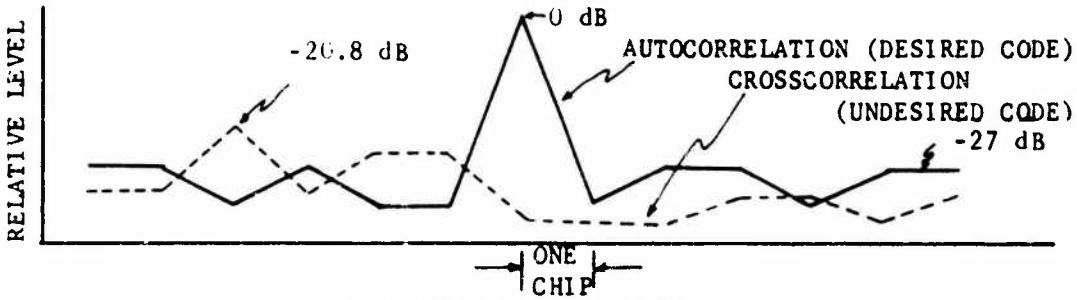
Note that if any or all of the Ω_i 's are equal to zero, the level of interference is equal to K_i. In addition, if any or all of the Ω_i 's are equal to $\frac{2\pi}{T}$, the level of interference has a peak equal to $\frac{1}{2}\alpha_1$. Similarly, if any or all of the Ω_i's are equal to $\frac{2\pi n}{T}$, the level of interference has a peak equal to $\frac{1}{2}\alpha_n$. For a chip rate of 1.023 Mbps and a code length of 1023, the $\frac{2\pi n}{T}$ are equal to 1.0 kHz, 2.0 kHz, 3.0 kHz, and so on. Since the maximum doppler shift is in the order of 4 kHz, the sideband cross correlation is important.

The above model will be used to determine performance in a multiple access environment in the next sections.

3.2.2 Acquisition Performance

3.2.2.1 Lock-Up to Undesired Code.- The desired code is acquired by correlating the received signal with the locally generated desired signal. The locally generated code is advanced one-half chip position until a lock condition is detected. This inhibits the advancing pulses and allows the loop to track. After about ten loop time constants, if the lock detector still indicates lock, the code inlock condition is indicated. The second lock condition is necessary in order to reject false lock conditions due to thermal noise or an undesired code.

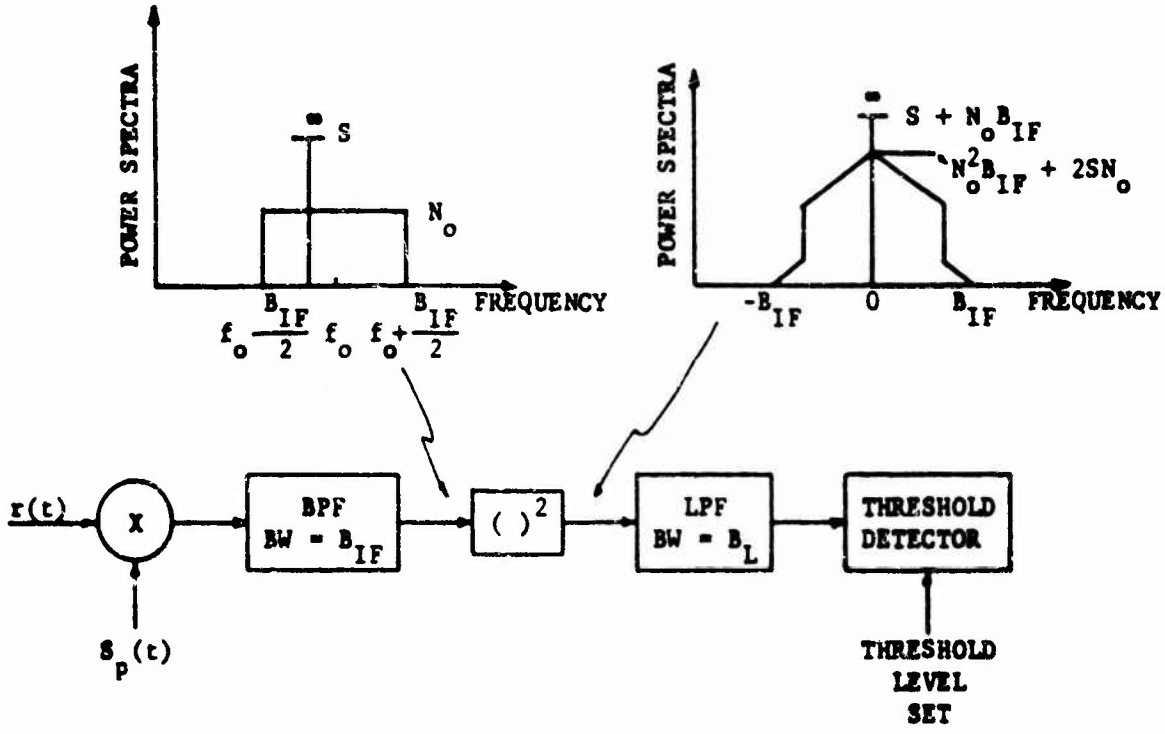
Let us discuss how an undesired code is rejected. Typical autocorrelation and cross-correlation functions for the 1023 Gold codes are sketched below.



Referring to the above sketch, we see that there is a potential false-lock condition, especially since the undesired signal level can be 11.2 dB above the desired signal level. That is, there can be a cross-correlation peak as large as -20.8 + 11.2 = -9.6 dB (9.6 dB below the desired signal peak). This undesired peak will, with high probability, inhibit the advancing pulses. However, by choosing the code loop noise bandwidth, B_L, such that the interference signal-to-thermal noise ratios in B_L is below loop threshold, the thermal noise will, with very high probability, cause loss of lock after ten loop time constants. For example, if the minimum desired signal-to-noise density is +33 dB-Hz, the maximum interference signal-to-noise density is +23.4 dB-Hz. Therefore, choose B_L such that, for a signal-to-noise density of +30 dB-Hz, the loop is at threshold. For this case, B_L = 27.8 Hz for a noncoherent DLL. It may be desirable to use a narrower bandwidth during tracking in order to reduce the ranging error.

The above discussion assumes one undesired code. This should be a reasonable assumption since the probability of the composite cross-correlation function (the cross-correlation function resulting from all undesired codes) resulting in a potential false-lock condition is very small.

3.2.2.2 Performance of Inlock Detector. - A block diagram of a typical inlock detector is presented below. The power spectra at various points are sketched for the signal-plus-noise case (Reference 3) neglecting data modulation.



The threshold level is chosen somewhere between (depends upon the desired probability detection and probability of false alarm) the signal-plus-noise level (correlation) and the noise-only level (no correlation). From the above sketch, the LPF outputs are given by:

dc voltage with correlation = $N_o B_{IF} + S$

rms noise voltage with correlation = $\sqrt{2 B_L N_o (N_o B_{IF} + 2 S)}$

dc voltage with no correlation = $N_o B_{IF}$

rms noise voltage with no correlation = $\sqrt{2 B_L N_o^2 B_{IF}}$

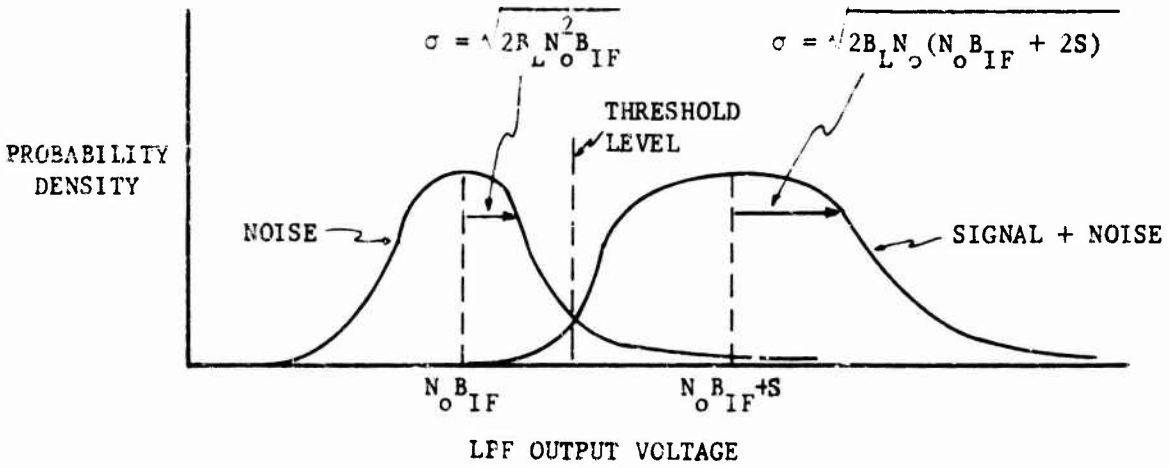
where S is the signal power

N_o is the one-side noise density

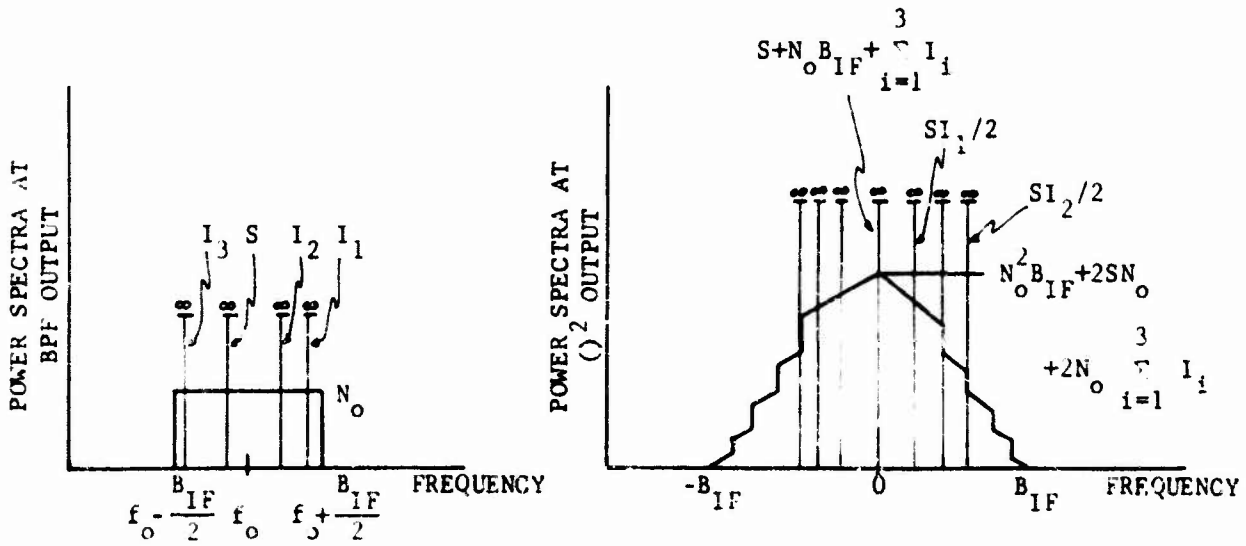
B_{IF} is bandpass filter bandwidth in Hz

B_L is the lowpass filter bandwidth in Hz

For $B_L B_{IF} > 10$, the output voltages are approximately Gaussian. For this approximation, the resulting probability density functions are sketched below.



In the multiple-access environment, there will be a desired signal, N undesired signals, and noise. The power spectra at the output of the BPF and the square-low device are sketched below.



The LPF outputs for this situation are:

dc voltage with correlation = $N_o B_{IF} + S + I_T$

rms noise voltage with correlation = $\sqrt{2\ B_L N_o\ (N_o B_{IF} + 2\ S + 2\ I_T)}$

dc voltage with no correlation = $N_o B_{IF} + I_T$

rms noise voltage with no correlation = $\sqrt{2\ B_L N_o\ (N_o B_{IF} + 2\ I_T)}$

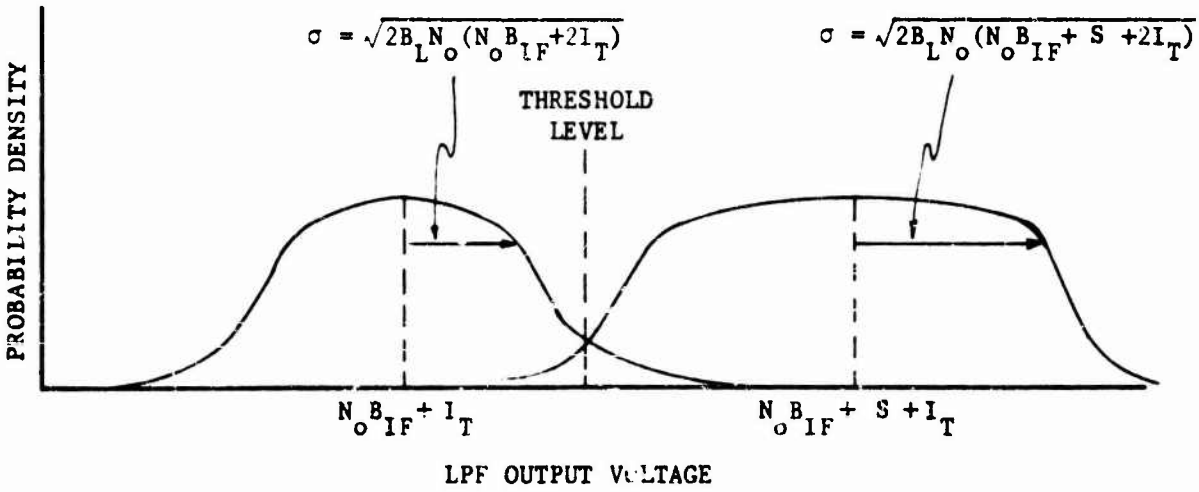
where S is the desired signal power

I_T is total interference power at the correlator output from all undesired signals and is equal to

$$I_T = \sum_{i=1}^{N} I_i$$

where I_i is the interference power of the correlator output from the i^{th} undesired signal.

The resulting probability functions with multiple-access interference are sketched below.



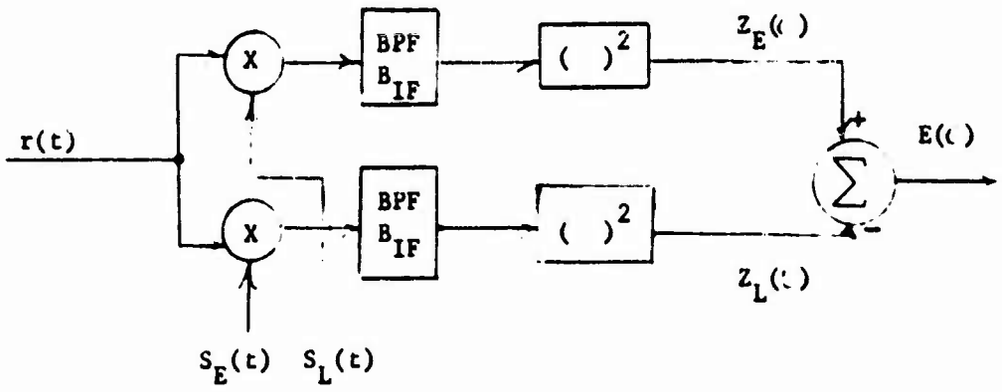
Comparing the probability density functions with and without multiple-access interference, we see there are two effects. The first is that the mean output voltage (with and without correlation) increases in direct proportion to the

total interference level. The second is that the noise power also increases in direct proportion to the total interference level. The first effect requires that the threshold level be related to the uncorrelated output level but doesn't affect the code-acquisition time. However, the second effect requires the dwell time per decision to be increased (B_L decreased) in order to maintain the same probability of detection and probability of false alarm as with no interference. This does affect code-acquisition time since the dwell time is the time required to test each code position. For example, if B_L must be reduced by a factor of 1.5, the dwell time would increase by 1.5 and the total code-acquisition time increases by the same factor. The effect upon time to first fix will be less than the 1.5 factor since code acquisition is only one component of the time to first fix.

Figure 9 presents relative dwell time-per-search cell as a function of signal-to-interference ratio at the correlator output (with ratio of desired signal power-to-noise density as a parameter). For a given ratio of signal to interference, the increase in dwell time is less for lower ratios of signal-to-noise density (greater noise power) since the interference power must be greater to cause an increase in dwell time.

3.2.3 Noncoherent DLL Tracking Performance

A block diagram of the correlation network for a noncoherent DLL is presented below.



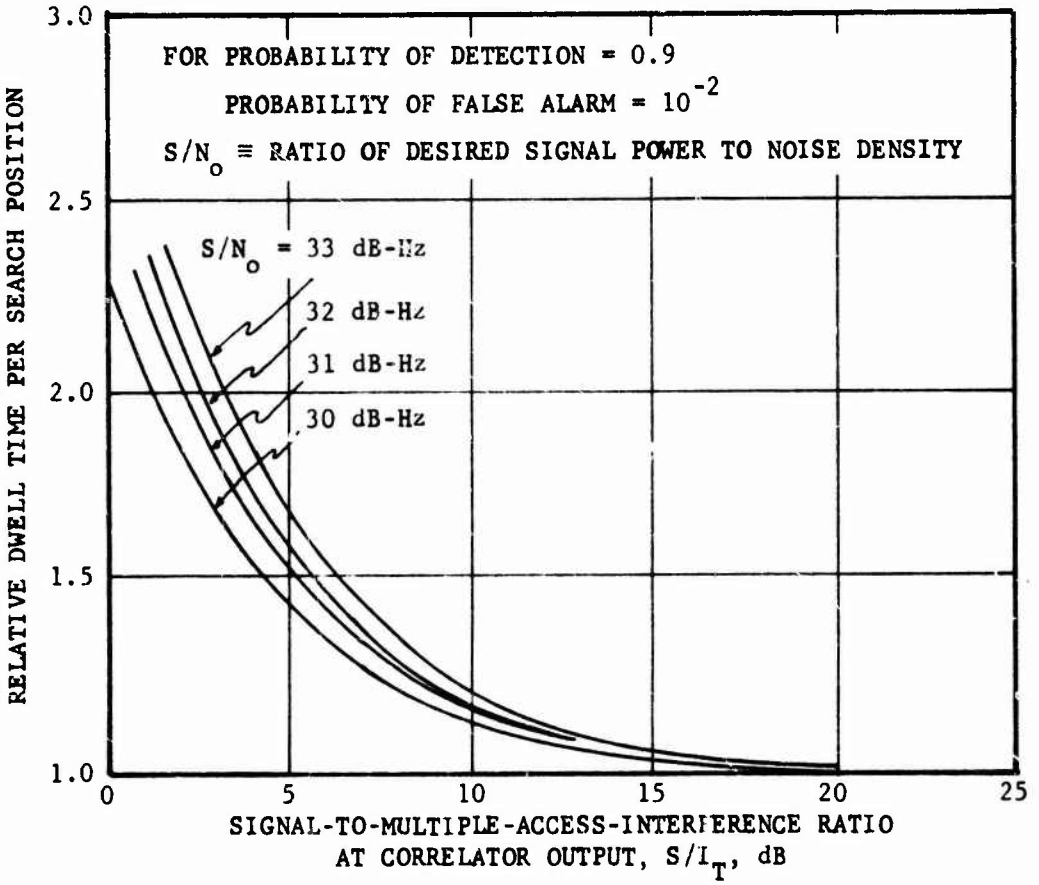


Figure 9 Relative Dwell Time vs. Signal-to-Multiple-Access-Interference Ratio

When the loop is locked:

$$Z_E(\epsilon) = 1/2\ S \left(1/2 \pm \frac{\epsilon}{\Delta} \right)^2 + I_{T_E}$$

$$Z_L(\epsilon) = 1/2\ S \left(1/2 \pm \frac{\epsilon}{\Delta} \right)^2 + I_{T_L} \quad \text{(neglecting noise terms)}$$

where ϵ is the tracking error and $|\epsilon| \leq \frac{\Delta}{2}$

Δ is the chip duration

Thus,

$$E(\epsilon) = S \frac{\epsilon}{\Delta} + (I_{T_E} - I_{T_L})$$

However, on the average, the total interference in each channel should be approximately equal (that is $I_{T_E} \simeq I_{T_L}$) since there are at least four interfering signals in each channel at all times.

Thus,

$$E(\epsilon) \simeq S \frac{\epsilon}{\Delta}$$

In this case, there is no offset voltage caused by the multiple-access interference; however, it can be shown that the noise density with interference has increased and is given by

$$G_n(f) = 1/2\ S\ N_o + 1/4\ N_o\ I_T + 1/2\ N_o^2\ B_{IF}$$

The second term results from the interference. The normalized tracking error is then given by (Reference 4)

$$\left[\frac{\sigma}{\Delta}\right]^2 = \frac{N_o B_L}{S} \left[1/2 \left(1 + \frac{I_T}{S}\right) + \frac{N_o B_{IF}}{S} \right]$$

Note that $I_T/S = 0$ corresponds to the no-interference case. Therefore, one can define a degradation resulting from multiple-access interference, D_I, by

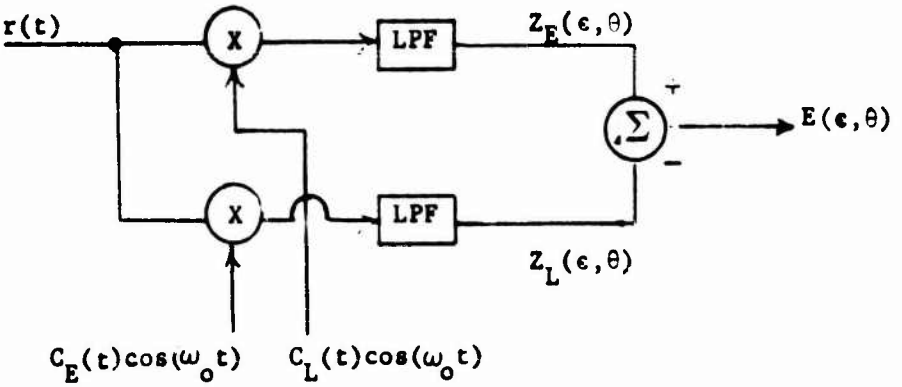
$$D_I = \frac{1/2 + \frac{N_o B_{IF}}{S}}{1/2\ (1+I_T/S) + \frac{N_o B_{IF}}{S}}$$

This degradation, D_I, is presented in Figure 10, as a function of the signal-to-total-interference ratio at the correlator output. Note that, for a signal-to-interference ratio greater than 6 dB, the resulting degradation is less than 0.5 dB. This means that the code loop bandwidth, B_L, must be reduced by a factor of 1.12 in order to provide the same 1σ range error as with no interference.

3.2.4 Coherent DLL Tracking Performance

The performance of the coherent DLL will be based upon one interfering signal since it is reasonable to assume that, at any given instant of time, only one interfering signal is within the loop bandwidth. The loop bandwidth is in the order of 20 Hz.

A block diagram of the correlation network for a 1-Δ coherent DLL is presented below.



When the loop is locked:

$$Z_E(\epsilon,\theta) = 1/2\ V_{S_E} \pm V_{S_E}\ \frac{\epsilon}{\Delta} \pm V_{I_E} \cos\theta$$

$$Z_L(\epsilon,\theta) = 1/2\ V_{S_L} + V_{S_L}\ \frac{\epsilon}{\Delta} \pm V_{I_L} \cos\theta$$

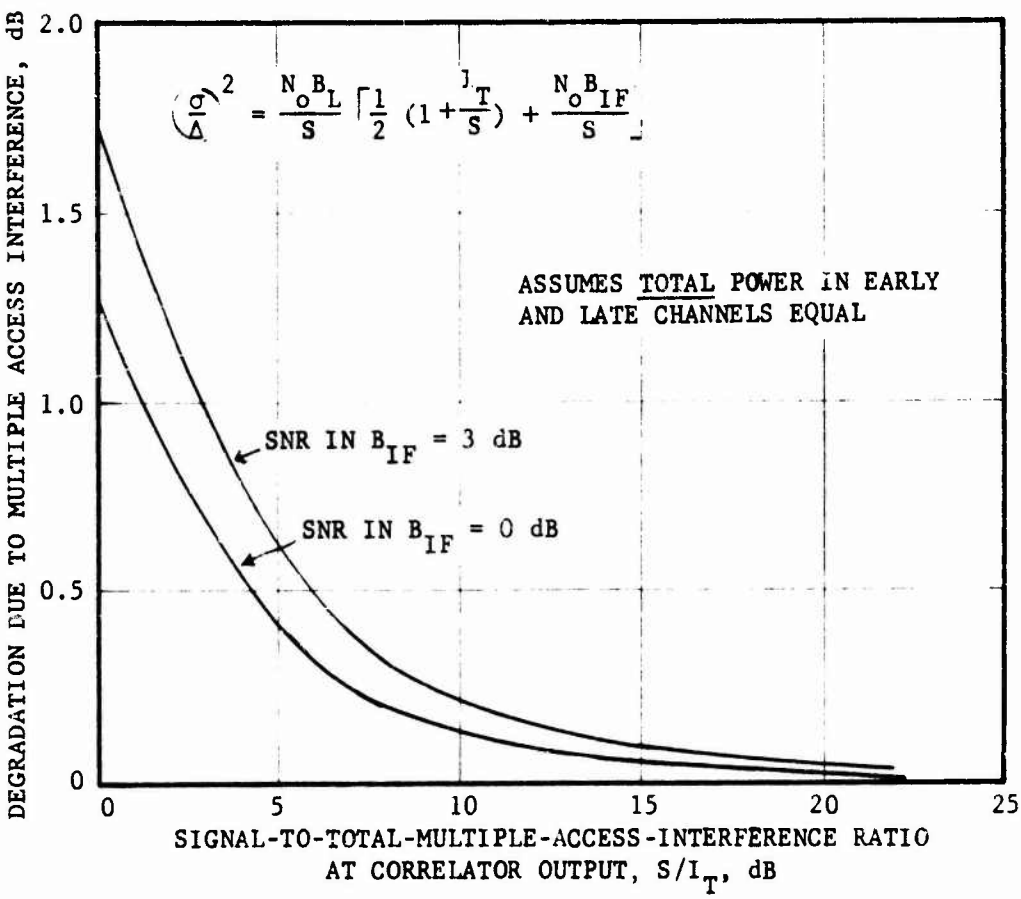


Figure 10 Multiple-Access Performance of Narrowband Noncoherent DLL

where θ is the carrier phase of the interference relative to the desired carrier

ϵ is the tracking error and $|\epsilon| \leq \Delta/2$

The early and later outputs are correlated for the desired code but are uncorrelated for the undesired code. Therefore,

$$V_{S_E} = V_{S_L} = V_S$$

but V_{I_E} does not always equal V_{I_L} since they are uncorrelated.

Thus,

$$E(\epsilon,\phi) = 2V_S \frac{\epsilon}{\Delta} \pm (V_{I_E} - V_{I_L}) \cos \phi$$

Therefore, the effect of multiple-access interference is to inject an offset voltage which results in a range offset. In addition, until the loop responds to this voltage offset, a smaller noise voltage can cause a loss of lock. This results in an increase in the DLL noise threshold.

The worst-case range offset occurs when either $V_{I_E} \gg V_{I_L}$ or $V_{I_L} \gg V_{I_E}$. For this case,

$$E(\epsilon, \phi) = \frac{2 V_S \epsilon}{\Delta} \pm V_I \cos \phi$$

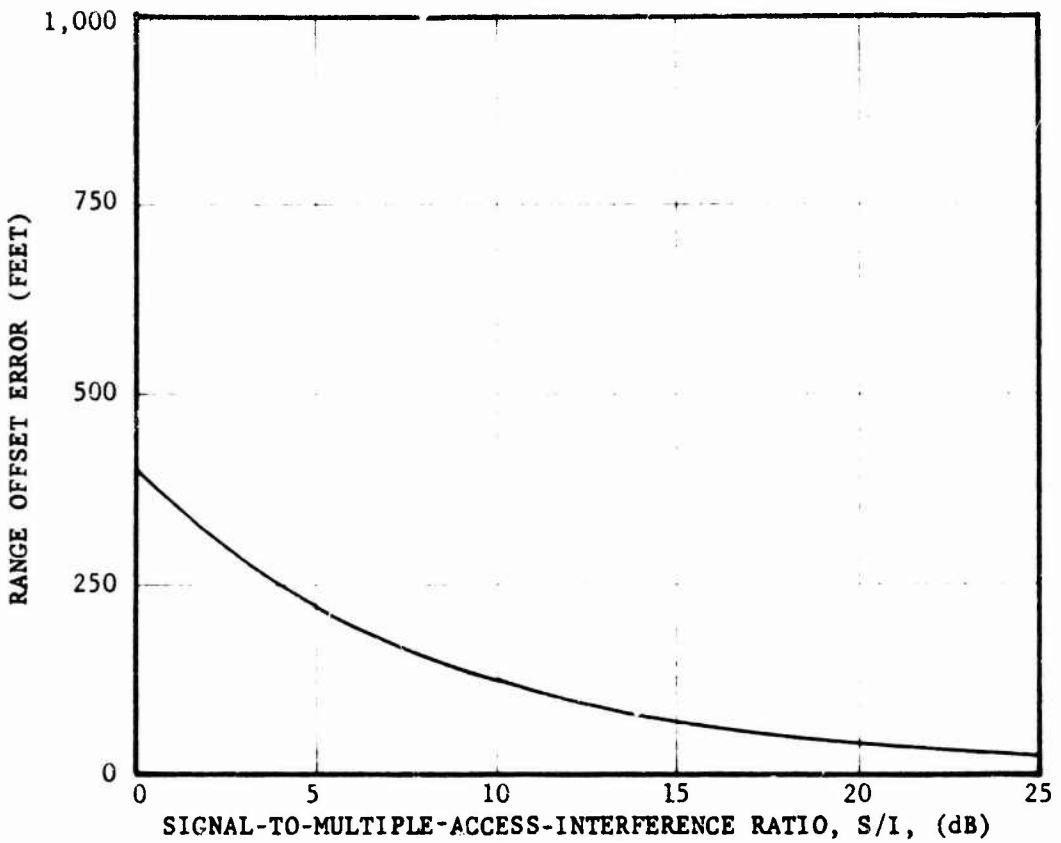
If ϕ is assumed to be random (that is, $p(\phi) = \frac{1}{\pi}$, $-\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2}$ *, then the expected value of $E(\epsilon)$ over ϕ is given by

$$E(\epsilon) = 2 V_S \frac{\epsilon}{\Delta} \pm \frac{2}{\pi} V_I$$

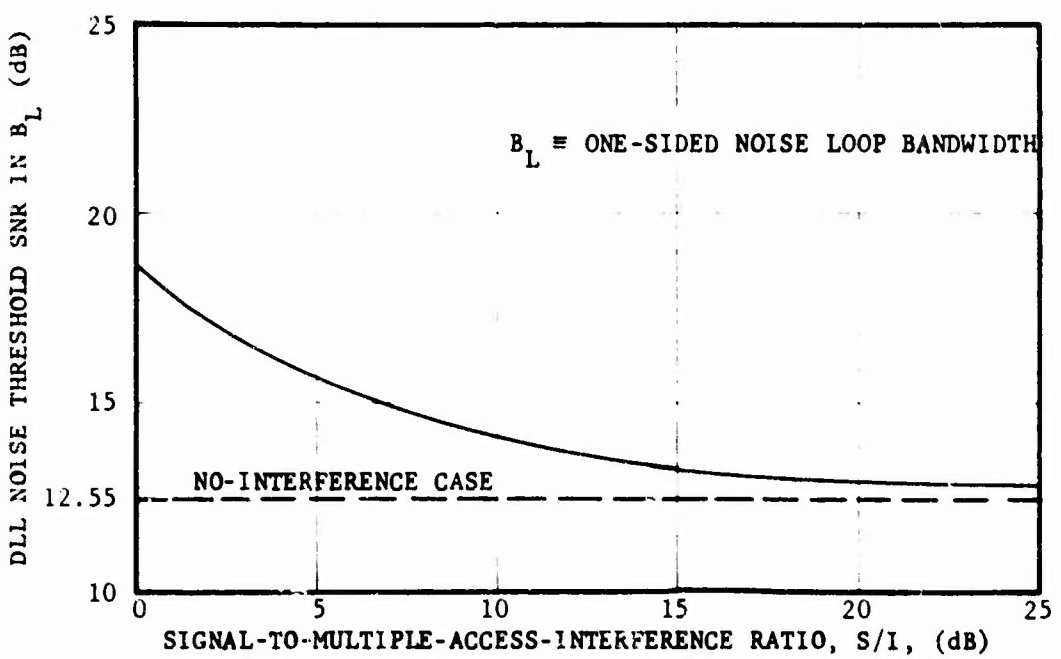
Figures 11 a and b present range offset error and DLL noise threshold, respectively, as a function of signal-to-interference ratio at the correlator output. DLL noise threshold is defined as the signal to noise in the loop bandwidth when the range offset plus the 3σ of the noise equals $\Delta/2$. Note that for a signal-to-interference ratio greater than 10 dB, the resulting range offset is less than 160 ft which is equal to 0.1 of a chip duration.

---

* The phase ϕ is assumed to be random between $-\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2}$ since both plus and minus values of $V_I \cos \phi$ result in a range offset.



(a) Range Offset

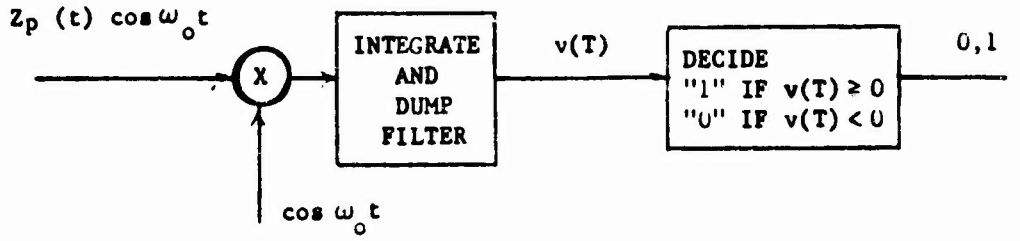


(b) DLL Noise Threshold

Figure 11 Multiple-Access Performance of Narrowband Coherent DLL

3.2.5 Coherent Data Detection

The performance of the coherent data detector will also be based upon one interfering signal since it is again reasonable to assume that, at any given instant of time, only one interfering signal is within the 50-Hz data bandwidth.



The decision variable, v(T), is given by

$$v(T) = \int_0^T (V_s + V_I \cos \phi)\, dt$$

$$= V_s T + V_I T \cos \phi = T (V_s + V_I \cos \phi)$$ where ϕ is assumed uniformly distributed from 0 to 2π.

The 0-to-2π range is used since the interference signal may increase the desired signal levels (both data bits are the same) or may decrease the desired signal levels (data bits are different). The desired and interfering signal are assumed to be in bit synchronization and at approximately the same carrier frequency or at a multiple of the code epoch rate.

The resulting bit error rate is given by

$$P_E = \frac{1}{2\pi} \int_0^{2\pi} P_E \left(\frac{ST}{N_o}, \frac{S}{I}, \phi \right) P(\phi)\, d\phi$$

where $P(\phi) = \frac{1}{2\pi}$; $0 \le \phi \le 2\pi$

$$P_E \left(\frac{ST}{N_o}, \frac{S}{I}, \phi \right) = \frac{1}{2} \operatorname{erfc} \sqrt{W \left(\frac{ST}{N_o}, \frac{S}{I}, \phi \right)}$$

$$W \left(\frac{ST}{N_o}, \frac{S}{I}, \phi \right) = \frac{ST}{N_o} + \frac{2\,ST}{N_o} \sqrt{\frac{I}{S}} \cos \phi + \frac{ST}{N_o} \left(\frac{I}{S} \right) \cos^2 \phi$$

Figure 12 presents the bit error rate as a function of ST/N_o for several values of S/I. Referring to Figure 12 , the increase in energy-to-noise density due to multiple-access interference will be 1 dB, or less, for a signal-to-interference ratio of 15 dB, or greater. Note that for a S/I of 10 dB, the required increase in energy-to-noise density is 2.3 dB.

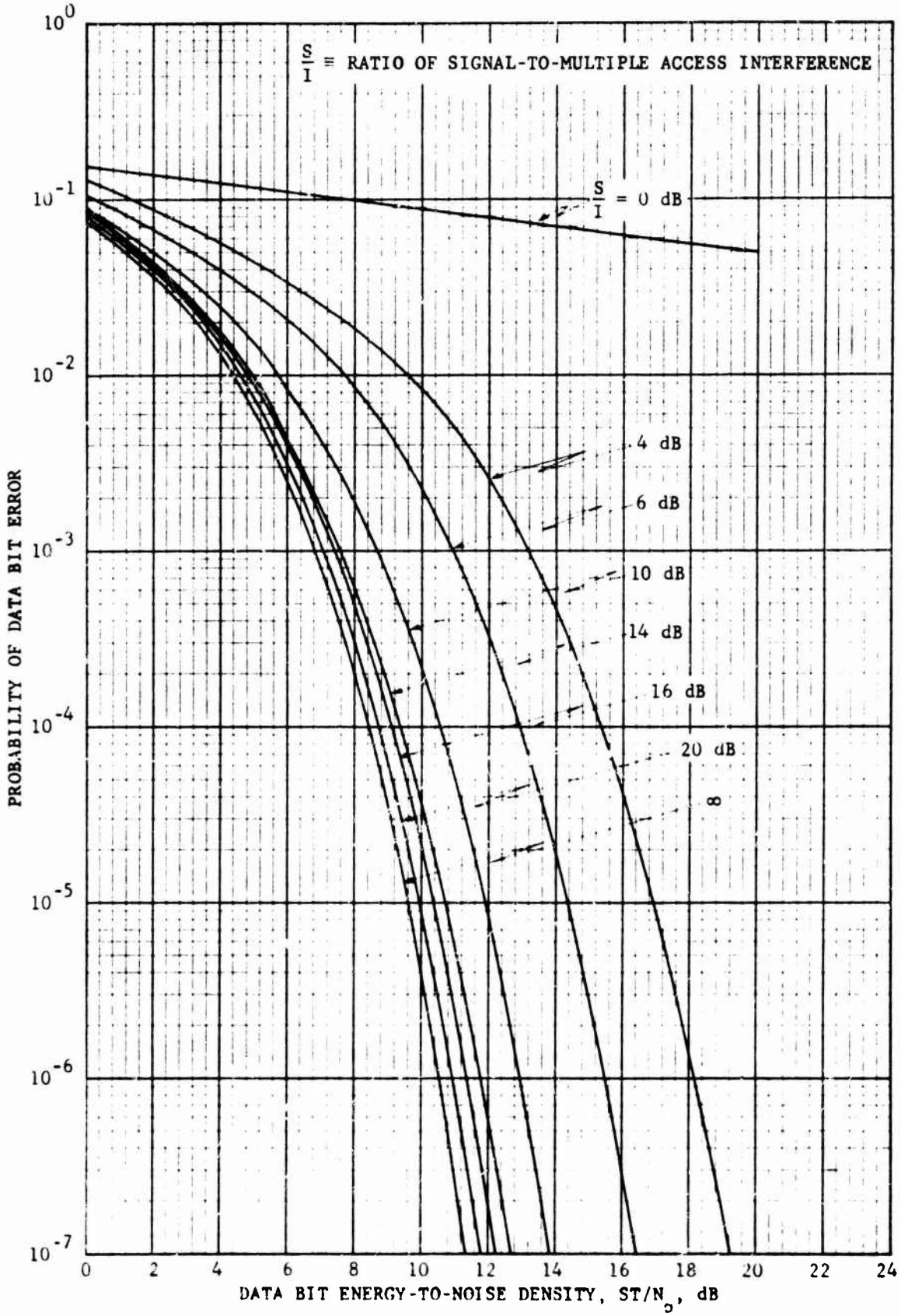


Figure 12 Multiple-Access Performance of Coherent Data Detector

3.2.6 Summary of Multiple-Access Performance

In the previous sections, the various performance parameters have been developed as a function of desired signal-to-multiple-access interference ratio at the correlator output, S/I. The ratio, S/I, is affected by correlation properties of the code and the relative levels between the desired signal and the undesired signal(s). That is,

$$\frac{S}{I} = \frac{1}{K^2} \left(\frac{S}{S_I} \right)$$

where S/S_I is the desired signal-to-interference power ratio at the correlator input

K is the cross-correlation voltage between any two Gold codes.

The probability distribution functions for K^2 (in dB) were presented for Gold code lengths of 511, 1023, and 2047, respectively. The maximum difference in signal level between the desired signal and any undesired signal is 11.2 dB. The various parameters which result in this 11.2 dB difference are summarized below.

Satellite Antenna Gain Plus Space Loss Difference	≤ 4.1 dB
Satellite EIRP Difference	≤ 2.6 dB
Propagation Loss Difference	≤ 1.5 dB
User Antenna Gain Difference	≤ 3.0 dB
TOTAL DIFFERENCE	≤11.2 dB

We are now in a position to evaluate S/I for various conditions. For coherent DLL tracking and coherent data detection, it is reasonable to assume that only one interfering signal is within the detection bandwidth at any given instant of time. Thus, at the maximum interference level and for the 1023 Gold code, the minimum signal-to-interference ratio is 9.6 dB (20.8-11.2) for coherent tracking and coherent data detection. This interference level only occurs if one of the undesired satellite signals is at a carrier frequency equal to a multiple of the code ephoch rate and 11.2 dB above the desired signal.

Rather than consider only the worst-case cross-correlation levels, let us determine the probability that the signal-to-interference ratio will be greater than some value from Figures 5, 6, and 7. Table 15 presents the values of S/I for probabilities of 90%, 95%, and 99% for Gold-code lengths of 511, 1023, and 2047. The maximum signal difference of 11.2 dB is assumed.

TABLE 15

S/I for 90%, 95%, and 99% Probabilities

CODE LENGTH	VALUE OF S/I WHICH IS EXCEEDED x% OF TIME		
	x = 90%	x = 95%	x = 99%
511	13.8 dB	12.4 dB	10.1 dB
1023	16.8 dB	15.4 dB	13.1 dB
2047	19.8 dB	18.4 dB	16.1 dB

For noncoherent DLL tracking and inlock detector performance, each satellite in view will contribute to the total interference. That is, we must determine the desired signal-to-total-interference ratio at the correlator output, S/I_T. The maximum number of undesired signals is ten. Thus,

$$I_T = \sum_{i=1}^{10} I_i$$

where I_i is the interference level at the correlator output from the i^{th} satellite.

For equal and independent I_i's,

$$E[I_T] = \sum_{i=1}^{10} E[I_i] = 10\, E[I]$$

$$Var[I_T] = \sum_{i=1}^{10} Var[I_i] = 10\, Var[I]$$

$E[I]$ and $\sqrt{Var[I]}$ can be estimated from Figures 5, 6, and 7. For example

$$E[I] \cong -30.5 \text{ dB} \quad \text{and} \quad \sqrt{Var[I]} \cong 4.5 \text{ dB}$$

Then

$$E[I_T] = -20.5 \text{ dB} \quad \text{and} \quad \sqrt{Var[I_T]} \cong 9.5 \text{ dB}$$

If the 11.2-dB power difference is included,

$$E[I_T] = -9.3 \text{ dB} \quad \text{and} \quad \sqrt{Var[I_T]} = 9.5 \text{ dB}$$

The probability distribution function of I_T can be estimated by assuming I_T to be Gaussian. Table 16 presents the values of S/I_T for probabilities of 90%, 95%, and 99% for Gold-code lengths of 511, 1023, and 2047.

TABLE 16

S/I_T for 90%, 95%, and 99% Probabilities

CODE LENGTH	VALUE OF S/I_T WHICH IS EXCEEDED x% OF TIME		
	x = 90%	x = 95%	x = 99%
511	8.9 dB	7.8 dB	6.1 dB
1023	11.9 dB	10.8 dB	9.1 dB
2047	12.9 dB*	11.8 dB*	10.1 dB*

* TOTAL INTERFERENCE IN 1-kHz BANDWIDTH

By combining Tables 15 and 16 with the various performance curves, a comparison of the 511, 1023 and 2047 Gold codes can be made. Table 17 presents a performance comparison of the three Gold-code lengths. 95% of the time, performance is equal to or better than that stated in Table 17.

TABLE 17

PERFORMANCE COMPARISON FOR 95% PROBABILITY

GOLD CODE LENGTH	PERFORMANCE PARAMETER				
	RANGING		DATA	ACQUISITION	
	COHERENT (OFFSET)	NONCOHERENT (DEGRADATION)	DEGRADATION IN E/N_o	RELATIVE INCREASE IN DWELL TIME	RELATIVE ACQUISITION TIME
511	100 ft.	0.5 dB	1.5 dB	1.22 - 1.30	1.00 - 1.07
1023	75 ft.	0.4 dB	1.0 dB	1.18 - 1.25	1.93 - 2.05
2047	50 ft.	0.3 dB	0.6 dB	1.15 - 1.20	3.77 - 3.94

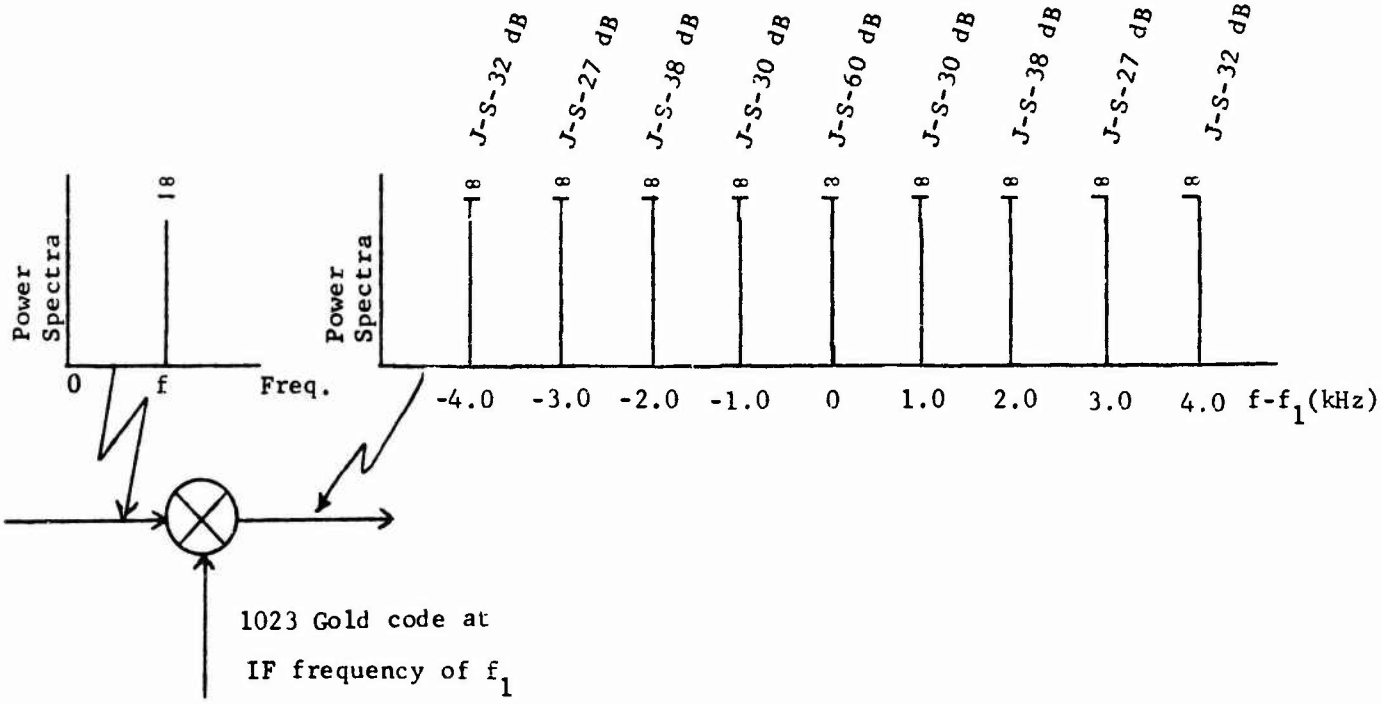
As expected, the performance improves as the Gold-code length is increased but so does code acquisition time since more code position must be searched. Referring to Table 17, the data degradation due to multiple-access interference is 1.0 dB for the 1023 Gold code.

4.0 C-SIGNAL JAMMING PERFORMANCE

The C-signal is not required to operate with intentional jamming as is the P-signal; however, there will always be unintentional jamming present. The effects of sinewave and "wideband" jammers are considered in this section.

4.1 Sinewave Jammer

For a sinewave jammer, the correlator output is just the convolution of the sinewave spectrum and the Gold-code spectrum. A typical correlator output spectrum is illustrated below.



where J is the total jammer power

S is the desired signal power

Referring to the above sketch, a sinewave jammer results in a line spectra at the correlator output. The power in each line will, in general, be different for each Gold code; however, there are at least 30 Gold codes of length 1023 which have a dc level of -60 dB. The average level of the other line spectra is approximately -30 dB since all the jamming power is contained in the 1022-line spectra about the dc component. Note that, as with multiple access interference, a jammer can only affect performance if his carrier frequency causes one of the line spectra to fall within the information bandwidth.

The effect of a sinewave jammer can be determined from the performance curves developed for multiple-access interference by redefining the signal-to-interference as follows:

$$\left(\frac{S}{I}\right)_J = \frac{J}{S} - 27 \qquad \text{for 511 Gold code}$$

$$= \frac{J}{S} - 30 \qquad \text{for 1023 Gold code}$$

$$= \frac{J}{S} - 33 \qquad \text{for 2047 Gold code}$$

where

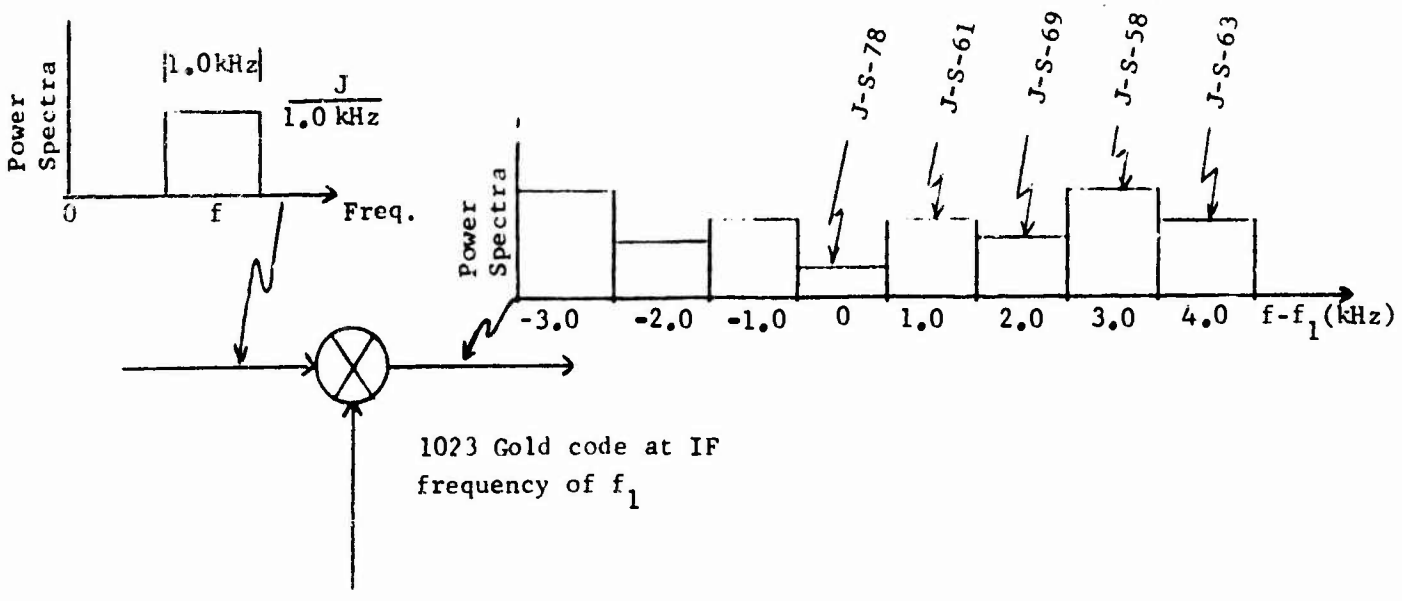
$\left(\frac{S}{I}\right)_J$ is the average signal-to-jammer ratio at correlator output in dB

$\frac{J}{S}$ is the jammer-to-signal ratios in dB

The average data degradation due to sinewave jamming can be evaluated by combining the above expressions and Figure 17. Figure 17 presents the degradation in E_b/N_o as a function of jammer-to-signal ratio for Gold-code lengths of 511, 1023 and 2047. Note that, for jammer-to-signal ratios less than 10 dB, the degradation is less than 0.6 dB.

4.2 "Wideband" Jammer

The preceding section assumed that the jammer had a line spectra; however, the unintentional jammer will have some bandwidth since he is transmitting some type of information. For the 1023 Gold code, a jammer bandwidth as small as 1.0 kHz can be considered a "wideband" jammer. A typical correlator output spectrum is sketched below.



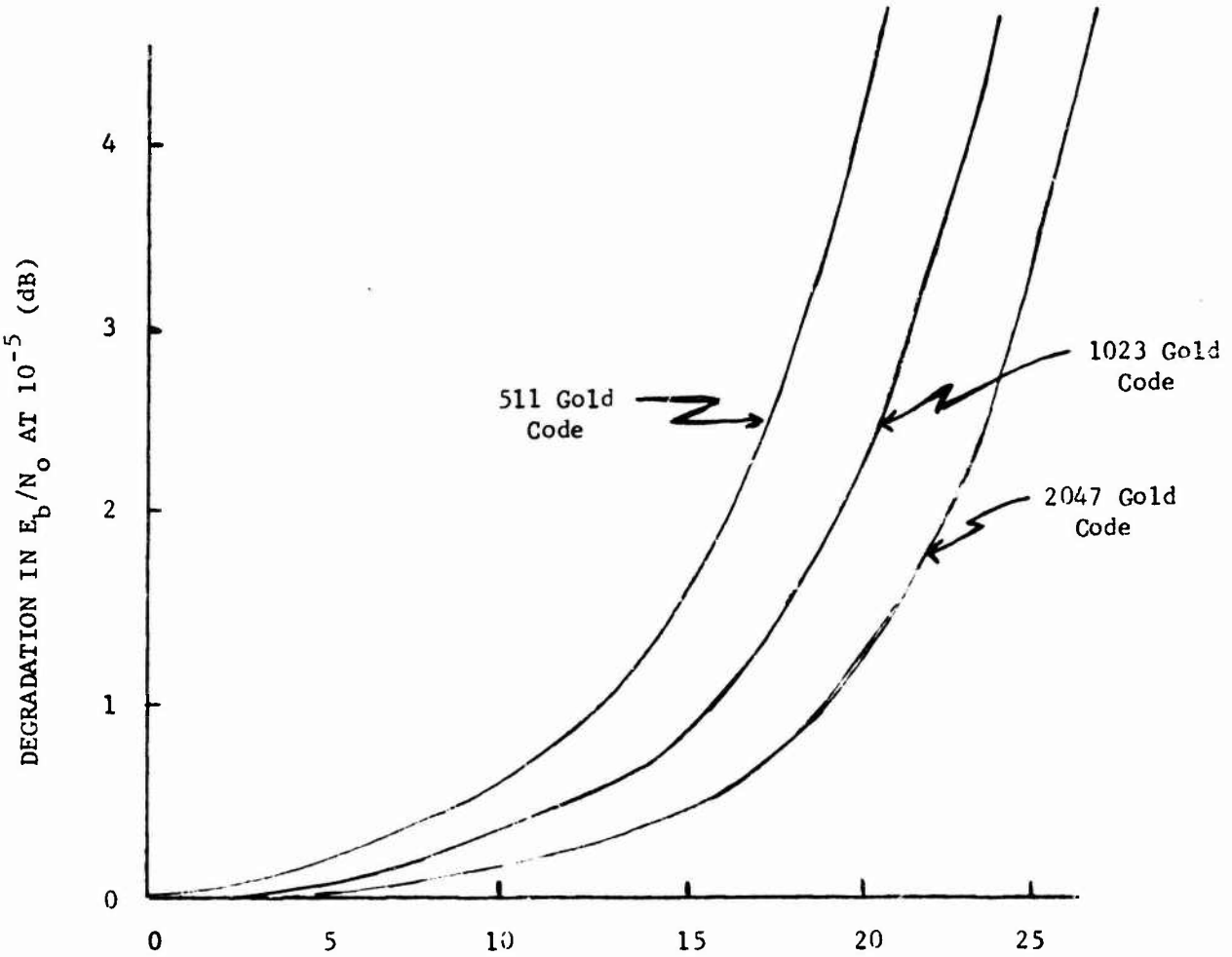


FIGURE 13 DATA DEGRADATION DUE TO SINEWAVE JAMMER

Referring to the above sketch, a 1.0-kHz bandwidth jammer results in an equivalent noise density given by

$$\left(N_o\right)_J = \frac{J}{S} - 30 - K^2$$

where K^2 is the relative Gold code line spectra causing the interference.

The worst-case value of K^2 is 20.8 dB, therefore,

$$\left(N_o\right)_J \leq \frac{J}{S} - 50.8$$

The largest bandwidth in the system is 1 kHz. Thus, the maximum noise due to jamming is

$$\left(N_o\right)_{B_{IF}} \leq \frac{J}{S} - 20.8$$

In order to affect performance, $(N_o)_J\ B_{IF}$ must be greater than -6 dB*, therefor, J/S must be greater than 14.8 dB before performance is affected.

For length 511 and 2047 Gold codes, the bandwidth of the jammer needs to be 2 kHz and 500 Hz, respectively, in order to be considered a "wideband" jammer. In addition, the worst-case jammer noise density would be increased by 3 dB and reduced 3 dB for the 511 and 2047 Gold-code lengths, respectively.

In summary, the 1023 Gold code can tolerate, as a minimum, a jammer level 13 dB greater than the signal level with essentially no loss in performance. The 511 and 2047 Gold codes can tolerate a jammer level 10 dB and 16 dB greater than the signal level, respectively.

---

*Assume a signal-to-noise ratio of 3 dB in the IF bandwidth.

5.0 P-Signal Jamming Performance

The jamming performance of the P-signal nas been analyzed (Reference 5) and experimentally measured (Reference 6) by Philco-Ford on previous contracts for SAMSO. The following summarizes the basic performance predicted for the P-signal receiver.

The equivalent C/N_o under jamming conditions can be expressed as:

$$Q = \frac{C}{N_o + J\Delta} \qquad (1)$$

where

Q = reduced C/N_o with jamming included

C = received signal power (dBW)

N_o = system noise density at receiver input (dBW/Hz)

J = incident jamming level (dBW)

Δ = chip time = 10^{-7} seconds (10 Mbps)

Solving for C

$$C = (N_o + J\Delta)Q \qquad (2)$$

For a coherent code tracking loop

$$\frac{\sigma}{\Delta} = \sqrt{\frac{B_L}{2Q}} \qquad (3)$$

where

σ = RMS jitter of the code in the loop

B_L = One-sided noise bandwidth of the loop

Experience has shown that the loop will break lock in the vicinity of $\frac{\sigma}{\Delta} = 0.15$. To be conservative, we assumed a value of $\frac{\sigma}{\Delta} = 0.12$. For this analysis, the loop bandwidth is assumed to be 3 Hz. From (3) a value of Q = 100 (20 dB/Hz) is computed. This value of Q is defined as a minimum value of C/N_o below which the performance of the user equipment is no longer acceptable.

We assume $N_o = 0.8 \times 10^{-20}$ watts/Hz nominal and 1.0×10^{-20} watts worst case. Substituting these values into equation (2) with Q = 100, $\Delta = 10^{-7}$ sec, and different values of J, values of required received signal

power are derived. Equating these values with Line 10 of Table 1, Repeat #2 of this Volume, plus Lines 1 and 2 of Table 2, we obtain the satellite transmitter power, P, required to develop Q = 100 in the code tracking loop. The results are plotted in Figure 14. Note that in the absence of jamming, from Line 5 of Table 2, a satellite power of 0.27 watts nominal and 1.9 watts worst case is required to obtain Q = 100 (C/N_o = 20 dB/Hz). The curves in Figure 14 are asymptotic to these values of transmitter power.

Figure 15 relates the received jamming power, J, to three jamming threats. The received jamming powers are indicated for a 25 watt satellite transmitter at L_1 for the P-signal.

It should be noted that improved jamming protection can be obtained by decreasing B_L. Techniques for doing this require more complex user equipments, however.

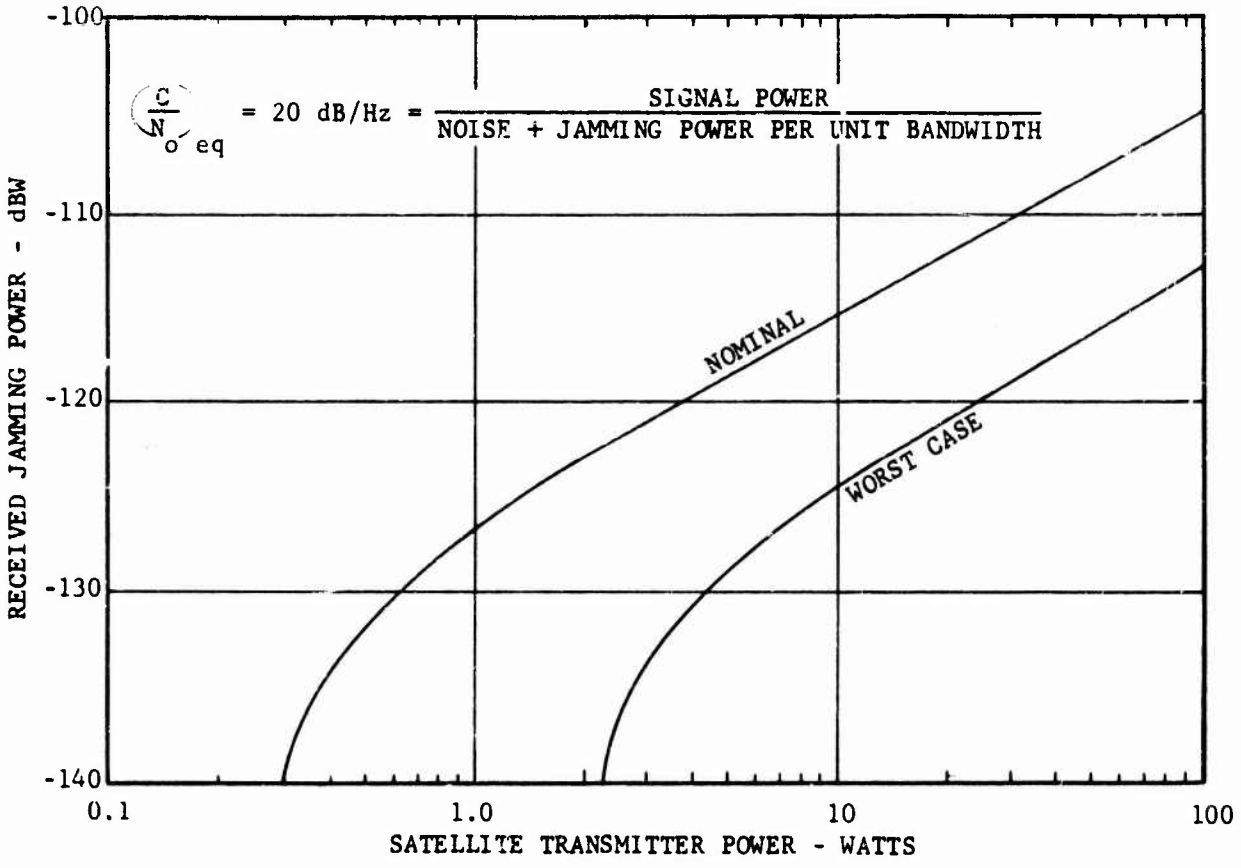


Figure 14 Required Satellite Transmitter Power vs. Jamming Threat

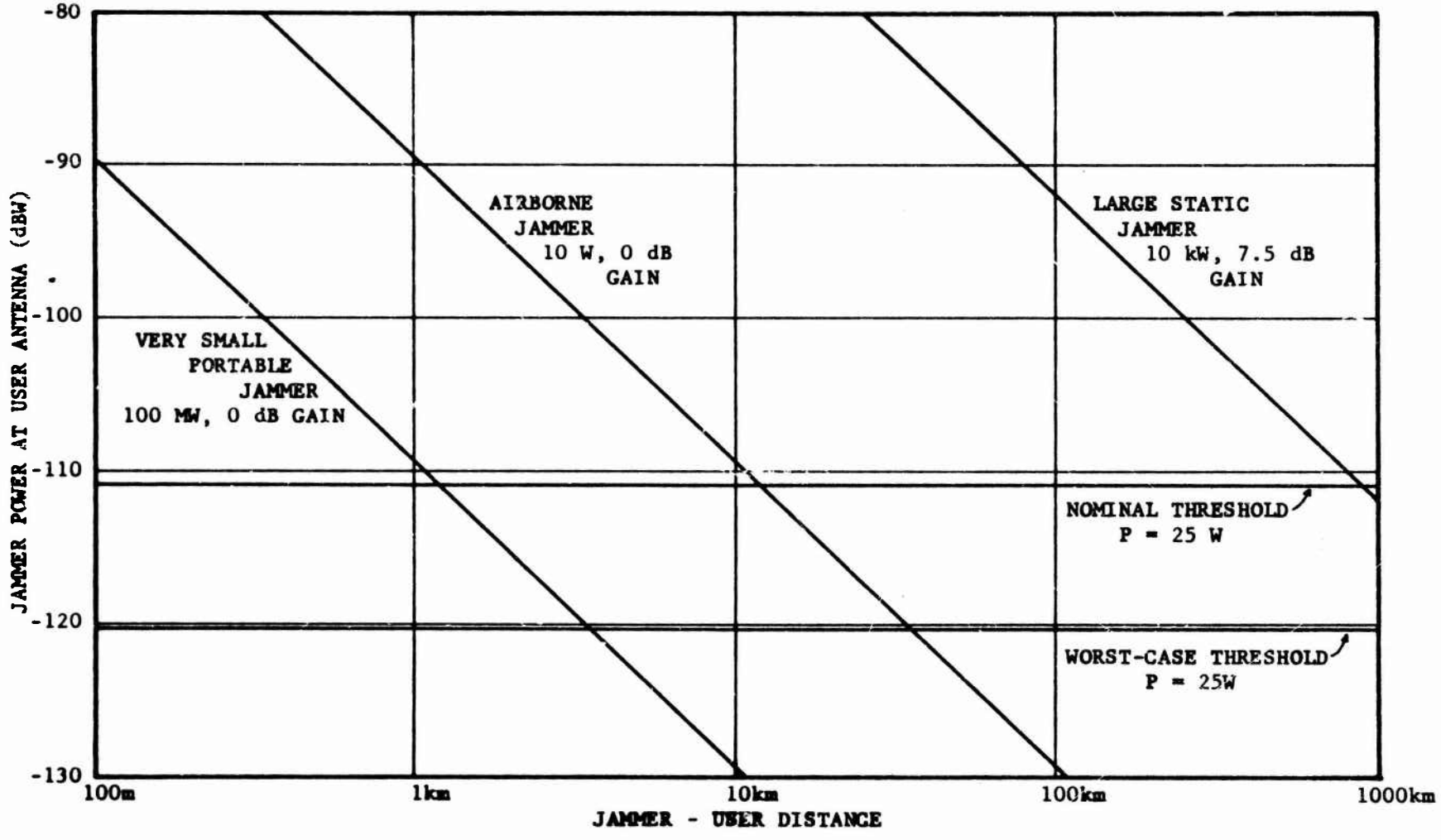


Figure 15 Jamming Power

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

REFERENCES

1. M. Deckett, "Analysis and Simulation of the C-Code Acquisition Time", TRW Memo No. DNSDP-MD-160, 9 November 1973

2. R. Gold, "Maximal Recursive Sequences with 3-Valued Recursive Cross Correlation Functions," IEEE Transactions on Information Theory, January 1968, pp. 154-156.

3. Davenport and Root "Random Signals and Noise", McGraw-Hill, 1958.

4. W. Gill, "A Comparison of Binary Delay-Lock Tracking-Loop Implementations", IEEE Trans. on Aerospace and Electronic Systems, July 1966, pp. 415-424.

5. "C-Band Signal and Reception Techniques Study", Philco-Ford, SAMSO-TR-70-434, January 1971, (CONF)

6. "System 621-B Noncoherent Receiver Development and Laboratory Test Program, Philco-Ford, SAMSO-TR-85, June 1972, (CONF)

Report No. 2

GPS LINK ANALYSIS

CONTENTS

SECTION		PAGE
1.0	Scope	2-1
2.0	Link Budgets	2-1
3.0	Comparison with GPS Specification	2-12
4.0	Alternative Data Rates	2-12
5.0	Telemetry and Command Link Analysis (SGLS)	2-16
Appendix		
A	Geometry Considerations for GPS	2-21

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Variation of C/N_o With Elevation Angle	2-11
2	Command and Telemetry Link	2-18

TABLES

TABLES		PAGE
1	Reference Link For PRN Ranging Signal at L_1	2-3
2	Received Power Requirement for User Receiver at L_1	2-4
3	Reference Link for PRN Ranging Signal at L_2	2-7
4	Received Power Requirement for User Receiver at L_2	2-9
5	Summary of Link Analysis	2-13
6	Comparison of Received Signal Level Requirements	2-14
7	Performance at GPS Specification Received Signal Levels	2-14
8	TT and C Requirements	2-16
9	Uplink SGLS S-Band Calculation (1700 MHz)	2-19
10	Downlink SGLS S-Band Calculation (2250 MHz)	2-20

GPS LINK ANALYSIS

1.0 SCOPE

This report presents the RF link analysis for the PRN ranging signals for the GPS. In addition, in Section 5, the link budgets for the S-Band TT&C (SGLS) are presented.

The objective of the analysis is to determine satellite, user, and ground terminal transmitted power (EIRP) and receiver sensitivity specifications which will meet the objectives of the GPS. These parameters will be used in specifying the interface between the control, satellite, and user segments.

Key issues which will be addressed in this report are:

1. What margin shall be included in the link budget?
2. How shall the available satellite transmitter power be distributed between the C/A and P-signals, and between the L_1 and L_2 band?

2.0 LINK BUDGETS FOR L_1 AND L_2

Tables 1 through 4 present link budgets for the GPS PRN signals. In these tables a most probable value of each item is calculated or estimated as described in the comments following each table. We also calculate or estimate a worst case tolerance for each item. The statistics of these tolerances are not generally precisely known but we attempt to estimate them such that the probability of the parameter exceeding the worst case value is small (i.e., 5%).

In the computation of received signal power or C/N_o, a worst case tolerance is derived by simply adding arithmetically the worst case tolerances of each of the individual parameters. The resultant worst case tolerance is the link margin. The tolerances aren't computed as the RMS sum of the individual tolerances because the tolerances aren't necessarily independent. To be conservative, the arithmetic sum is used. Therefore, the probability of exceeding the total margin should be quite small.

The link budgets are presented with the satellite transmitter power, P, carried as an unknown. In Tables 2 and 4, a value is derived for P which will be sufficient to generate the required C/N_o (carrier-to-noise power density ratio at the receiver input), plus the sum of worst case tolerances. Thus the link is designed with a margin equal to the sum of the worst case tolerances.

The link budgets presented in Tables 1 and 3 are for a user-to-satellite elevation angle of 10 degrees. An allowance is provided (Line 9) for the satellite at a 5 degree elevation. A 10 degree elevation was selected for the nominal link because: (1) elevation angles will generally be low, less than 30 degrees 50% of the time, (2) performance below 10 degrees will generally be poor, due to poor user antenna performance, multipath interference from nearby structures, and increased propagation loss.

Figure 1 shows how the C/N_o varies with elevation angle. The top curve shows the effect of increased path loss (due to increased slant range, and reduction in satellite antenna gain toward the user as the satellite-to-user path moves off the antenna axis. The lower curve includes a 3 dB user antenna gain variation (0 dB overhead to -3 dB at 5 degree elevation). The lower curve is also displaced by the sum of worst case tolerances.

Table 1 provides the budget for the satellite-to-user link at L_1. Table 3 provides the same thing at L_2. Given a value for P, the satellite transmitter power, the satellite EIRP (Line 6), and the received signal power at the user (Line 10) are specified. These two parameters are sufficient to specify the satellite and user segments of the RF link.

(Strictly speaking, the satellite EIRP should be specified as a function of the angle off the earth vertical since a 1.94 dB variation in path loss exists between a user at 90^o elevation and a user at 5^o elevation. See Appendix A.)

Tables 2 and 4 contain the user receiver portion of the link budgets at L_1 and L_2. Separate C/N_o requirements exist for data reception, range measurement, and signal acquisition. The signal acquisition requirement applies only to the C/A signal.

TABLE 1

REFERENCE LINK FOR PRN RANGING SIGNAL AT L_1

Satellite Elevation Angle = 10°

	ITEM	NOMINAL	WORST CASE TOLERANCE
1.	Satellite RF Transmit Power (dBW)	P	-1.0
2.	Satellite Line Losses (dB)	-0.5	-0.4
3.	Satellite BPF Correlation Loss (dB)	-0.3	-0.3
4.	Satellite Antenna Gain Toward User (dB)	12.4	-0.5
5.	Satellite Antenna Polarization Loss (dB)	-0.1	-0.1
6.	Satellite EIRP (dBW)	P-11.5	-2.3
7.	Path Loss (f = 1575 MHz) (dB)	-184.3	-0.0
8.	Propagation Loss (dB)	-2.0	-1.0
9.	Tolerance for 5° elevation (dB)	----	-0.3
10.	Received Signal Power (dBW)	P-174.8	-3.6

NOTES

LINE

1. Worst case over on-orbit temperature range at end-of-life.
2. Design uncertainty
3. To be calculated
4. Peak Gain = 15.4 dB
   Half-Power Beamwidth = 27.6 degrees
   Aperture Efficiency = 55%
   User Elevation Angle = 10 degrees
   Antenna Pointing Error 0 degree nominal, 1 degree worst case
5. Axial ratio less than 3 dB over satellite antenna half-power beamwidth
6. Lines 1 + 2 + 3 + 4 + 5
7. $L = 37.8 + 20 \log F \text{ MHz} + 20 \log R_{NM}$

   $F = 1575$ MHz

   $R = 13340$ NM
8. Losses due to scintillation atmospheric attenuation and multipath. Nominal value unknown; 2.0 dB assumed. 3 dB worst case from SAMSO/ Aerospace
9. Calculated with 1° pointing error
10. Lines 6 + 7 + 8 + 9 received power at output of 0 dB gain circular polarized antenna.

TABLE 2

RECEIVED POWER REQUIREMENT FOR USER RECEIVER AT L_1

ITEM	NOMINAL	WORST CASE TOLERANCE
1. Antenna Pointing/Polarization Loss - dB	0	-3.0
2. Line Loss - dB	-1.0	-0.5
3. System Noise Power Density dBW/Hz	-200.2	+0.6
4. Received Power at Receiver - dBW	P-175.8	-7.1
5. Carrier/Noise Density - dB/Hz	P+24.4	-7.7
Data Reception		
6. Data Rate (50 bps)	17.0	----
7. E_b/N_o $(P_e = 10^{-5})$	9.8	----
8. Demodulation Loss	2.0	-0.3
9. Multiple Access Loss *	1.0	-0.2
10. Required C/N_o - dB/Hz **	(29.8)	+0.5
11. Margin - Data Reception - dB	P-5.4	-8.2
12. Satellite Power - dBW (Margin = 8.2 dB)	13.6 (22.9W)	
13. Minimum Receive Power - dBW	-164.8	
Range Measurement		
14. Required C/N_o - dB/Hz	(27.8)	-1.0
15. Margin - Range Measurement	P-3.4	-8.7
16. Satellite Power - dBW (Margin = 8.7 dB)	12.1 (16.2W)	
17. Minimum Receive Power - dBW	-166.3	
Signal Acquisition		
18. Required C/N_o - dB/Hz	(32.0)	----
19. Margin Acquisition	P-7.6	-7.7
20. Satellite Power - dBW (Margin = 7.7 dB)	15.3 (33.9)	
21. Minimum Receive Power - dBW	-163.1	

* Value for C/A-Signal. Value for P-Signal ≅ 0.18

** Value for P-Signal. Value for P-Signal = 28.8 dB-Hz

TABLE 2 COMMENTS

LINE

1. Antenna gain is measured relative to circular polarized wave. Off-angle pointing loss is caused by pattern directivity and axial ratio variations.

2. Antenna coupling and line losses between feed and receiver preamp.

3. System noise temperature referred to receiver preamp input.

$T = LT_A + (1-L)\,T_L + T_R$

$T_R = (F-1)\ 290^{\circ}K$

F = Receiver noise figure (including bandpass filter and noise contribution of preamp and subsequent stages)

$T_A = 130^{\circ}K$ - A conservative estimate of antenna temperature of hemispheric beam looking upward with some radiation from earth coming in via sidelobes. (More data needed on this.)

L = -1.0 dB = 0.8 nominal

= -1.5 dB = 0.7 worst case

$T_L = 290^{\circ}K$ = temperature of antenna coupling circuit and line.

F = 4.5 dB nominal

= 5.0 dB worst case

Receiver Noise Density = N_o

$N_o = KT$

$K = -228.6$ dBW/Hz - $^{\circ}K$

4. Line 10 of Table 1, plus Lines 1 and 2 of Table 2.

5. Line 4 - Line 3

7. Direct Carrier Bi-Phase Modulation (PSK) with differential encoding

8. Includes losses due to carrier and bit sync jitter, plus filter detector mismatch and tau-dither loss of 1.3 dB.

9. See signal structure analysis. (Correlation loss caused by receiver negligible as RF band limiting is done in satellite. See Line 3 of Table 1.)

TABLE 2 COMMENTS (Continued)

LINE

10. Lines 6 + 7 + 8 + 9

11. Line 5 - Line 10

12. P-5.4 = +8.2

13. Line 10 of Table 1 with P = 13.6 dBW

14. For P-Signal: σ = 5ft with 4g's acceleration

15. Line 5 - line 14

16. P - 3.4 = 8.7

17. Line 10 of Table 1 with P = 12.1 dBW

18. Based on C/A-code acquisition time = 50 sec. with 90% probability and two acquisition circuits.
See Report No. 1 of this volume resulting time-to-first fix. (TTTF)

19. Line 5 - Line 18

20. P - 7.6 = 7.7

21. Line 10 of Table 1 with P = 15.3 dBW

TABLE 3

REFERENCE LINK FOR PRN RANGING SIGNAL AT L_2

User Elevation Angle = 10^{o}

	ITEM	NOMINAL	WORST CASE TOLERANCE
1.	Satellite RF Transmit Power (dBW)	P	-1.0
2.	Satellite Line Losses (dB)	-0.5	-0.4
3.	Satellite BPF Correlation Loss (dB)	-0.3	-0.3
4.	Satellite Antenna Gain Toward User (dB)	10.9	-0.2
5.	Satellite Antenna Polarization Loss (dB)	-0.1	-0.1
6.	Satellite EIRP (dBW)	P +10.0	-2.0
7.	Path Loss (f = 1232.5 MHz)	-182.1	-0.0
8.	Propagation Loss (dB)	-2.0	-1.0
9.	Tolerance for 5^{o} Elevation (dB)	----	-0.3
10.	Received Signal Power (dBW)	P-174.1	-3.3

TABLE 3 COMMENTS

LINE

1. Same as Table 1

2. Same as Table 1

3. Same as Table 1

4. Peak Gain = $15.4 - 20 \log \left(\frac{1575}{1232.5} \right)$

   $= 15.4 - 2.2 = 13.2$ dB

   User Elevation Angle = 10°

   Antenna Pointing Error $\approx$ 0° nominal, 1° worst case

5. Same as Table 1

6. Same as Table 1

7. L = Space Loss at L_1 - $20 \log \left(\frac{1575}{1232.5} \right)$

8. Same as Table 1

9. Same as Table 1

10. Same as Table 1

TABLE 4

RECEIVED POWER REQUIREMENT FOR USER RECEIVER AT L_2

	ITEM	NOMINAL		WORST CASE TOLERANCE
1.	Same as Table 2			
2.	Same as Table 2			
3.	Same as Table 2			
4.	Received Power at Receiver - dBW	P -175.1		-6.8
5.	Carrier/Noise Density - dB/Hz	P + 25.1		-7.4
	Data Reception			
6.	Same as Table 2			
7.	Same as Table 2			
8.	Same as Table 2			
9.	Same as Table 2			
10.	Required C/N_o - dB/Hz	28.8		+0.5
11.	Margin - Data Reception	P -3.7		-7.9
12.	Satellite Power - dBW (Margin = 7.9 dB)		11.6(14.5W)	
13.	Minimum Receive Signal Power - dBW		-165.8	
	Range Measurement			
14.	Required C/N_o - dB/Hz	27.8		-1.0
15.	Margin - Range Measurement	P -2.7		-8.4
16.	Satellite Power - dBW (Margin = 8.4 dB)		11.1(12.9W)	
17.	Minimum Receive Signal Power - dBW		-166.3	

TABLE 4 COMMENTS

LINE

4. Line 10 of Table 3, plus Lines 1 and 2 of Table 4 (Table 2)

5. Line 4 - Line 3

10. From Line 10 of Table 2

11. Line 5 - Line 10

12. P - 4.7 = 7.9

13. Line 10 of Table 3 with P = 12.6 dBW

14. Same as Line 14 of Table 2

15. Line 5 - Line 14

16. P-2.7 = 8.4

17. Line 10 of Table 3 with P = 11.1 dBW

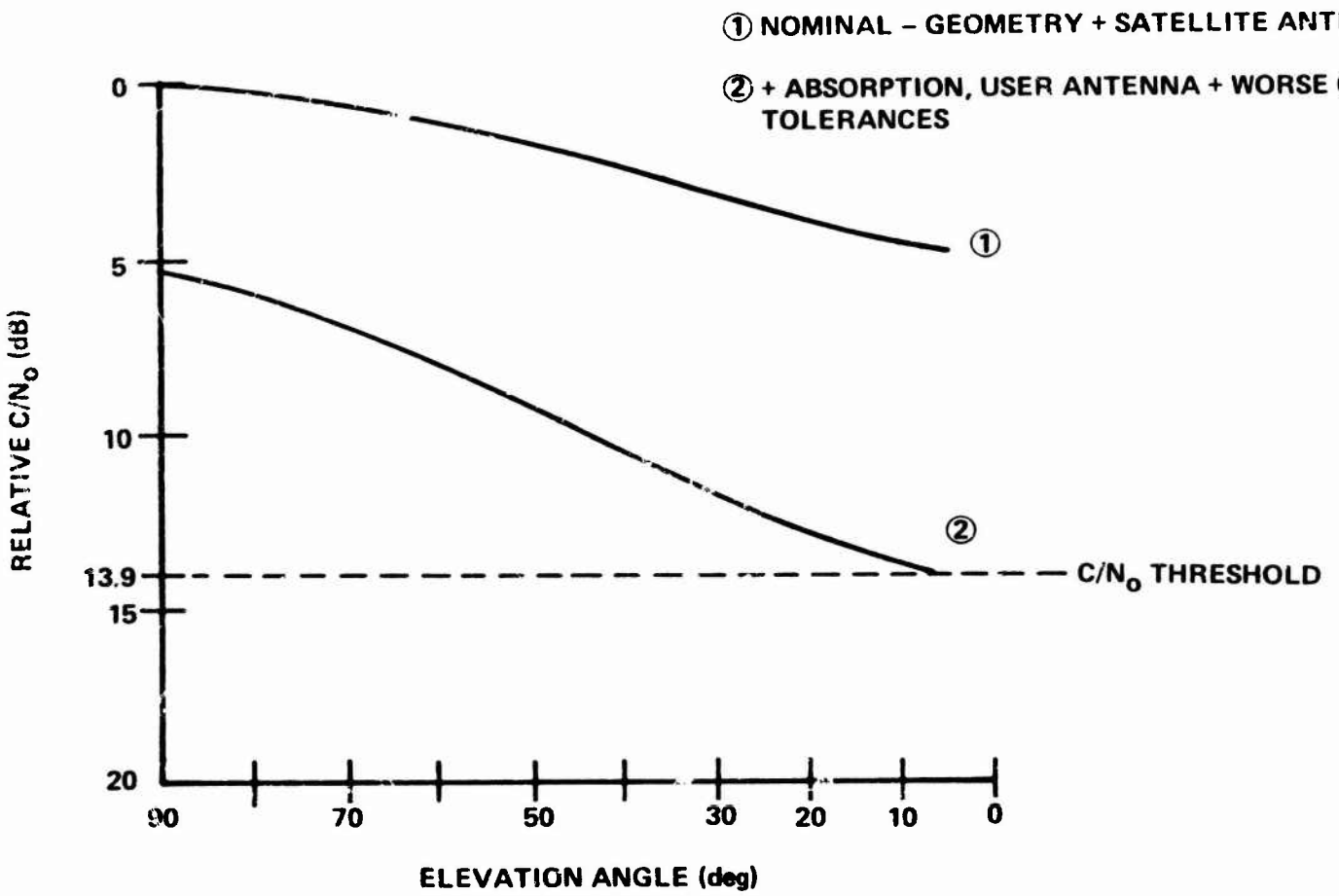


Figure 1 Variation of C/N$_o$ with Elevation Angle

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

Finally, a value for required satellite transmitter power and receive signal power is derived for each case. The case requiring the maximum satellite power is the controlling requirement used in specifying the link. These results are summarized in Table 5. Note that the satellite power requirements are based upon Table 5, a total of 69.7 watts of RF power must be generated by the satellite to perform the L_1 and L_2 functions.

3.0 COMPARISON WITH GPS SPECIFICATIONS

Table 6 presents a comparison of the required minimum receive signal levels to meet the desired performance and the received signal levels from the GPS specification document SS-GPS-101A, dated 29 January 1974. Note that in all cases the required minimum levels are below the specified levels. Table 7 presents the resulting performance of a user receiver at the GPS specification levels. Alternatively, the performance of the user equipment could be degraded by the amount shown under "excess margin" in Table 6 to reduce user equipment cost.

4.0 ALTERNATIVE DATA RATES

Given the GPS minimum received signal powers specified levels shown in Table 7, a data rate of 166.6 bps can be supported over the C/A signal channel. The advantage of the higher data rate is a significant reduction in acquisition time and a greater capacity to handle special data messages, such as remote rekeying data. Furthermore, the need for Manchester coding to resolve ambiguities is eliminated. Comparing line 21 of Table 1 with the first line of Table 7, we note an increase in received signal power of 3.1 dB. This increase leads to an increase of required C/N_o to 35.1 dB-Hz (line 18 of Table 2 plus 3.1 dB). An increase of data rate from 50 bps to 166.6 bps increases the required C/N_o for data reception from 29.8 dB-Hz to 35.0 dB-Hz (line 10 of Table 2). These numbers apply to the C/A signal.

TABLE 5

SUMMARY OF LINK ANALYSIS

LINK	CONTROLLING REQUIREMENT	REQUIRED SATELLITE POWER - WATTS	REQUIRED MINIMUM RECEIVE SIGNAL POWER-dBW	REFERENCE
L_1 - C/A	Code Acquisition Time ≤ 50 Sec*	33.9 (Margin = 7.7 dB)	-163.1	Table 2 Lines 20,21
L_1 - P	BER < 10^{-5}	22.9 (Margin = 8.7 dB)	-164.8	Table 2 Lines 12,13
L_2 - P	1 σ Range Error 5 Feet**	12.9 (Margin = 8.4 dB)	-166.3	Table 4 Lines 16,17

* Assumes Dual Acquistion Circuits

** Assumes 10^{-5} Ber Not Required

TABLE- 6 - COMPARISON OF RECEIVED SIGNAL LEVEL REQUIREMENTS

LINK	REQUIRED MINIMUM RECEIVE SIGNAL LEVELS (dBW)	RECEIVE SIGNAL LEVELS FROM GPS SYSTEM SPEC. (dB)	EXCESS MARGIN (dB)
L_1-C/A	-163.1	-160.0	+3.1
L_1-P	-164.8	-163.0	+1.8
L_2	-166.3	-166.0	+0.3

TABLE 7 - PERFORMANCE AT GPS SPECIFICATION

RECEIVED SIGNAL LEVELS

LINK	RECEIVED SIGNAL LEVELS FROM GPS SPEC. (dBW)	PERFORMANCE
L_1 - C/A	-160.0	(a) CODE ACQ. TIME $\simeq$ 50 sec with single acquisition circuit
		(b) CODE ACQ. TIME $\simeq$ 25 sec with dual acquisition circuit
L_1 - P	-163.0	BIT ERROR RATE $\simeq 10^{-7}$ at 50 bps
L_2 - P	-166.0	1 σ RANGE ERROR $\simeq$ 5 feet

The data rate for the P-signal might be reduced to 10 bps. This would reduce the required C/N_o from 28.8 dB-Hz (@ 50 bps) to 21.8 dB-Hz (line 10 of Table 2). This reduction in C/N_o is advantageous when operating in a jamming environment. In Section 5, Report 1 of this volume, the P-signal jamming performance was estimated on the basis of a C/N_o = 20 dB-Hz threshold. At 10 bps, a bit error probability of 10^{-3} is achievable at this threshold, which still may be adequate for useful navigation data transmission. Another advantage of 10 bps is the reduction in spectrum width (after code decorrelation) permits a narrower IF bandwidth which improves the jamming performance of some types of receivers. These alternative data rates should be investigated further.

Pretty damn good idea!

5.0 Telemetry and Command Link Analysis (SGLS)

The GPS satellite will receive and transmit S-band SGLS compatible command, telemetry, and ranging support from the Remote Tracking Stations (RTS) of the SCF network. This section calculates performance so that satellite characteristics can be defined which will meet system specifications. Table 8 presents the TT and C requirements. Downlink margin may have two values as interpreted by Philco Ford; 6 dB over the 27° earth coverage angle and 0 dB over the remainder of the sphere in the event the satellite is not in stable orbit. Other values in the table are self-explanatory. In all the calculations that follow it is assumed that only one baseband function occurs at a time. For example, command and ranging or telemetry and ranging will not occur simultaneously.

TABLE 8

TT AND C REQUIREMENTS

Item	Uplink	Downlink
Information	• Housekeeping Commands • Navigation Updates • Ranging	• Housekeeping Telemetry • Navigation Update Verification • Ranging
Data	• 3 Tone FSK - 1000 b/s • 1 Mb/s Ranging Code	• 1.024 MHz Subcarrier - 250 b/s • 1 Mb/s Ranging Code
Performance	• $P_e = 10^{-5}$ • σ_r = 60 feet	• $P_e = 10^{-5}$ • σ_r = 60 feet
RF Power (RTS)	• 1000 watts	Not applicable
Antenna (RTS)	• 15 feet (T_x)	• 15 feet (R_x)
Margin	• 6 dB	• 6 dB - over 27° coverage • 0 dB - remainder of sphere
Minimum Elevation Angle	• 5 degrees	• 5 degrees

Figure 2 presents a summary of the link performance. The uplink command and ranging are transmitting individually from the 15 foot RTS with 1 kilowatt of RF power. A minimum margin of 9.5 dB results, which more than satisfies the specification. In an emergency mode where the satellite attitude may not be stable, antenna gain is reduced to -6 dB, thus reducing margin to approximately 3 dB. If 6 dB of margin (under all conditions) were a firm requirement, the RTS transmitter power could be raised to 2000 watts. The downlink telemetry and ranging are transmitted individually from the satellite 1 watt power amplifier. When the satellite is performing its navigation mission, a 6 dB margin results in a 27° cone covering the earth (0 dB satellite antenna gain). In an emergency mode, the margin reduces to approximately 0 dB as specified (-6 dB satellite gain).

Table 9 presents the uplink SGLS S-band calculation for both the Nadir point and 5° elevation angle. The calculations show that the command data is the restraining link, but even that provides 9.5 dB margin with a 0 dB satellite antenna. Before satellite antenna requirements are considered, the downlink should be presented. Table 10 shows that the telemetry (250 bps) is the constraining downlink factor and provides 6.2 dB margin with a 1 watt RF power amplifier and a 0 dB satellite antenna gain (over 27° coverage). Since only 0 dB margin is required outside 27°, a -6.2 dB antenna gain is necessary over that region.

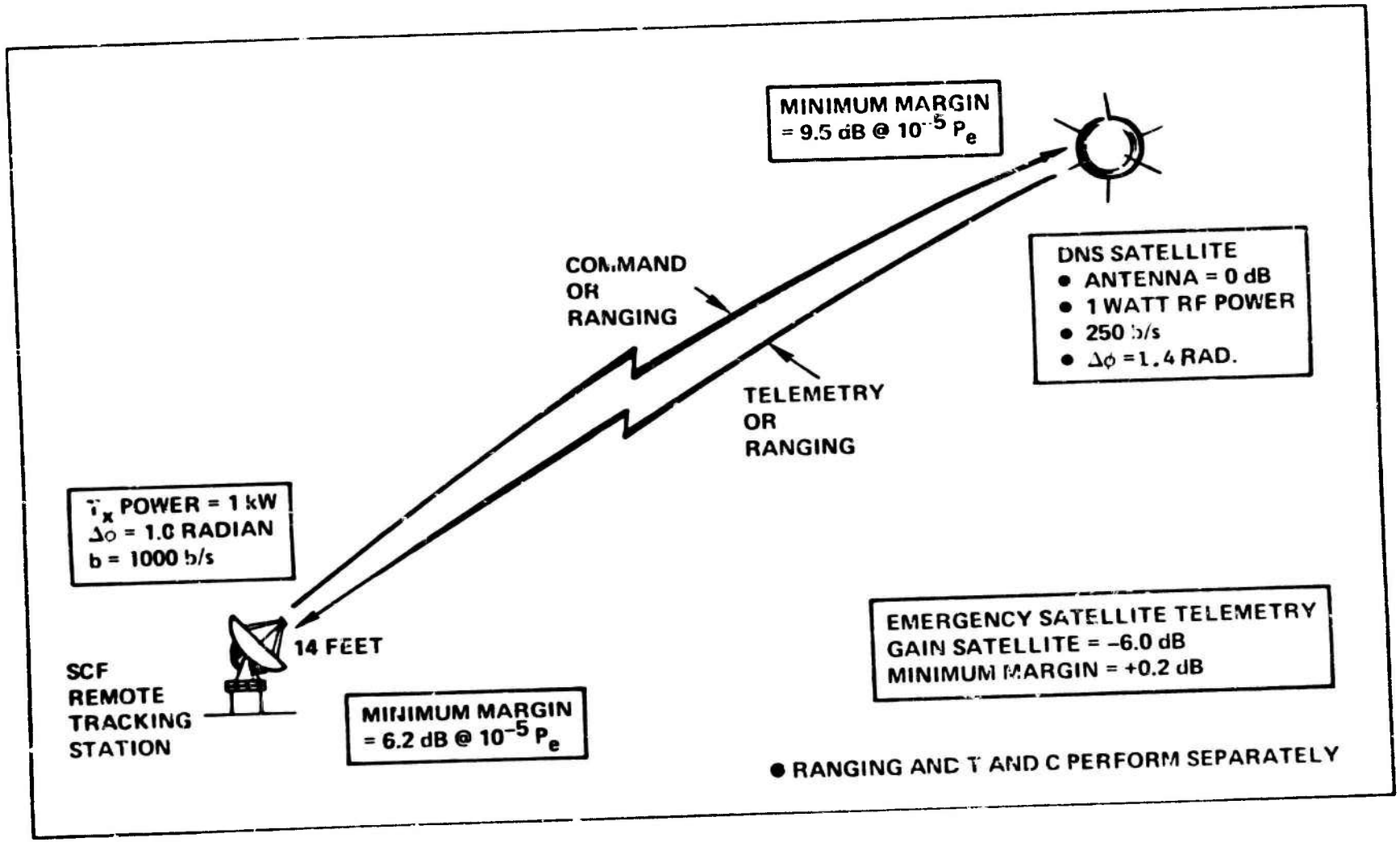


Figure 2 Command and Telemetry Link

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

TABLE 9

UPLINK SGLS S-BAND CALCULATION (1700 MHz)

	Nadir Point	5° Elevation
Transmitter Power (dBm)	60.0	60.0
VSWR Loss	-0.8	-0.8
Line Loss	-0.9	-0.9
Rotary Joint Losses	-0.2	-0.2
Radome Loss	-0.7	-0.7
Antenna Gain (15')	+35.5	+35.5
EIRP (dBm)	92.9	92.9
T_x Point Error (dB)	-0.1	-0.1
Path Loss (dB)	-183.6	-135.6
Polarization (AR = 2)	-0.3	-0.3
Atmospheric Loss (dB)	-0.4	-0.4
R_x Antenna Gain (dB)	0.0	0.0
R_x Line (dB)	-0.5	-0.5
R_x Coupler Loss (dB)	-4.5	-4.5
R_x Power (dBm)	-98.3	-100.3
R_x Noise Density (dBm/Hz)	-162.3	-162.3
R_x C/N_o (dB/Hz)	+65.8	+63.8
Data		
Required E/N_o ($P_e = 10^{-5}$) (dB)	20.1	20.1
Data Rate (1000 b/s)	30.0	30.0
Modulation Loss - 1.0 Radian (dB)	4.2	4.2
Required C/N_o (dB/Hz)	54.3	54.3
Margin (dB)	+11.5	+9.5
Ranging		
Required S/N_o (dB/Hz)	26.6	26.6
Modulation Loss (0.3 Radian) (dB)	11.5	11.5
Required C/N_o (dB/Hz)	38.1	38.1
Margin (dB)	27.7	25.7

TABLE 10

DOWNLINK SGLS S-BAND CALCULATION (2250 MHz)

	Parameter	Carrier I(1) Nadir Point	Carrier I(1) 5° Elevation
1.	Modulation	PCM/PSK/PM	PCM/PSK/PM
2.	Frequency (MHz)	22.50	22.50
3.	Spacecraft Transmit Power (dBm) 2 Watts	33.0	33.0
4.	Transmit Line Losses (dB)	-1.0	-1.0
5.	Satellite Diplexer Loss (dB)	-0.5	-0.5
6.	Satellite T_x Antenna Gain (dB)	-0.0	-0.0
7.	Satellite EIRP (dBm)	28.5	28.5
8.	Polarization Loss (dB) (AR = 2.5 dB)	-0.3	-0.3
9.	Atmospheric Losses at 5° (dB)	-0.4	-0.4
10.	Path Loss - 10,925 n.mi. Altitude (dB)	-185.4	-187.4
11.	Terminal Antenna Gain - 15' (dB)	34.0	34.0
12.	Terminal Line Loss (dB)	-0.4	-0.4
13.	Total Rec'd Power (dBm)	-124.0	-126.0
14.	Terminal System Noise Temp (°K)	400	500
15.	Noise Spectral Density (dBm/Hz)	-172.6	-171.6
16.	R_x Power/Noise Spectral Density (dB/Hz)	48.6	45.6

		Carrier	Sub-Carrier	Carrier	Sub-Carrier	Ranging(2)
17.	Modulation Loss (dB) $\Delta\theta$ = 1.4 Rad., 1.0 Rad. Ranging	-4.9	-2.3	-4.9	-2.3	-2.4
18.	Bandwidth or Data Rate (kHz or Kbps)	0.20	0.25	0.20	0.25	---
19.	R_x S/N or E/N_o (dB)	20.7	22.3	17.7	19.3	43.2
20.	Req'd. Theoretical S/N or E/N_o (dB) $P_e = 10^{-5}$	10.0	9.6	10.0	9.6	26.6
21.	Degradation from Theoretical (dB)	0	3.5	0	3.5	0
22.	Req'd. S/N or E/N_o (dB)	10.0	13.1	10.0	13.1	26.6
23.	Margin (dB)	10.7	9.2	7.7	6.2	16.6

(1) Data on 1.024 MHz subcarrier

(2) Ranging and telemetry not simultaneous

APPENDIX A

GEOMETRY CONSIDERATIONS FOR GPS

This appendix calculates the variation in path loss and satellite antenna gain as a function of user elevation angle and satellite look angle. Also calculated is the area coverage ratio as a function of the angles. Since both angles are functions of the earth central angle, it is convenient to use it as the variable. The geometry is presented in Figure A-1. Slant range, S, may be expressed from the law of cosines as;

$$S = \left((R+h)^2 + R^2 - 2R\,(R+h)\cos\theta\right)^{\frac{1}{2}} \qquad \text{A-1}$$

where the parameters are defined in Figure A-1. Satellite look angle and user elevation angle can be calculated as;

$$\alpha = \sin^{-1}\left(\frac{R\sin\theta}{S}\right) \qquad \text{A-2}$$

$$\beta = 90 - \alpha - \theta \qquad \text{A-3}$$

Using equations A-1 to A-3 α and β can be calculated as a function of θ. These relations are expressed in Figure A-2(a).

The variation in path loss from zenith can be written as;

$$\Delta L = -20\log_{10} S + 20\log_{10} h \qquad \text{A-4}$$

Likewise, the variation in satellite antenna gain from zenith can be written as;

$$\Delta G = 10\log_{10}\left[e^{-0.693\left(\frac{\alpha+\epsilon}{\alpha_{3\ dB}}\right)^2}\right] \qquad \text{A-5}$$

where;

$\alpha_{3\ dB}$ = half power beamwidth = that angle when = 5°

ϵ = pointing error = 1 degree

θ = Earth Central Angle
α = Satellite Look Angle
β = User Elevation Angle

R = Earth Radius - 3440 nm
h = Satellite Altitude = 10,902 nm
S = Slant Range

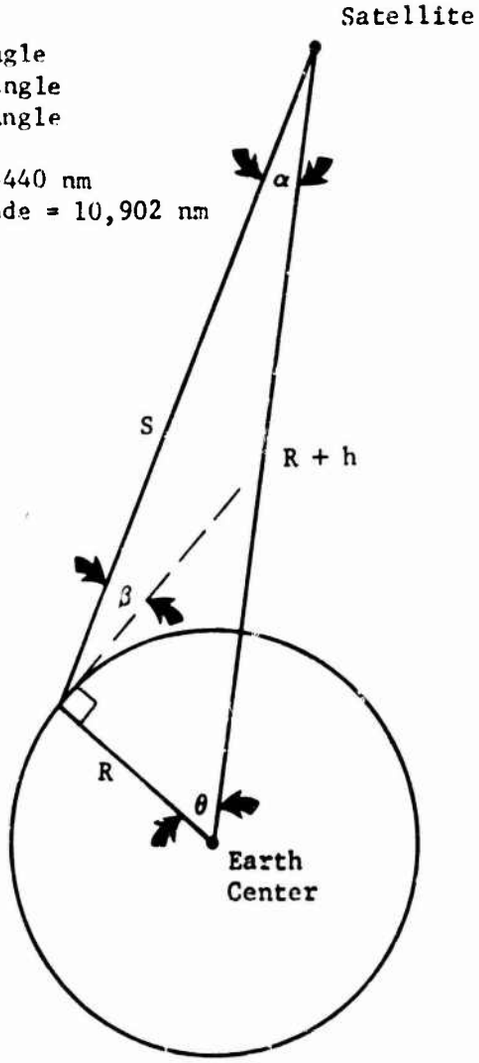


Figure A-1 GPS Geometry

Using equations A-1 through A-5 differencial space loss and satellite antenna gain have been calculated and are presented in Figure A-2(b). The sum of both effects is presented in Figure A-2(c). Also presented in the last figure is the area coverage ratio, which expresses the percentage of the area covered at 5^{o} elevation angle. Note that it is plotted as a function of earth central angle which can be converted to α or β with the use of Figure A-2(a). The expression for the area coverage ratio is simply;

$$\frac{A}{A_{3\ dB}} = \frac{1 - \cos\theta}{1 - \cos\theta_{3\ dB}} \qquad \text{A-6}$$

The printouts for the curves presented in Figure 2 are given for reference in Tables A-1 and A-2.

To determine space loss as a function of elevation angle, use the relation:

$$L = L_o + \Delta L \quad \text{dB} \qquad \text{A-7}$$

where

$$L_o = 20 \log\left|\frac{\lambda}{4\pi R}\right| \qquad \text{A-8}$$

$$\lambda = \frac{c}{f}$$

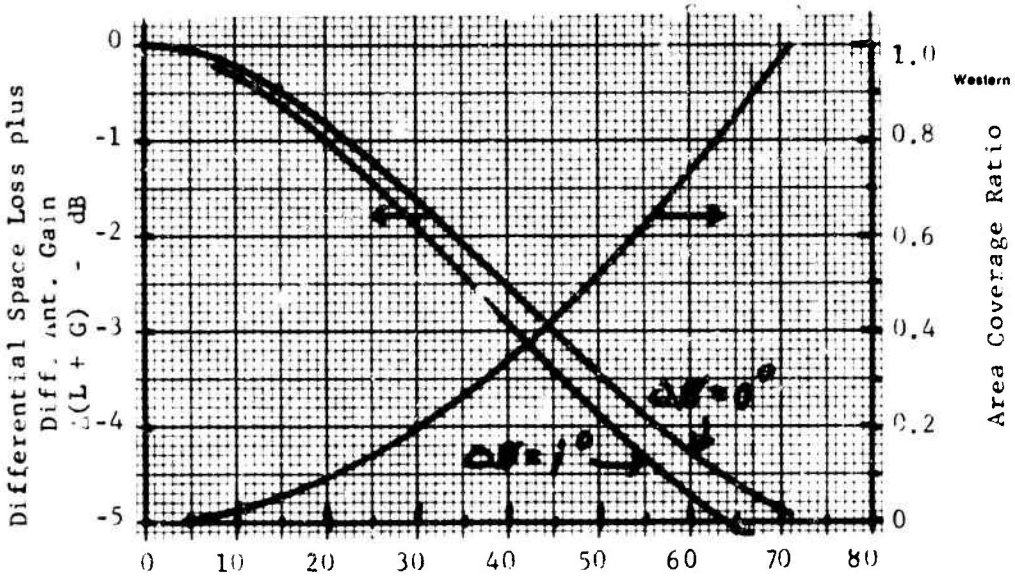
$c = 0.9836 \times 10^{9}$ ft/sec (velocity of light)

$R = 10{,}902$ nm $\times$ 6076 ft/nm

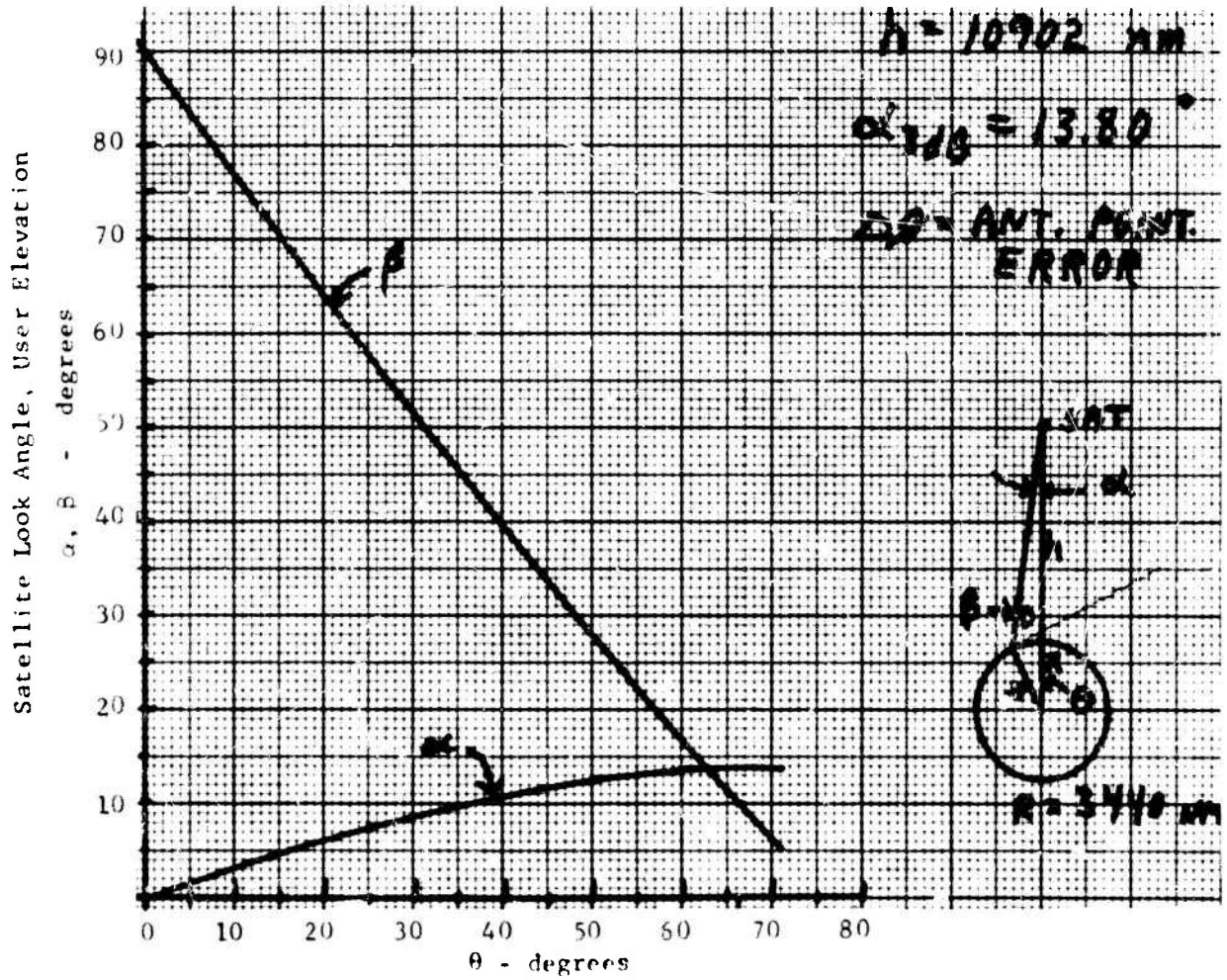
$L_o = -182.50$ dB for f = 1575 MHz

$L_o = -180.35$ dB for f = 1230 MHz

ΔL given in Tables A-1 and A-2



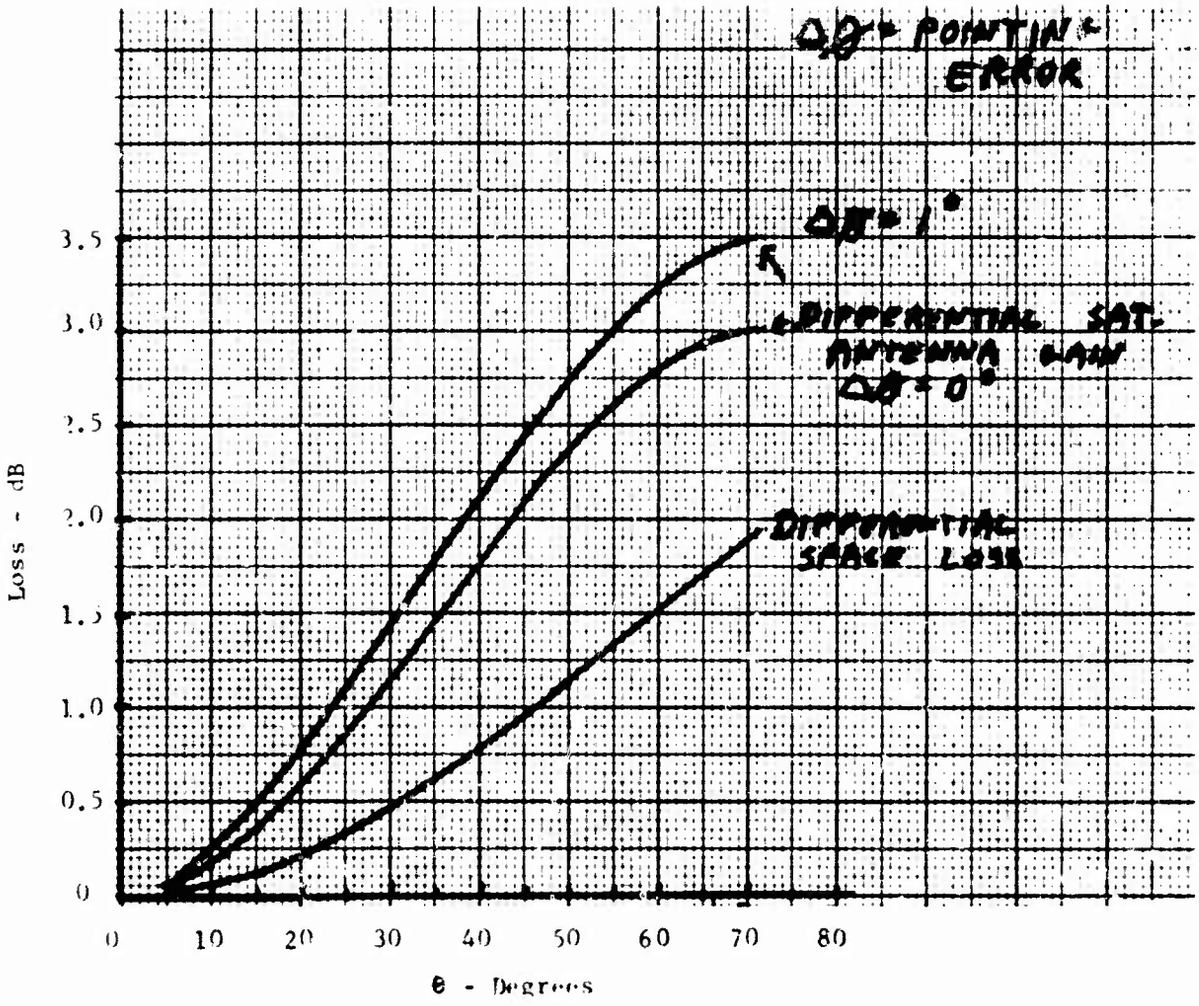
(c) Δ(L + G) and Area versus Central Angle



EARTH CENTRAL ANGLE

(a) Look Angle and Elevation Angle versus Central Angle

Figure A-2 GPS Geometric Parametrics



EARTH CENTRAL ANGLE

(b) ΔL and ΔG versus Central Angle

Figure A-2 GPS Geometric Parametrics (Continued)

TABLE A-1 PRINTOUT $\epsilon = 0^{o}$

THETA DEG	S	ALPHA DEG	BETA DEG	L DB	G DB	A/A2	L+G DB
0.	10902.	0.	90.00	0.	0.	0.0000	0.
5.000	10919.	1.57	83.43	-0.01	-0.04	0.0056	-0.05
10.000	10971.	3.12	76.88	-0.05	-0.15	0.0224	-0.21
15.000	11055.	4.62	70.38	-0.12	-0.34	0.0503	-0.46
20.000	11172.	6.05	63.95	-0.21	-0.58	0.0890	-0.79
25.000	11318.	7.38	57.62	-0.33	-0.86	0.1383	-1.18
30.000	11492.	8.61	51.39	-0.46	-1.17	0.1978	-1.64
35.000	11692.	9.72	45.28	-0.61	-1.49	0.2670	-2.09
40.000	11914.	10.70	39.30	-0.77	-1.80	0.3454	-2.57
45.000	12155.	11.54	33.46	-0.95	-2.10	0.4324	-3.04
50.000	12414.	12.26	27.74	-1.13	-2.37	0.5274	-3.49
55.000	12686.	12.83	22.17	-1.32	-2.59	0.6296	-3.91
60.000	12969.	13.28	16.72	-1.51	-2.78	0.7382	-4.29
65.000	13260.	13.60	11.40	-1.70	-2.91	0.8524	-4.61
66.300	13357.	13.66	10.04	-1.75	-2.94	0.8829	-4.69
70.000	13556.	13.80	6.20	-1.89	-3.00	0.9714	-4.89
71.176	13627.	13.82	5.00	-1.94	-3.01	1.0000	-4.95

TABLE A-2 PRINTOUT $\epsilon = 1^{\circ}$

THETA DEG	S	ALPHA DEG	BETA DEG	L DB	G DB	A/A2	L+G DB
0.	10902.	0.	90.00	0.	-0.02	0.0000	-0.02
5.000	10919.	1.57	83.43	-0.01	-0.10	0.0056	-0.12
10.000	10971.	3.12	76.88	-0.05	-0.27	0.0224	-0.32
15.000	11055.	4.62	70.38	-0.12	-0.50	0.0503	-0.62
20.000	11172.	6.05	63.95	-0.21	-0.78	0.0890	-0.99
25.000	11318.	7.38	57.62	-0.33	-1.11	0.1383	-1.43
30.000	11492.	8.61	51.39	-0.46	-1.45	0.1978	-1.91
35.000	11692.	9.72	45.28	-0.61	-1.81	0.2670	-2.42
40.000	11914.	10.70	39.30	-0.77	-2.15	0.3454	-2.93
45.000	12155.	11.54	33.46	-0.95	-2.48	0.4324	-3.42
50.000	12414.	12.26	27.74	-1.13	-2.77	0.5274	-3.90
55.000	12686.	12.83	22.17	-1.32	-3.01	0.6296	-4.33
60.000	12969.	13.28	16.72	-1.51	-3.21	0.7382	-4.72
65.000	13260.	13.60	11.40	-1.70	-3.36	0.8524	-5.06
66.300	13337.	13.66	10.84	-1.75	-3.39	0.8797	-5.14
70.000	13556.	13.80	6.20	-1.89	-3.45	0.9714	-5.34
71.176	13627.	13.82	5.00	-1.94	-3.46	1.0000	-5.40

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

REPORT NO. 3

MODELING STABLE CLOCKS IN THE GPS

TABLE OF CONTENTS

SECTION		PAGE
1.0	SCOPE	3-1
2.0	CAUSES OF TIME DISPERSION	3-4
	2.1 Clock Noise	3-5
3.0	TIME DISPERSION PREDICTION ALGORITHMS	3-7
	3.1 The Linear Predictor	3-9
	3.2 The Exponential Predictor	3-10
	3.3 Frequency and Time Jump Detection	3-11
	3.4 Summary	3-12
4.0	THE EFFECTS OF THE ENVIRONMENT AND PHYSICAL UPDATE ON CLOCK BEHAVIOR	3-15
	4.1 Environment	3-16
	4.2 Treatment of Environmental Effects	3-20
	4.3 Effects of Physical Update of the Clock Frequency	3-21
5.0	LABORATORY MEASUREMENTS AND COMPUTED SIMULATION OF USER ERRORS	3-23
	5.1 Laboratory Measurements and Derived Clock Models	3-23
	5.2 Calculation of Short Term Stability	3-28
	5.3 Computer Simulation of Effect of Clock Errors on User Position	3-31
	5.3.1 Simulation by the Program TRACE	3-31
	5.3.2 Model Adequacy	3-32
	5.3.3 TRACE Simulations	3-35
	5.3.4 Results of the Runs	3-38
	5.3.5 Users Parameters	3-42
	5.4 Concluding Remarks	3-42
6.0	EXPECTED PERFORMANCE AND CONCLUSIONS	3-47

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Clock Timing Problem	3-2
2	Typical Stability Plot	3-8
3a	Finite Difference of the Time	
3b	Differences	3-13
4	Clock Prediction Algorithms	3-14
5	Temperature Stability Tests - Two Efratom Units	3-18
6	Histograms Showing Distribution of Frequency Stability Versus Temperature for Collins AFS-81	3-18
7	Irradiation Test Results, Unit 120 Efratom	3-19
8	Laboratory Set-Up	3-24
9	Flow Diagram for Accumulative Phase	3-25
10	Error in Prediction of Time Dispersion of Two Rubidium Clocks	3-27
11	Short-Term Stability Versus Averaging Time	3-30
12	Allan Variance of Real and Simulated Clocks	3-34
13	Error on Range Bias	3-39
14	Error in Range Drift	3-40
15	Clock Timing Error - Vehicle #1	3-41
16	Longitude Error	3-43
17	Latitude Error	3-44
18	Altitude Error	3-45
19	Bias Error	3-46

LIST OF TABLES

TABLE		PAGE
1	Stability Characteristics of Atomic Standards	3-7
2	Summary of Environmental Effects	3-17
3	Values of the Noise Coefficient	3-33
4	Summary of Timing Error	3-48

MODELING STABLE CLOCKS IN THE GPS

1.0 SCOPE

The Global Positioning System operates on the ability of a user to determine time of arrival (or time differences) of ranging signals to a high degree of accuracy. These signals originating from medium altitude satellites and controlled by stable clocks are monitored at numerous locations. The monitoring locations use these signals for many purposes including determination of ephemeris and update of the time between satellites in the system. Figure 1 shows a simple block diagram of the timing system. From this figure it can be seen that for some period of time the satellite clocks are in operation independent of the ground and each other.

Ideal timing systems for this application (and many others) would then be a set of clocks on the ground and in the satellites which would never change time relative to each other. However, no clocks exist which do not exhibit time change from each other a a function of time. We will define this time difference as time dispersion. Time dispersion is simply the time change between some clock i and the perfect clock and is expressed in equation 1.

$$T_i(t) = t - t_i \qquad (1)$$

It is $T_i(t)$ which we will model and attempt to predict for some time Δt between updates. If we assume a master timing reference or primary standard located at the master station, then the problem summarizes to predicting the behavior of all clocks in the system (satellites, monitor stations, SCF, etc.) relative to this primary standard.

The clock model to the user for prediction of time dispersion is $T_{ij}(t) = T_{ij}(t_o) + \left[R_{ij}(t_o) + \hat{R}_{ij}(t)\right] \times_v (t-t_o) + \frac{1}{2} D_{ij} \times (t-t_o)^2$.

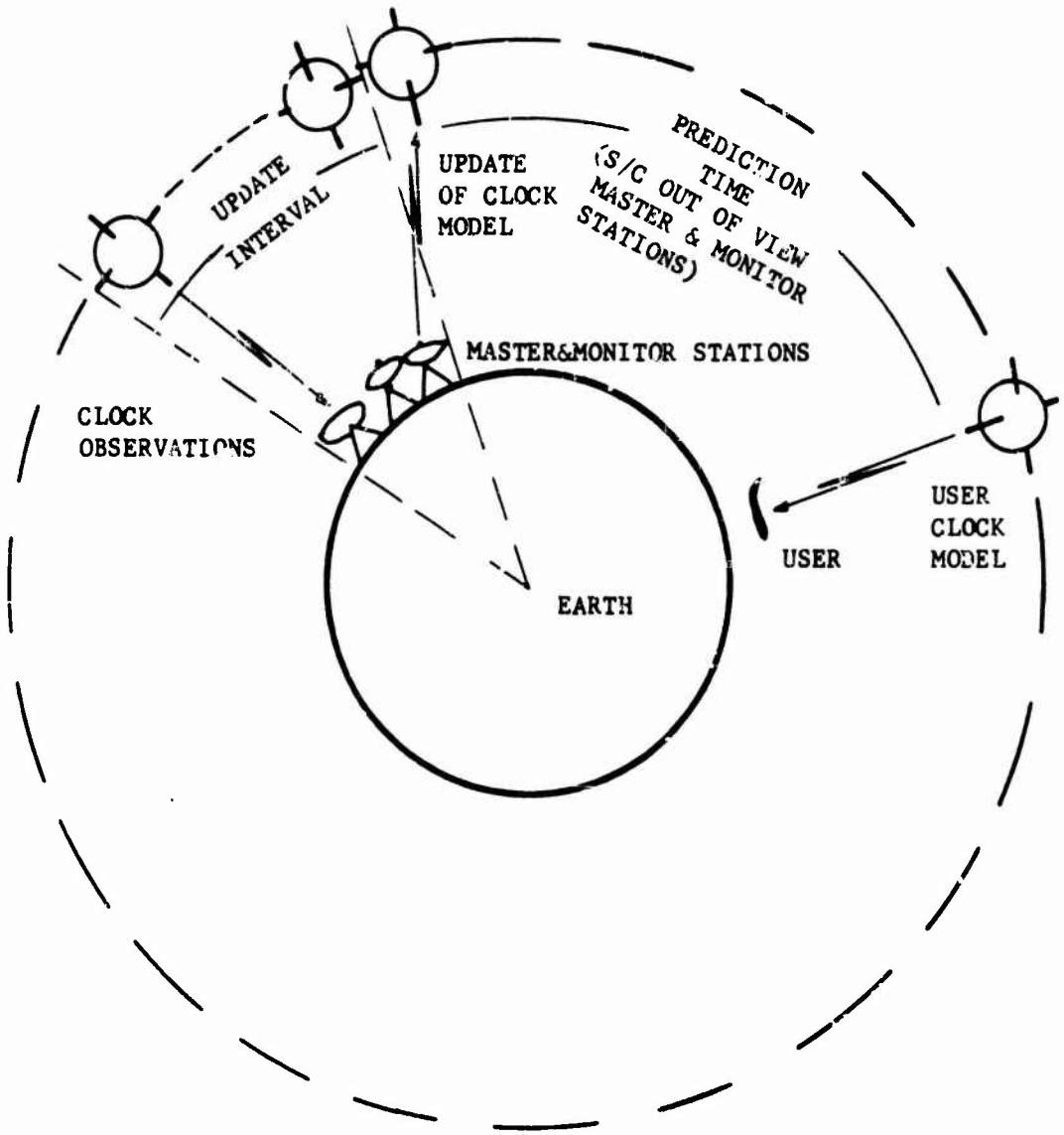


Figure 1 Clock Timing Problem

In this expression $T_{ij}(t_o)$, $R_{ij}(t_o)$ and $D_{ij}(t_o)$ are the time offset, rate offset and rate drift offset respectively of clock i relative to clock j. $\hat{R}_{ij}(t)$ is the optimum estimate of the residual of the measured frequency offset and the estimated frequency offset at time t. This estimate is a result of recursive algorithm (exponential) weighting the old residuals to present residuals The weighting of the recursive filter is chosen to be optimized for the flicker noise FM process of the clocks and the desired prediction time. The prediction process (Fig. 4) summarizes to: a least squares fit to the frequency data to obtain an initial estimate of R_{ij} and D_{ij} and following this a recursive filtering of the residuals to optimize the estimate of R_{ij} in the presences of white and flicker noise FM.

It is possible that the functions described in this paper and shown in Figure 4 could be assumed by a Carlson Triangular filter algorithm. However, caution should be exercized in applying this algorithm in that long term measurements of clock bias will be polluted to predominantly by flicker noise FM not white noise. Thus, any application of the Carlson triangular filter which depends on white noise statistics at the input could be distorted due to the flicker noise.

The purpose of this memo will be to describe the reasons for time dispersion in ultra stable clocks and techniques for modelling this time dispersion and, thus, predicting $T_i(t)$ for some future time period. We will also discuss some laboratory measurements performed, using two rubidium standards and the results of modelling this data and predicting their time dispersion. Lastly, we will discuss the sensitivity of the prediction algorithm to the update interval and the effects any updating mechanism applied to the satellite clock would have on the prediction algorithm.

2.0 CAUSES OF TIME DISPERSION

The basic causes of time dispersion between stable clocks are listed below:

a. Synchronization errors (time offset)
b. Rate error (time drift)
c. Frequency drift
d. Clock noise
e. Steps in time
f. Steps in frequency

Synchronization errors, rate errors and frequency drift are all deterministic properties of the time dispersion in clocks. As such, they are measureable quantities which can be used in a clock model to predict the time dispersion. Clock noise is a non-deterministic property of clocks. Since noise is a random process, its effect on the time dispersion of the clock is not predictable. However, it will be shown that by knowing the properties of the clock noise, it will be possible to make an estimate as to what the time error will be due to the noise. Lastly, it has been observed in some clocks that steps in time or frequency do occur. These may have causes internal, or external to the clock. However the effect is to change the deterministic terms of the time dispersion model. Techniques exist such that it is possible to detect any jumps in the time and frequency offset of the clock. This information is then used to correct the time dispersion prediction algorithm (Ref. 1).

From the above information, we can write a simple expression for the time dispersion between clocks i and j as follows:

$$T_{ij}(t) = T_{ij}(t_o) + R_{ij}(t_o) \times (t-t_o) + \frac{1}{2} D_{ij} \times (t-t_o)^2 + n_i(t) + n_j(t) \quad (2)$$

In this expression, $T_{ij}(t_o)$ is the time offset and $R_{ij}(t_o)$ is the frequency offset between clocks i and j at time t_o. D_{ij} is the frequency drift offset between clocks i and j, and u(t) is the clock noise (Reference 1). This expression constitutes the basic clock prediction model. The problem is to measure or determine accurately T_{ij}, R_{ij}, D_{ij} in the presence of $n_i(t)$ and

$n_j(t)$ to make some estimate on the contributions of $n_i(t)$ and $n_j(t)$ to the time error between clocks i and j. It is, therefore, necessary to make some characterization of the clock noise n(t).

Before proceeding to discuss clock noise, some words should be said concerning the use of equation 2 as the clock model. At this time, this expression represents the known physical behavior for time error between clocks, deterministic and random. It is conceivable that long polynomials or Fourier series expansion could be used to describe clock behavior, the coefficients of such expansions being based on past measurements. However, there is no reason to believe such a model would accurately predict future time dispersion. Indeed, any higher order polynomial fit (greater than degree 2) to accumulated time error data would indeed be attempting to fit the clock noise, rather than the known deterministic properties of the clock. Since we are dealing with noise, a deterministic time series representation of past data will bear no relationship on future behavior. Thus, our approach has been to model the deterministic behavior of the clock and to obtain an estimation technique of time error for the various types of clock noise.

2.1 Clock Noise

A clock is defined as a device which counts the cycles of some periodic phenomenon. Thus the clock is a frequency device and in determining the stability (or instability) of this device, we are mainly concerned with noise which affects the frequency of the clock. Typical types of noise observed in clocks are (a) white noise FM, (b) flicker noise FM and (c) random walk FM.

At this point, it is appropriate to define one of the most accepted measures of clock noise, the Allan variance. The Allan variance $\sigma^2_{2,\tau}\left(\frac{\Delta f}{f}\right)$ is a measure of the instability of the clock frequency. It differs from the conventional definition of variance only in the number of samples used to calculate the variance. Calculation of an Allan variance involves only two measurements of frequency. The variance is a function of the averaging time τ involved in the frequency measurement. Since only two samples are used in the calculation, the confidence level on the measurement is low. To increase this confidence level a number of separate calculations of Allan variance are averaged together. The expression for this averaged Allan variance

$\left\langle \sigma^2_{2,\tau}\left(\frac{\Delta f}{f}\right)\right\rangle$ is given in equation 3.

$$\left\langle \sigma^2_{2,\tau}\left(\frac{\Delta f}{f}\right)\right\rangle \frac{1}{f^2}\sum_{n=1}^{N}\frac{\left(f_n - f_{n-1}\right)^2}{2n} \qquad (3)$$

where τ is the measurement time of f_n

In our expression for time dispersion, we have defined the noise term as phase noise n(t). As stated previously the most often observed noise in clocks is noise on the frequency of the clock. Therefore, we will concern ourselves only with FM noise and define a new variable $y_i(t) = \dot{n}_i(t)$. (Although there exists a contribution to clock instability from white noise PM and flicker noise PM for averaging time < 1 second, this contribution is not predominant.) The time dispersion contributed by the clock noise will depend on the spectrum of this noise. Work performed at National Bureau of Standards principally by Allan and Barnes has shown that clock noise can be adequately modeled by power law spectral densities. Therefore, we represent the spectral density of $y_i(t)$, $S_{y_i}(f)$ by the following expression.

$$S_{y_i}(f) = h_o f^o + h_{-1} f^{-1} + h_{-2} f^{-2} \qquad f \leq 1 \text{ Hz} \qquad (4)$$

In this expression, the h_o term represents white noise FM, h_{-1} is the flicker noise FM, and h_{-2} is the random walk FM. The h_o term is directly related to physical quantities associated with the particular clock, such as resonance linewidth, signal-to-noise ratio of the detector, etc. The cause of the flicker noise contributions are not known at this time, and evaluation of h_{-1} can only be made through direct measurement (Ref. 2). Likewise, the random walk h_{-2} term may be related to changes in the physical environment of the clock (i.e., temperature, pressure, magnetic field, etc.). A direct relationship to physical quantities is not possible and, as such, determination of h_{-2} must be accomplished through measurements of the clock in its operational environment. Equation 5 relates the time domain measure of instability, the Allan variance, to expression 4 for the noise spectral density.

$$\sigma^2_{2,\tau}\left(\frac{\Delta f}{f}\right)\ \sigma^2(\tau) = \frac{h_o}{2\tau} + 2h_{-1}\ln 2 + \frac{(2\pi)^2}{6} h_{-2}\tau \quad \text{(Reference 2)} \qquad (5)$$

This expression is plotted in Figure 2, and coupled with Table 1, represents the typical behavior of rubidium and cesium standards. Note, since the effects of environment on each clock will vary with the clock, and with the particular environment of operation, it is not possible to assign a value a priori for h_{-2} and τ_2 as shown in Figure 2.

The value of measurements of h_o, h_{-1}, h_{-2}, τ_1, τ_2 are used in determining the proper estimation algorithm for the desired prediction period. In addition, measurement of $\sigma^2(\tau)$ is necessary to determine the time dispersion uncertainty given a particular prediction algorithm.

TABLE 1 STABILITY CHARACTERISTICS OF ATOMIC STANDARDS

	$\sqrt{\frac{h_o}{2}}$	$\sqrt{2 \ln (2) h_{-1}}$	$\tau_1 = \frac{h_o}{4h_1 \ln 2}$
Rubidium	$\approx 10^{-11} (\text{sec}^{1/2})$	$3 \times 10^{-13} - 10^{-12}$	15 Min. to 1 Hr.
Cesium	$\approx 3 \times 10^{-11}$	$3 \times 10^{-14} - 3 \times 10^{-13}$	2-5 days to 1 Mo.
HP Super Tube	$\approx 5 \times 10^{-12}$	3×10^{-14}	< 1 day

3.0 TIME DISPERSION PREDICTION ALGORITHMS

The following section describes a technique for prediction of the time dispersion of each clock in the GPS system relative to some primary standard at the master station. This algorithm would provide the user with a means to correct the error in his range difference measurement due to drift in the satellite clocks relative to some system time (assumed to be the primary standard at the master station). In this algorithm, we have assumed that the user will have a clock which has good short term stability over the time period of his measurement. We are not considering the effects of the long term stability of the user's clock under the assumption that the user will solve for system time from measurements on a fourth satellite.

Equation (2) models the time dispersion of two clocks in terms of their deterministic and random components. $T_i(t_o)$, $R_i(t_o)$, D_{ij} are masked by the clock noise. So, in essence, we have a situation

$$\tau_1 = \frac{h_0}{4\ \ln(2)\ h_{-1}}$$

$$\tau_2 = \frac{12\ h_{-1}\ \ln(2)}{(2\pi)^2\ h_{-2}}$$

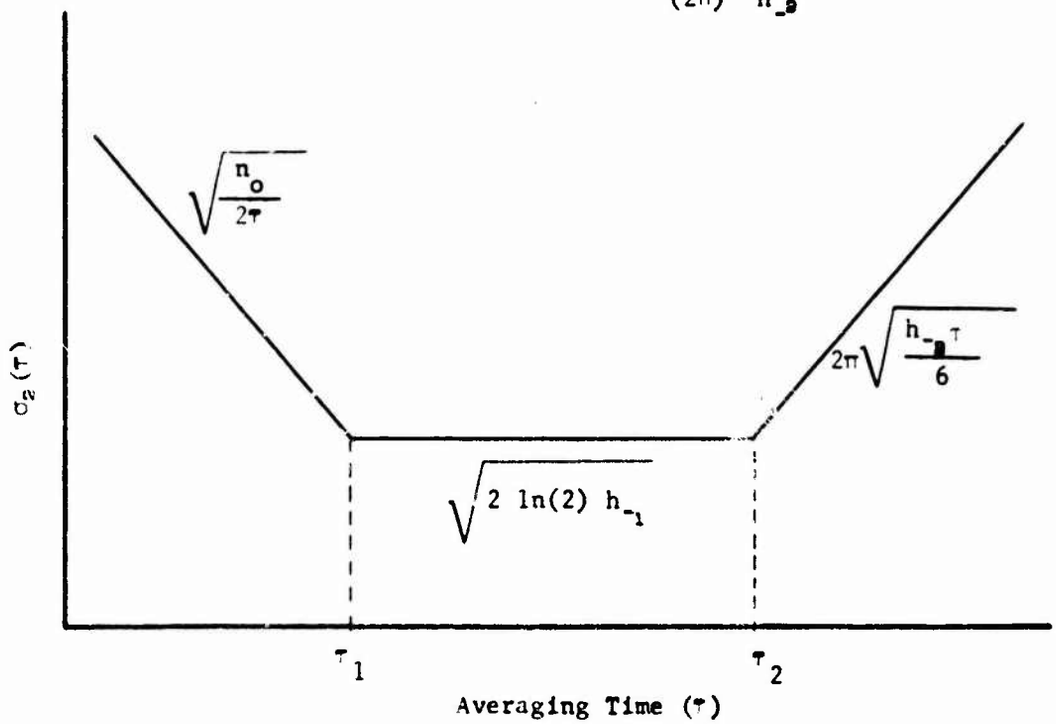


Figure 2 Typical Stability Plot

of signal detection to determine T_{ij}, R_{ij}, D_{ij}. The time dispersion prediction algorithm then is simply:

$$T_{ij}(t) = T_{ij}(t_o) + R_{ij}(t - t_o)\, xt + \frac{1}{2} D_{ij} t^2 \quad \text{(Ref. 1)} \tag{6}$$

The coefficients of this expression can be either stored with the users or sent to him, with update. Some form of update will be required as the estimate of these coefficients improves or as they change.

The problem then is given: we can make an initial estimate of the coefficients for equation (6); by what technique do we better our estimate; and what will be the optimum estimation (in the mean square error sense) for future estimates of time dispersion caused by the clock noise This problem has been studied extensively by Barnes and Allan and, although it can be approached by conventional optimal estimation theory techniques, they have developed a near optimum technique for estimation of R_{ij} in the presence of flicker and white noise using an exponential filter. They have also studied the problem using a linear predictor (which is optimum for white noise), and an optimum filter (over 5 decades) flicker noise FM. In this memo, we will discuss only the linear and exponential predictors, since they are most easily implemented in the GPS, and this analysis follows directly from the work performed at NBS.

3.1 The Linear Predictor

If we consider the case of only white noise FM on the clock frequency (i.e., $S_y(f) = h_o f^o$), then it can be shown that the optimum estimate of the clock frequency is as given below (Ref. 2).

$$\hat{R}_{ij}\left(t - \frac{\tau_c}{2}\right) = \frac{T_i(t) - T_i(t - \tau_c)}{\tau_c} \tag{7}$$

In this expression $\hat{R}_{ij}$ indicates an estimate on R_{ij} and τ_c is the calibration interval. That is, we are estimating the clock freqeuency by merely differencing the clock "time" over the entire calibration interval. The longer the calibration interval the better the estimate of R_{ij} for the case white noise FM only. Given that we now have an estimate of R_{ij} which can be used to

predict time dispersion (Equation 6), we would like to know what the mean squared timing error of this prediction would be. Equation 8 defines this error.

$$< \epsilon_{ij}^2(\tau) > = < \left[\hat{T}_{ij}(t+\tau) - T_{ij}(t+\tau)\right]^2 > \qquad \text{(Ref. 2)} \qquad (8)$$

$\hat{T}_{ij}(t)$ is the time dispersion obtained from using the rate estimate from Equation 7 and T_{ij} is the actual time dispersion. From Allan, (REF 1) this mean square error can be related to the clock noise by expression

$$\left\langle \epsilon_{ij}^2(\tau) \right\rangle = \frac{h_o}{2\tau_c} (\tau_c - \tau) \cdot \tau$$

$$+ h_{-1} \tau^2 \left[\frac{\tau}{\tau_c} \left(\frac{\tau_c}{\tau} + 1 \right)^2 \ln \left(\frac{\tau_c}{\tau} + 1 \right) - \left(\frac{\tau_c}{\tau} + 1 \right) \ln \frac{\tau_c}{\tau} \right] \qquad (9a)$$

$$\left\langle \epsilon_{ij}(\tau) \right\rangle \simeq \frac{h_o}{2} \tau + h_{-1} \left(1 + \ln \left(\frac{\tau_c}{\tau} \right) \right) \tau^2 \qquad \tau \ll \tau_c \qquad (9b)$$

$$\left\langle \epsilon_{ij}^2(\tau) \right\rangle = h_o \tau + 4 h_{-1} \ln(2) \tau^2 \qquad \tau = \tau_c \qquad (9c)$$

$$\left\langle \epsilon_{ij}^2(\tau) \right\rangle \simeq \frac{h_o}{2\tau_c} \tau^2 + h_{-1} \left(1 - \ln \left(\frac{\tau_c}{\tau} \right) \right) \tau^2 \qquad \tau \gg \tau_c \qquad (9d)$$

Expressions 9b through 9d follow intuitively that as the prediction interval τ increases relative to the calibration interval τ_c, the timing uncertainty or error likewise increases. In these expressions h_o, h_{-1} are the measures of white noise FM and flicker noise FM as given in Section 2.0.

3.2 The Exponential Predictor

A second form of predictor suited for estimation of the frequency in the presence of flicker noise is the exponential weighting filter. In this case, given a sequence of measurements of the frequency offset of the clock, the observations are weighted by the recursive relation given in Equation 10.

$$\hat{RM}_{ij}(n\Delta t) = \frac{1}{m+1}\left[RM_{ij}(n\Delta t) + m\,\hat{RM}_{ij}((n-1)\,\Delta t)\right] \qquad (10)$$

The values of RM_{ij} may be actual measurements of the fractional frequency offset of the i and j clocks, or may be derived from time measurements made on these clocks. The value of $m\Delta t$ is the time constant of this filter. Its numerical value is chosen such that the mean square time error is at a minimum. This value will depend on the particular noise characteristic of the clocks and on the time interval of prediction desired. The mean square time error considering white noise FM and flicker noise FM is given in the following expression:

$$\left\langle \epsilon^2(\tau) \right\rangle = \frac{h_o(m+1)}{2m+1}\,\tau + \frac{h_{-1}\tau^2}{m^2(2m+1)} \sum_{k=1}^{\infty} \left(\frac{m}{m+1}\right)^k k^2 \ln(k). \text{(Ref. 2)} \qquad (11)$$

The infinite series given on the right in expression (11) has been tabulated by Allan et al, for various values of m. The mean square time uncertainty given in Equation 11 will be evaluated for typical values of h_o and h_{-1} in Section 6.0. For values of m chosen less than 1, expression 11 will result in less timing error than expression 9. Indeed, in 9d, for prediction times greater than τ_1 as given in Table 1, the predominant noise will be flicker noise and the exponential weighting filter is a better approach to handling the data.

3.3 Frequency and Time Jump Detection

It has been observed in rubidium and cessium standards that jumps in time and frequency occur, which are beyond the expected changes which can be attributed to the clock noise. These change in time and frequency terms must be accounted for in the deterministic portion of the clock model. Simple statistical techniques have been developed by J. Barnes at NBS to detect any frequency or time jump has occurred with a particular standard. Although they have encountered these jumps infrequently in the clocks of the NBS Atomic Time Scale (approximately 3 per year), it should be noted that these clocks are maintained in a highly stable environment (temperature controlled within 10 millidegrees C). It is quite conceivable that any clocks operating in a satellite environment would exhibit more frequent jumps in time and rate. For this reason, it is necessary to include detection of time and frequency steps in the GPS clock prediction algorithm.

The procedure to determine if a jump has occurred is described herein. This procedure, as with determining the coefficients of the clock model, requires the examination of the clock statistics, which have been gathered for a considerable period of time. In this case, a comparison is made between the actual measurements of time offset or rate offset against the predicted values. Since we are looking for steps in time T_{ij}, or frequency R_{ij}, the procedure described by Barnes (Ref. 2) examines $\dot{T}_{ij}$ and $\dot{R}_{ij}$ for an impulse change in these derivatives. This error between actual and predicted is then compared with the rms error to determine if a jump has occurred. This is illustrated in Figure 3 and Equation 12.

$$\left| T_{ij}(t)_{pred.} - T_{ij}(t)_{actual} \right| \begin{cases} \leq 3\sigma(\epsilon_{rate}) & \text{No jump} \\ \geq 3\sigma(\epsilon_{rate}) & \text{Jump} \end{cases} \qquad (12a)$$

$$\left| R_{ij}(t)_{pred.} = (t)_{actual} \right| \begin{cases} \leq 3x\sigma(\epsilon_{rate}) & \text{No jump} \\ \geq 3x\sigma(\epsilon_{rate}) & \text{Jump} \end{cases} \qquad (12b)$$

where $\sigma(\epsilon_{rate\ or\ time})$ is the rms error of predicted to accual.

It may be desireable to average (or smooth) a number of days of data prior to making the check given in expression 12 (to filter the noise on the data). The actual averaging time will depend on the characteristics of the particular clock and the effect of a space environment on this clock.

3.4 Summary

In the previous section, we have described a series of techniques from which a time dispersion prediction algorithm can be developed. This algorithm is functionally shown in Figure 4. This prediction technique consists of three separate phases.

The initial phase is a calibration period in which as much data is gathered on the time dispersion characteristics of the clock as equipment and availability of the satellite will allow (a minimum of 30 days in order to obtain any confidence in the model). In addition, prelaunch data on the clock in a simulated space environment should be part of the calibration phase. This calibration data is used for two purposes: First, to determine the coefficients of the

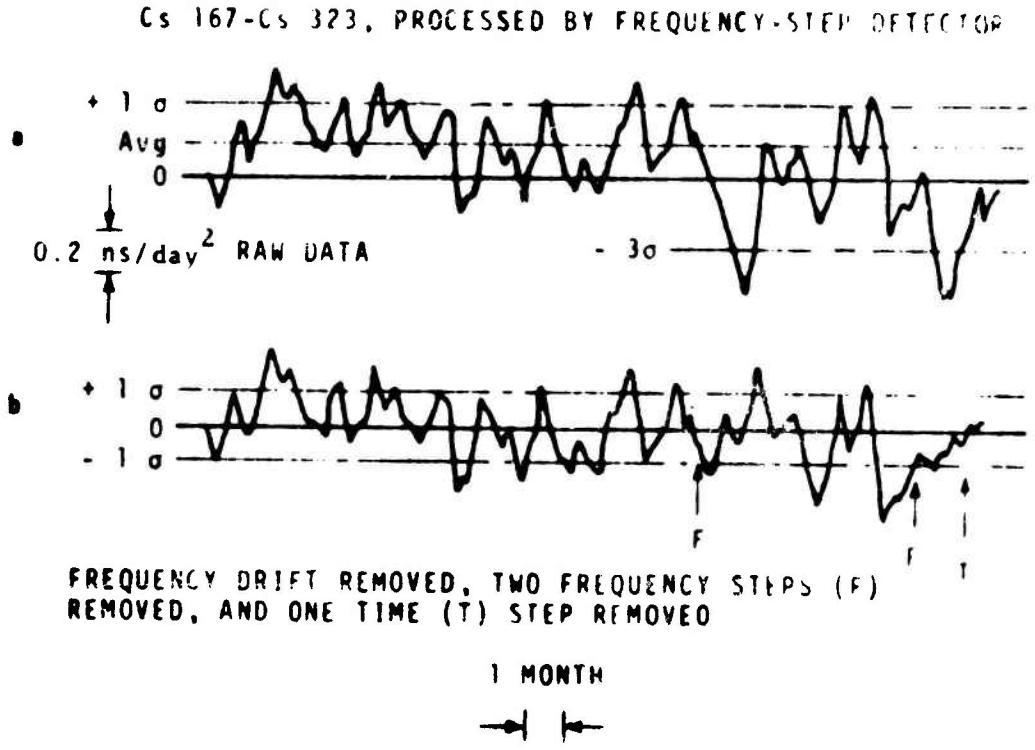


Figure 3a Second finite difference of the time differences (second derivative of the time differences) between Cs 167 - Cs 323 processed through an F-step detector to ascertain if frequency steps have occurred.

Figure 3b Same as 3a but with frequency drift 2 F-steps and 1 T-step removed.

Reference 1

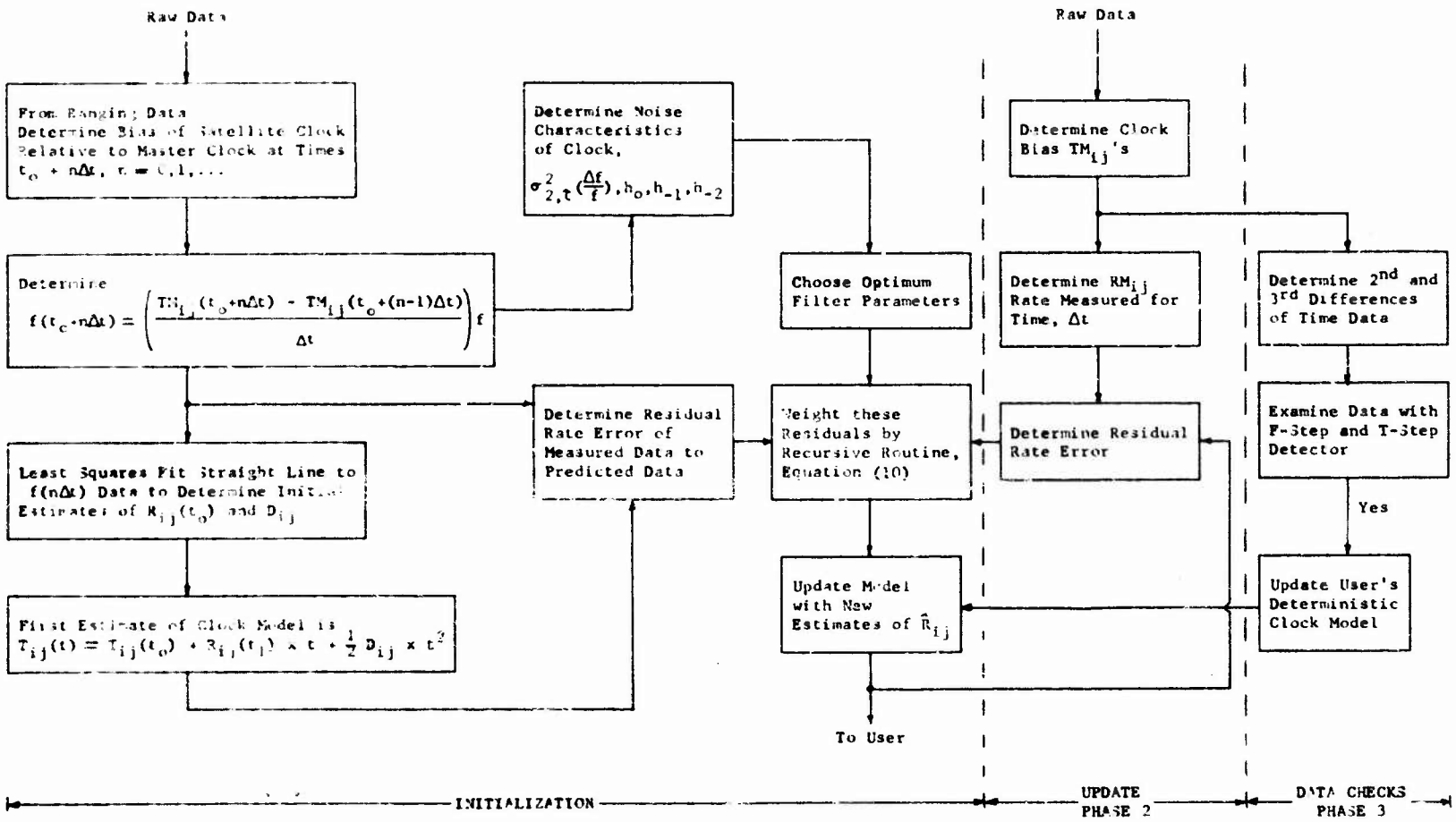


Figure 4 Clock Prediction Algorithm

PHILCO Ford
Philco-Ford Corporation
Western Development Laboratories Division

deterministic clock model, and second to determine the particular noise characteristics of the clock or clocks aboard each spacecraft. The actual technique for determining the clock model coefficients and the clock noise characteristics will be given in Section 5.0, Laboratory Measurements and Modeling of Clock Performance. Note the noise measurements made on the spacecraft clocks will be useful to determine if the clock is performing to specification and to change to the redundant unit, if necessary.

The second and third phases involve update and monitor of the clock behavior in the operational system. In the second phase, all the calibration data and all new measurements made on the clock are used to update or continually estimate the rate term of the clock model. This is done by filtering the residuals (predicted-actual) with the exponential recursive expression (1) to continuously improve the estimate of the clock rate in the prediction algorithm. This technique will also be more fully described in Section 5.0.

The last aspect to the clock prediction algorithm is monitoring for jumps or steps in frequency and time not accounted for by the clock noise. This also involves calculating the $\dot{x}$ rate and $\ddot{x}$ rate error and comparing them against the calculated rms errors to determine if a step has occurred. This comparison is done according to expression (12). When a step is detected, the deterministic clock model is adjusted accordingly and its new form is used for prediction and determining residuals for the exponential filter.

If all three techniques are employed in determining and maintaining the clock model, the user will be provided with a time dispersion prediction which represents a near optimum estimation of the clock behavior for the stable environment. Techniques for handling the spacecraft environment are given in the next section.

4.0 THE EFFECTS OF THE ENVIRONMENT AND PHYSICAL UPDATE ON CLOCK BEHAVIOR

In the previous sections, we have discussed the deterministic characteristics of clocks. Two additional influences will affect the actual time dispersion of the clock and the ability to predict this dispersion. These influences are the environment in which the clock is maintained and the physical updating (change of frequency or phase) of the clock.

4.1 Environment

The actual effects of the environment on time dispersion are highly dependent on the individual clock and the actual environment. Many different influences affect the frequency stability of a stable clock. Some of the more important environmental conditions, which have a direct influence on clock behavior include: (a) temperature, (b) pressure, (c) magnetic field, (d) gravitational fields, (e) radiation, (f) acceleration, and (g) changes in the supply voltage.

Of these different environmental influences, temperature changes have an extremely large effect on the stability and fractional frequency offset of any clock. It would, then, seem desirable to determine some temperature/frequency stability relationship and, together with, a model of spacecraft temperature, determine the changes in clock frequency due to temperature as a function of time. This approach is extremely complex and, when considering quartz or rubidium cell devices, is not feasible. For, although a particular unit may exhibit a positive temperature coefficient on days 1-20, it is perfectly possible that it may change to a negative coefficient on day 21. This is due to component ageing and/or changes in the operating conditions of the oscillator. This fact makes deterministic modeling of clock behavior as a function of temperature essentially impossible. Therefore highly precise temperature control is required.

Other environmental influences are quite analytic in their effect on the standard. Changes in magnetic field, pressure, and line voltage can all be related to physical parameters of the clock. But the magnitude of the effect of changes in these parameters when they are reasonably held constant is negligible when compared with the clock noise. What is true is that clocks will perform within a certain specification when the environment is likewise maintained within a certain specification. In addition, certain units may behave more reasonably than others over changes in environment, whereas other units may change very erratically with environmental changes.

In the following paragraphs, we will summarize some of the environmental testing performed on two different makes of rubidium clocks. One unit is the Efratom rubidium clock, and the other unit is the Collins AFS-81 rubidium standard. This testing was performed at Naval Research Laboratory and General Radio (builders of the basic rubidium unit for Collins), respectively. It was

from discussions with Mr. Herb Strademeyer of General Radio, who tested over 150 rubidium standards environmentally, that it was concluded that deterministic modeling of the temperature effects could easily prove futile.

Figure 5 illustrates the different behavior observed in two Efratom units as a function of temperature. As can be observed in this figure, unit 124 behaved quite linearly, whereas unit 120 exhibited extremely erratic behavior. Figure 6 is a pair of histograms indicating the distribution of frequency shift of approximately 50 Collins standards. This frequency shift is measured relative to $+25^{\circ}C$ and is shown for $+65^{\circ}C$ and $-55^{\circ}C$. The interesting aspect of this figure is to confirm that the performance of clocks is dependent on the individual unit.

Table 2 summarizes some of the results of the environmental testing at NRL and General Radio. It is given to illustrate the magnitude of change which has been observed in rubidium standards in non-ideal environments.

TABLE 2

SUMMARY OF ENVIRONMENTAL EFFECTS

	EFRATOM	COLLINS AFS-81	HP CESSIUM +
Temperature	Figure 4	5×10^{-13} $\Delta f/^{\circ}C$	$< \pm 5 \times 10^{-12}$ $0\text{-}50^{\circ}C$
Pressure	5×10^{-13}/MB	----	----
Magnetic Field	* $1 \times 10^{-12}/AM^{-1}$	----	$\pm 1 \times 10^{-13}$/GAUSS (dc field)
Line Voltage	----	6×10^{-13}/dc volt	----
Radiation	Figure 7	----	----
Acceleration	$4 \times 10^{-12}/G$	----	----

* Efratom Specification
\+ HP Specification

Figure 6 illustrates the effect of radiation on the Efratom unit as measured at NRL.

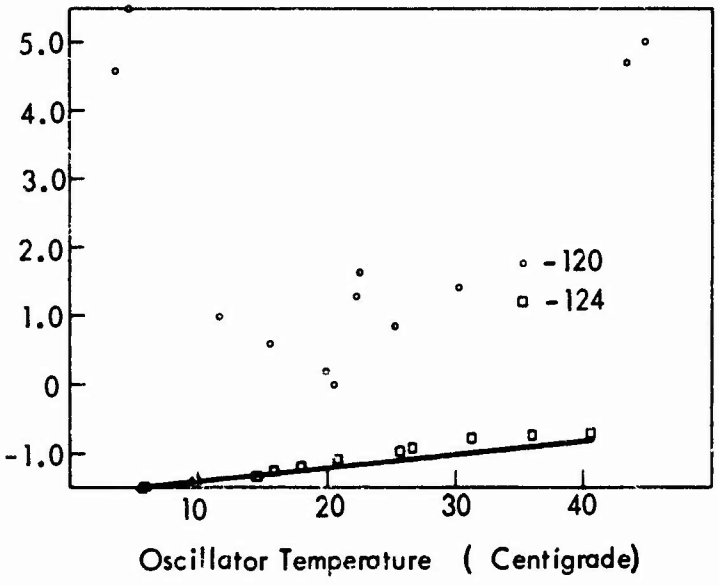
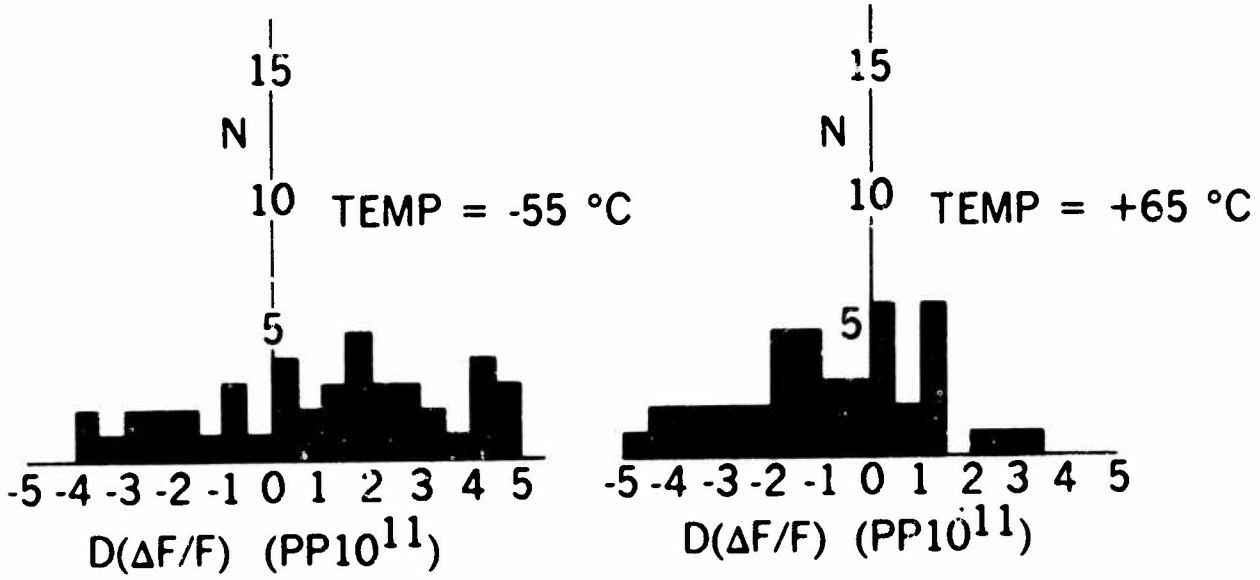


Figure 5 Temperature Stability Tests - Two Efratom Units (from Frequency Symposium 1973, Fort Monmouth)



QTY = 44 UNITS

N = NUMBER OF UNITS

D(ΔF/F) = FRACTIONAL FREQ SHIFT REFERENCED TO +25 °C

Figure 6 Histograms Showing Distribution of Frequency Stability vs. Temperature for Collis AFS-81. (from Collins Product Description - AFS-81)

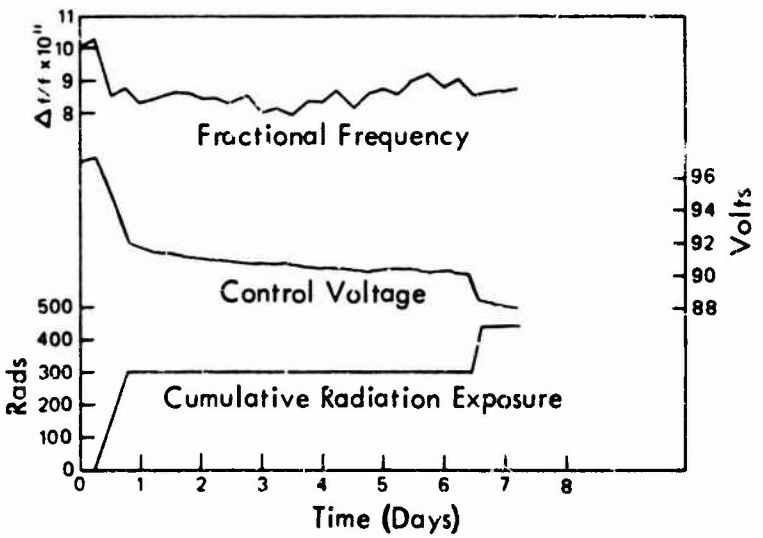


Figure 7 Irradiation test results, unit 120 Efratom (from Frequency Symposium 1973, Fort Monmouth)

(Reference 3)

4.2 Treatment of Environmental Effects

Thus far we have listed the kinds of environmental changes which affect clocks and given some examples of the magnitude of the effects. In this section we will describe what efforts can be applied to attempt to limit the impact of the environment on timing error. Two actions are proposed to accomplish this. First and most obvious is the spacecraft design should be such that any possible environmental changes are minimized. Good design practice should be able to nearly eliminate instability due to changes in supply voltage, magnetic field, radiation and hopefully temperature. Second the effect of environmental instabilities not controllable through spacecraft design initially will be modeled as random walk FM noise on the clock with a limited frequency spectrum. A routine for optimum prediction of the effects of random walk FM (or white noise $\dot{f}$) should be incorporated into the time dispersion prediction algorithm described in Section 3.0. As stated previously, deterministic modelling of environmental changes would require a great deal of telemetry and prediction of the environment for a small payoff. One of the largest environmental effects - temperature, cannot (for rubidium and quartz) be expected to always exhibit a constant temperature-frequency relationship. Treatment of the environmental effects as random walk FM is straight forward and should be easily implemented.

The optimum predictor for random walk FM or white noise $\dot{F}M$ is merely the linear predictor of the frequency rate term of expression 6, D_{ij}. Expression 13 is the optimum estimate of the frequency drift in the presence of random walk FM.

$$\hat{D}_{ij}(t - \frac{\tau_c}{2}) = \frac{R_{ij}(t) - R_{ij}(t - \tau_c)}{\tau_c^2} \qquad (13)$$

$\hat{D}_{ij}$ is the estimate on the frequency drift due to white noise $\dot{F}M$, and R_{ij} are measurements of the clock frequency difference over the calibration time τ_c. If random walk FM is present either due to environmental effects or present

in the noise of the oscillator then $\hat{D}_{ij}(t - \frac{\tau_c}{2})$ the mean value of D_{ij} over the interval τ_c will be the best prediction for the drift rate. The mean square timing error for this predictor has not been calculated but is available from Reference 2.

The question remains whether the sum of environmental effects (temperature, fields, radiation, etc.) approaches a random walk process over the period of prediction (i.e. between updates). The environment is bounded whereas random walk type noise would imply steadily increasing errors with time ($\sigma^2(\tau) = \frac{2\pi^2}{6} h_{-3}\tau$). However during the period between updates, it appears that although the environment may be somewhat controlled, environmental changes will be slow and somewhat random. Test data presented previously on temperature frequency relationships in rubidium standards indicates that the prediction of temperature influences may or may not be causal. The determination of environmental effects will not be answered until sufficient data is gathered under simulated and real conditions to ascertain the types of noise perturbing the clock behavior. In addition to the environment clock relationships the actual spacecraft environment will to a certain degree behave randomly (radiation for example).

It is in light of the extreme difficulties which would be placed on an operational system (in particular the user) in attempting to predict environmental effects, it seems more reasonable to control the environment and determine the noise characteristics of the clock in this environment. With this information it will then be possible to determine the optimum prediction scheme for the environment effects.

4.3 Effects of Physical Update of the Clock Frequency

In order to minimize the time differences between the various clocks in the GPS and in order to maintain magnitude of corrections which a user must apply to his range data for clock drift, it is tempting to provide in the satellite, some means for changing the clock frequency or phase. This would allow the master station a means of aligning all clocks in the system at periodic intervals. A great deal of caution should be exercised in the

techniques used in physically updating the satellite timing system. Precise knowledge of what changes are being made is absolutely necessary.

There are two approaches to the update of the satellite clock. The first approach would be to physically change the frequency of the clock itself. For instance, with a rubidium, the current of the C-Field might be changed to change the output frequency. The second approach would be to allow the clock to run freely and to use some external device (e.g. a phase stepper) at the output of the clock and periodically adjust this device to affect the frequency change. Whichever approach is chosen, any change to clock frequency will result in the necessity to update the clock model. If the resolution and knowledge of this frequency stepper is not sufficient (as defined by the desired error budget for clock drift) it could well require many days to recalibrate and determine the coefficients of the user clock model.

For example if the changes are of the order of 1×10^{-10} resolution in frequency, and the actual change is 1.01×10^{-10}, the resulting error in the clock prediction algorithm after 10^{+5} seconds would be 100 nanoseconds solely from imprecise knowledge of the rate offset of the system clocks. It may then require many days of measurements to determine that the actual frequency step was 1.01×10^{-10} not 1×10^{-10}. It is easily seen that it is necessary to be able to make frequency changes of 1×10^{-14} resolution with precise knowledge in order to keep this source of error negligible (1 nanosecond) over 10^{5} seconds.

Lastly although no direct proof exists one way or the other it does not seem inconceivable that using the approach of directly changing the operating conditions of the clock, that this would also change the noise characteristics of the clock too. This would require changes in the prediction schemes to optimize the prediction filters. Any technique which operates on the clock directly should be thoroughly tested to determine its effects on the clock noise characteristics of the unit being adjusted.

5.0 LABORATORY MEASUREMENTS AND COMPUTED SIMULATION OF USER ERRORS

This section presents the preliminary results of an effort to predict time dispersion based on measurements made between two rubidium clocks; and a computer simulation using TRACE to asses the clock effect on user position errors.

5.1 Laboratory Measurements and Derived Clock Model

A measurement of accumulated phase difference between a Efratom Rubidium standard and an HP 5065A Rubidium was recorded for a period of approximately 65 hours. The units were operated in an uncontrolled laboratory environment. The measurement set-up is shown in Figure 8. A flow diagram of the computer program used with the HP systems programmer is given in Figure 9. Approximately 10 hours of this measured data was then used to determine coefficients for a prediction algorithm of time dispersion as given in equation 6 and repeated here.

$$T_{ij}(t) = T_{ij}(t_o) + R_{ij}(t-t_o) \times t + \frac{1}{2} D_{ij} \times t^2 \qquad (6)$$

To evaluate the above coefficients, where $T_{ij}(t_o)$ - initial time difference, R_{ij} - frequency offset, and D_{ij} - aging difference, it was necessary to fit a straight line, equation 2, (using least squares) to $\Delta f(n\Delta t)$ data. Equation 3 was used to determine $\Delta f(n\Delta t)$ from the measured accumulated phase data.

$$\Delta f(t) = \hat{R}_{ij} + \hat{D}_{ij} t \qquad (14)$$

Where $\hat{R}$ and $\hat{D}$ are estimates of the frequency and aging offsets

$$\Delta f(n\Delta t) = \frac{(\Phi_m(n\Delta t) - \Phi_m((n-1)\Delta t))K}{\Delta t} \qquad (15)$$

Where $\Phi_m(n\Delta t)$ are measured values of accumulated phase, and K is a constant to convert to phase in degrees to frequency.

$T_{ij}(t_o)$ was chosen to equal $\Phi_m \cdot K$ for m = 0, i.e., the initial measured value of time offset.

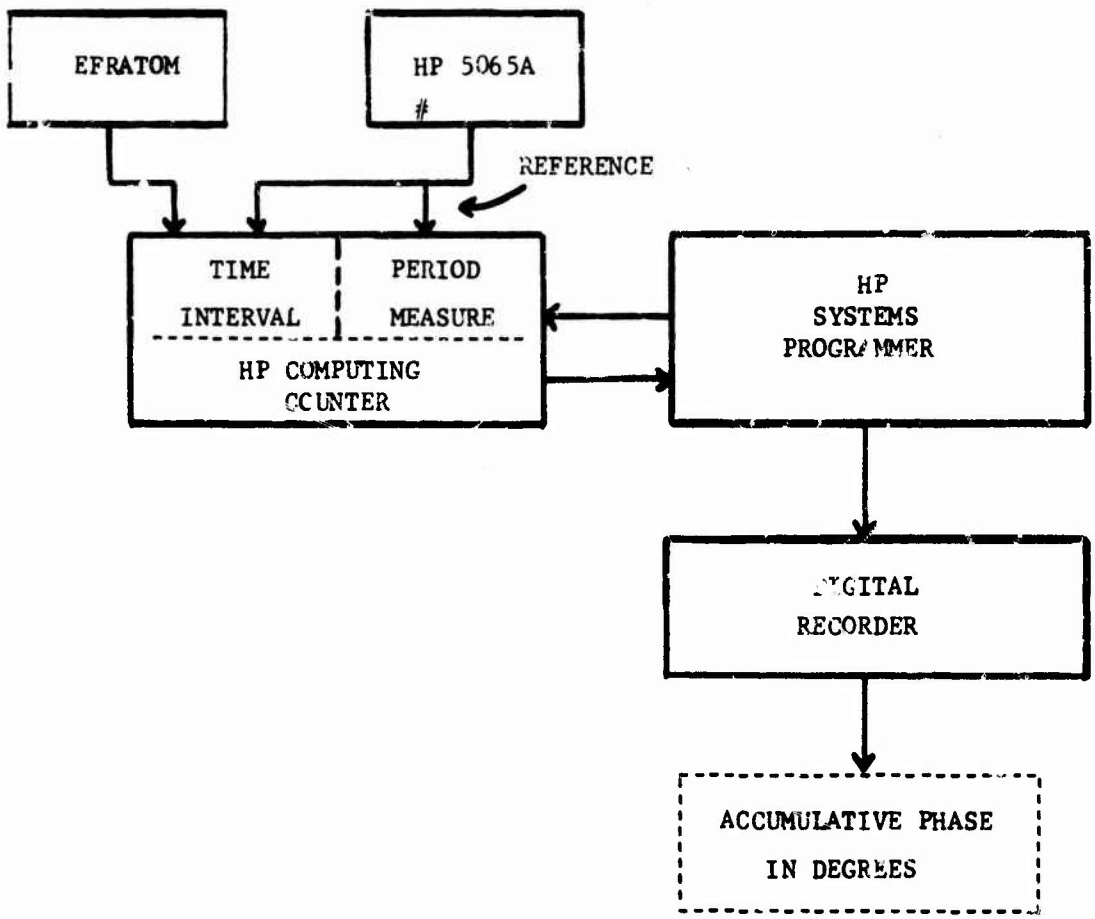


Figure 8 Laboratory Set-up

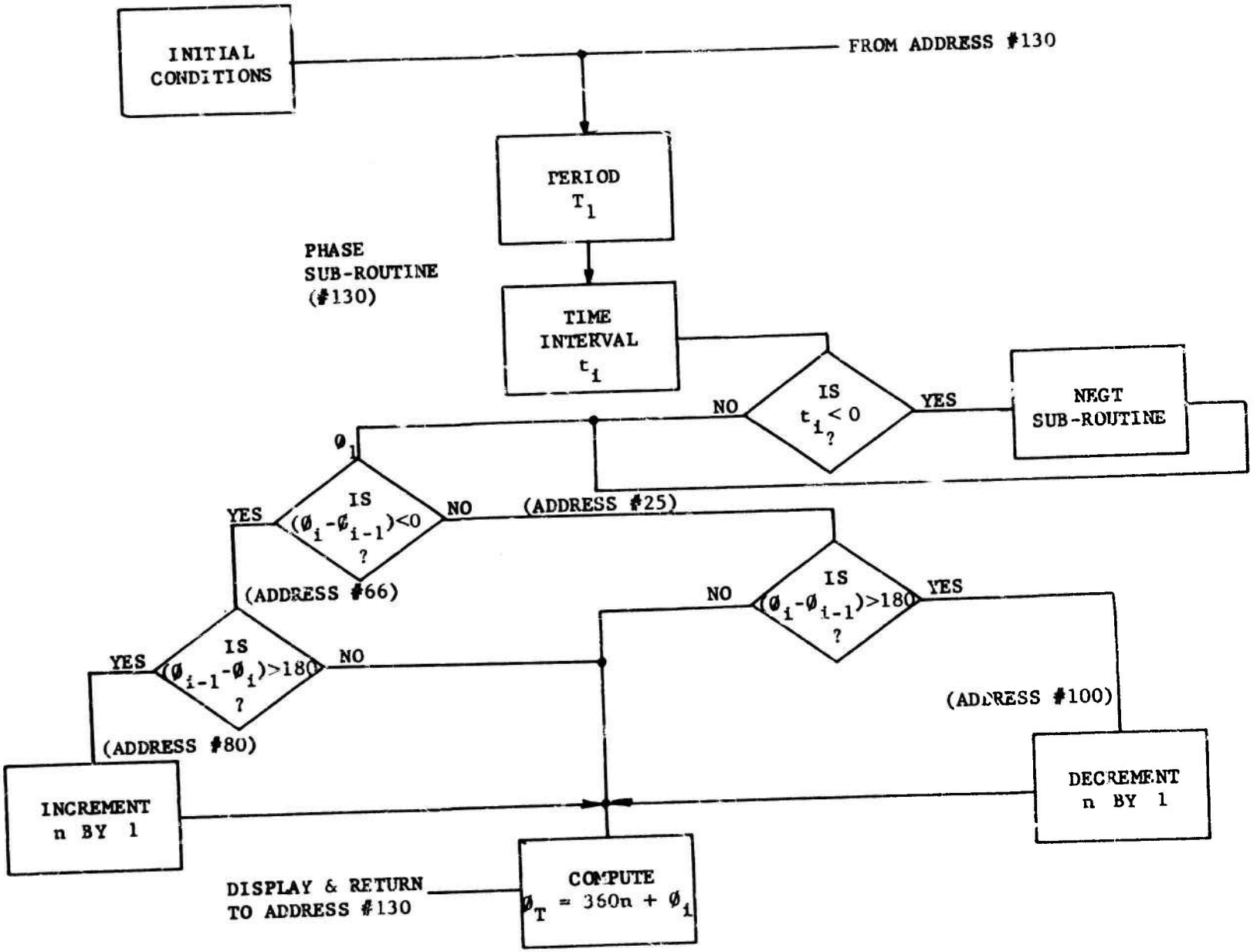


Figure 9 Flow Diagram for Accumulative Phase

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

The initial values determined for $\hat{R}_{ij}$ and $\hat{D}_{ij}$ from the least squares fit, is a good initial estimate on the deterministic characteristics of the clock. It was now desirable to improve this estimate through some means of filtering. Actually, the situation is a case of signal detection in the presence of noise. In our case, the desired signal is the frequency of the clock and the noise is a combination of white and flicker noise and as later measurements revealed, random walk FM.

As stated previously in Section 3, this problem has been studied extensively by Allan and Barnes, et.al., at NBS (Ref. 2). It was decided to use the exponential filter in order to improve our estimate of the rate offset R_{ij}. For this case no attempt was made to optimize the filter time constant $m\Delta t$ since we had not yet calculated the noise characteristics of these clocks. The exponential filter was used to weight the residuals of data and the initial estimate of the deterministic model. This was done as follows.

If we define the residual error in offset frequency as:

$$y(n\Delta t) = R - \Delta f(n\Delta t) \tag{16}$$

then the recursive filter expression weighting previous measurements to present measurements of frequency is given in equation 17.

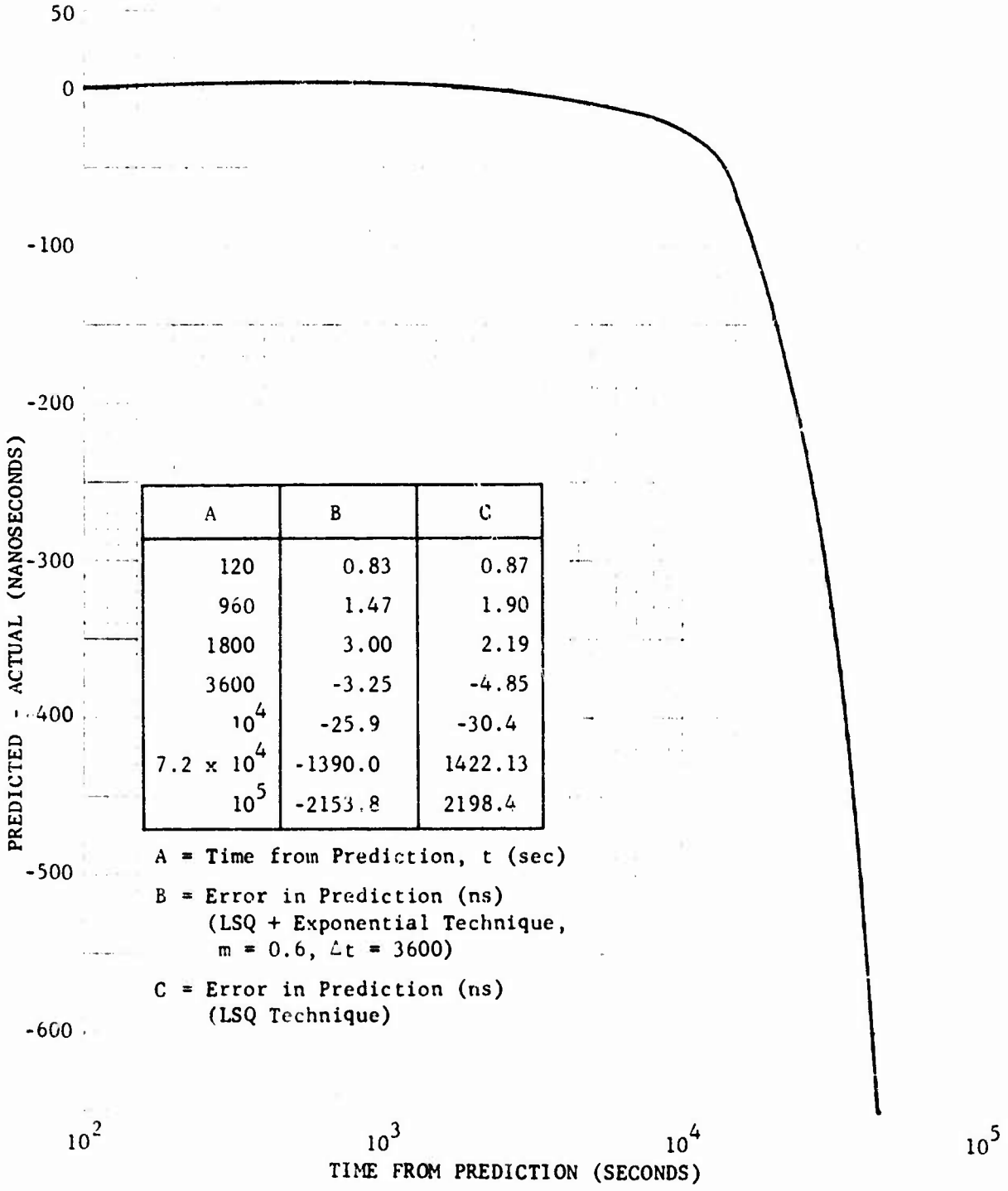
$$\hat{y}(n\Delta t) = \frac{1}{m+1}\left[y(n\Delta t) + m\,\hat{y}((n-1)\,\Delta t)\right] \qquad \begin{matrix} m = 0.6 \\ \Delta t = 3600 \text{ sec.} \end{matrix} \tag{17}$$

Thus, in equation 1, R - the frequency offset becomes $R(n\Delta t) = \hat{R} + \hat{y}(n\Delta t)$.

The above procedure (least squares fit to $\Delta f(n\Delta t)$) and the exponential filtering yielded a final estimate for $T_{ij}(t)$ given below:

$$T_{ij}(t) = -2.53 \times 10^{-6} - 3.849 \times 10^{-11} t + 1.129 \times 10^{-16} t^2 \tag{18}$$

This final expression was then used to predict ahead for 50 hours. The resulting errors $x_p(t) - \tilde{q}_m(n\Delta t)$ are plotted in Figure 10. In this case, no further data (beyond 10 hours) was used to update the estimate on R.



A	B	C
120	0.83	0.87
960	1.47	1.90
1800	3.00	2.19
3600	-3.25	-4.85
10^4	-25.9	-30.4
7.2×10^4	-1390.0	1422.13
10^5	-2153.8	2198.4

A = Time from Prediction, t (sec)

B = Error in Prediction (ns)
(LSQ + Exponential Technique,
m = 0.6, Δt = 3600)

C = Error in Prediction (ns)
(LSQ Technique)

Figure 10 Error in Prediction of Time Dispersion of Two Rubidium Clocks

Ideally, if measurements were to be made, say, every 10 hours, than any new values of $\Delta f(n\Delta t)$ should be used to improve knowledge of R in equation 1. Since we had only limited data, it was decided that it would be of more interest to examine the error associated with long-term prediction than to attempt to further improve the estimate of R.

It can be seen in figure 10 that for time greater than 10^4 seconds the timing error grew very rapidly. This rapid increase in timing error is due to the choice of time constant ($m\Delta t \cong 2000$ sec.) of the exponential filter. Had our data been processed by a bank of filters with increasing time constants, it is believed this error would have been considerably less and only limited by the amount of data used in calibration (i.e. 10 hours).

5.2 Calculation of Short Term Stability

Utilizing this same 65 hours of measured accumulative phase data obtained on the Efratom and the HP rubidium standards, the short term stability (Allan variance) was calculated for averaging times from 120 seconds to 5×10^4 seconds. Equation 1 expresses f_{offset}, the frequency offset between the HP rubidium and the Efratom, in terms of the measured data.

$$f_{offset} = \left[\frac{\left(\Phi_m(n\Delta t) - \Phi_m(n-1)\ \Delta t \right) K}{\Delta t} \right] f \qquad (19)$$

where $\Delta t = \tau$ (averaging time)
$K = 2.78$ nanoseconds/degree
$f = 1$ MHz

The Allan variance is defined again here in equation 20.

$$\sigma^2_{2,\tau} \left(\frac{\Delta f}{f} \right) = \frac{1}{f} \left(\frac{(f_1 - f_2)}{2} \right)^2 \qquad (20)$$

In this expression f_1 and f_2 are adjacent measurements of the oscillator, under test, against a perfect standard for a measurement time of τ seconds. In general, since only two frequency measurements are used in calculating

the Allan variance, it is necessary to average $\sigma^2_{2,\tau}\left(\frac{\Delta f}{f}\right)$. Thus the expression for $< \sigma_{2,\tau} \left(\frac{\Delta f}{f}\right)>$ is shown in equation 21.

$$< \sigma_{2,\tau} \left(\frac{\Delta f}{f}\right)> = \frac{1}{f} \sqrt{\frac{1}{2N} \sum_{i=1}^{N} (f_i - f_{i-1})^2} \qquad (21)$$

where f_i is the i^{th} frequency offset as given in equation 19.

A short computer program was written to calculate $< \sigma_{2,\tau}\left(\frac{\Delta f}{f}\right) >$ from the measured values of accumulative phase. The results of these calculations are plotted in Figure 11. Note that curve 1 represents the short term stability of both units under test. Curve 2 merely assumes that the statistics of the noise for both units are identical and independent of each other. In this case, we have plotted

$$\frac{< \sigma \left(\frac{\Delta f}{f}\right) >}{\sqrt{2}} .$$

This last assumption may not be totally valid for values of $\tau > 1000$.

The interesting observation from this figure is the presence of random walk FM noise. This is indicated th the upswing of the fractional frequency deviation $\sigma(t)$ at about 10^4 seconds. There are a number of possible causes for this upswing or apparent presences of random walk noise. For example both units were in an uncontrolled environment which might have caused this effect. Also both units were reasonably new and possibly not totally aged which may have also caused this behavior, that is, the deterministic behavior of the clocks may not have been totally eliminated based on the coefficients determined previously. And lastly, these calculations were made on a limited data base and for the long averaging times 10^4 seconds or more, few data pairs were available to calculate the variance. Thus the statistical confidence in the latter portion of this curve is limited.

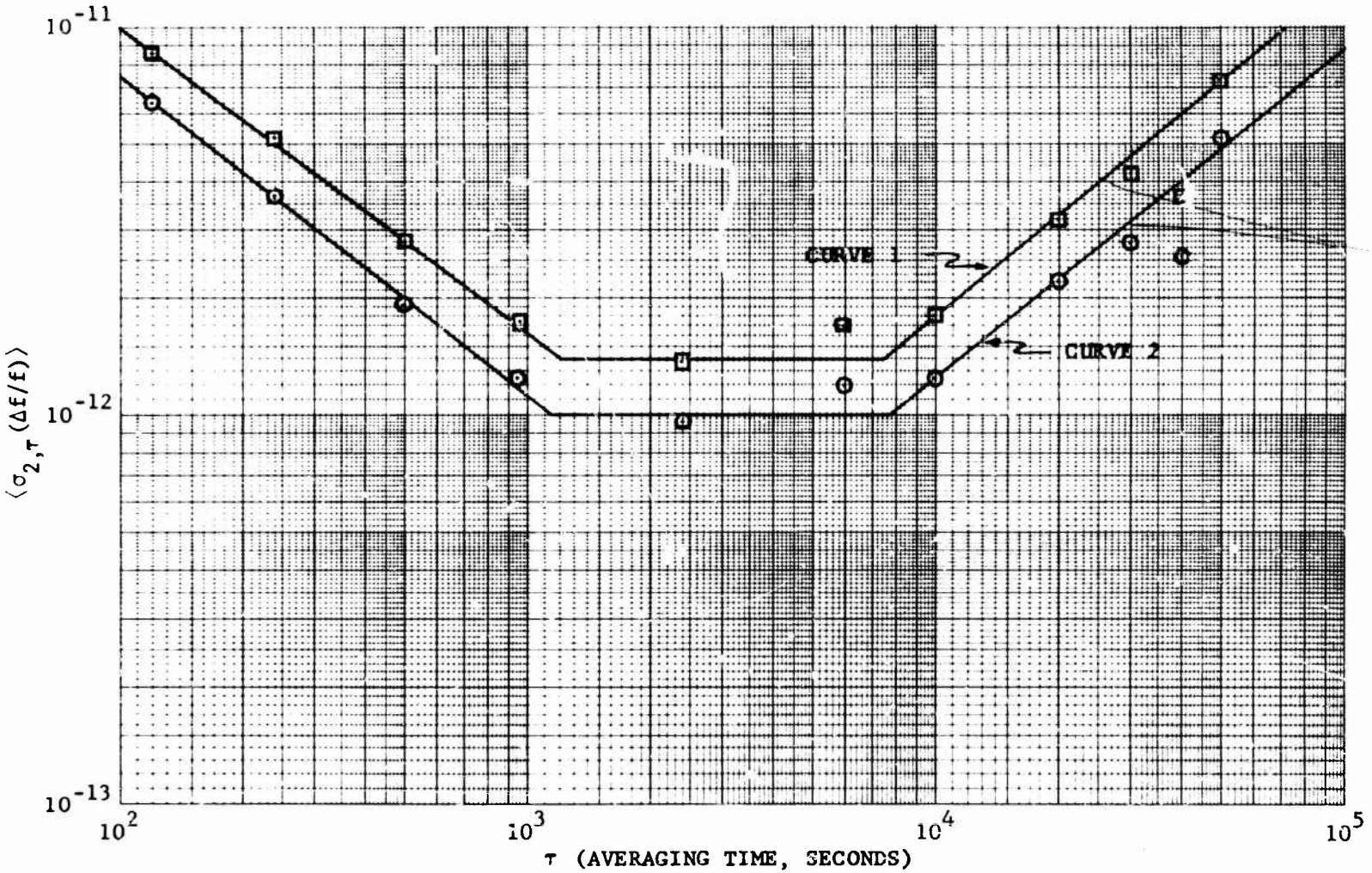


Figure 11 Short-Term Stability versus Averaging Time

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

5.3 Computer Simulation of Effect of Clock Errors on User Position

In order to assess the effects of realistic clock errors, and to relate these to user navigation error, a simulation has been performed using the TRACE orbit determination computer program in a sequence of cases. The sections which follow present a discussion of the clock error model, description of the TRACE simulations and a summary of the computed results.

It should be pointed out that the results obtained are for individual cases, with data generated to include the clock errors in a pseudo-random sequence of additive noise, with the appropriate characteristics, imposed on top of the ideal observations. The user navigation error was computed by taking actual deviations of the user solution with noisy data from the known user location. This is in contrast to the covariance analysis methods used in assessing other effects, where the propagation of errors through the system is modeled analytically, producing ensemble statistics for the error characteristics. In the present case, the results have been determined for a particular sequence of random effects, controlled in size and judged to be typical in nature. In order to guarantee statistical significance, several sample cases should be executed with different random sequences.

5.3.1 Simulation by the Program TRACE

The clock error model implemented in TRACE is defined by (6) where the random process Y(t) is given by:

$$Y(t) = Y_1(t) + Y_2(t) * Y_3(t)$$

where Y_1, Y_2, and Y_3 are random processes which are independently restricted to be either white, colored or integrated white noise. (Also the possibility of simply adding Y_1 to Y_2 is provided.)

Each of these three random processes is defined by two parameters. For instance Y_1 is defined by the parameters C_3 and C_4 according to the following conventions:

1) $C_3 = 0$ Process is white noise and C_4 is the variance of this white noise

2) $C_3 > 0$ Process is colored noise and C_4 is the variance of this noise. In this case the autocorrelation function is given by:

$$Y_1(t+\tau)\,Y_1 = C_4^2 \exp(-\tau/C_3)$$

Thus C_3 is in fact the correlation time.

3) $C_3 < 0$ Process is Integrated White Noise and C_4 is equal to $(h_0/2)^{\frac{1}{2}}$, the coefficient of $\sigma y(t)$. In this case C_3 has no meaning (used as a flag). At time t, the variance of Y_1 is in this case:

$$\sigma_Y(t) = \langle Y_i^2 \rangle^{M} = C_4\, t^{\frac{1}{2}}$$

The above three processes are generated in TRACE using a single step recursive algorithm with a random number generator as input to $\sigma y(t)$.

5.3.2 Model Adequacy

Since TRACE does not have the capability of generating Flicker noise, the question arises whether the available models may still be used to obtain some approximation of the clock behavior. As it appears on Figure 2, a IWN (Integrated White Noise) model is quite adequate for a given span of observation time, depending of the type of clock, before Flicker effects become predominant.

Since a period of 48 to 72 hours is considered, it is obvious that the IWN model will be appropriate for the cesium clocks exhibiting a moderate Flicker component. Simply ignoring the h_{-1} term will lead to a poor approximation in the case of High Performance cesium clock and cannot be reasonably thought of for Rubidium clocks.

However, by adjusting the value of C_4 so as to match some "average" of the variance, it may be possible to obtain some indication of the effects of clock errors. If C_4 is chosen equal to 3.10^{-11}, the cesium clock is well represented, and the HP cesium is to some extent. By increasing C_4 by an order of magnitude, some simulation, probably pessimistic in this case, of the rubidium clock noise can be obtained. On Figure 12, the lines are drawn corresponding to those two values of C_4, as they have been actually used in TRACE runs.

Assuming only white and flicker noise FM the variance of the clock noise is:

$$\sigma^2(t) = (h_0/2)\ t + (2Ln2h_{-1})\ t^2$$

while using a pure IWN model leads to:

$$\hat{\sigma}_Y^2(t) = C_4^2\ t$$

The choice of C_4 may be made by imposing the condition that real and approximated variances have the same average over the observation time T_o, ie:

$$\int_{t=0}^{t=T_o} \hat{\sigma}_Y^2(t)\ dt = \int_{t=0}^{t=T_o} \sigma_Y^2(t)\ dt$$

In this case the value of C_4 is as given in the expression below:

$$C_4 = (h_0/2 + (2/3)\ (2Ln2h_{-1})\ T_o)^{\frac{1}{2}}$$

Applying this to the values of Table I, with $T_o = 10^5$ seconds, leads to the following values for the noise coefficient C_4:

TABLE 3 Values of the Noise Coefficient

Clock Type	C_4
Rubidium	$1.1\ 10^{-10}$ to $3.5\ 10^{-10}$
Cesium	$3.0\ 10^{-11}$ to $1.1\ 10^{-10}$
HP cesium	$1.1\ 10^{-11}$

The units for C_4 are $(sec)^{\frac{1}{2}}$.

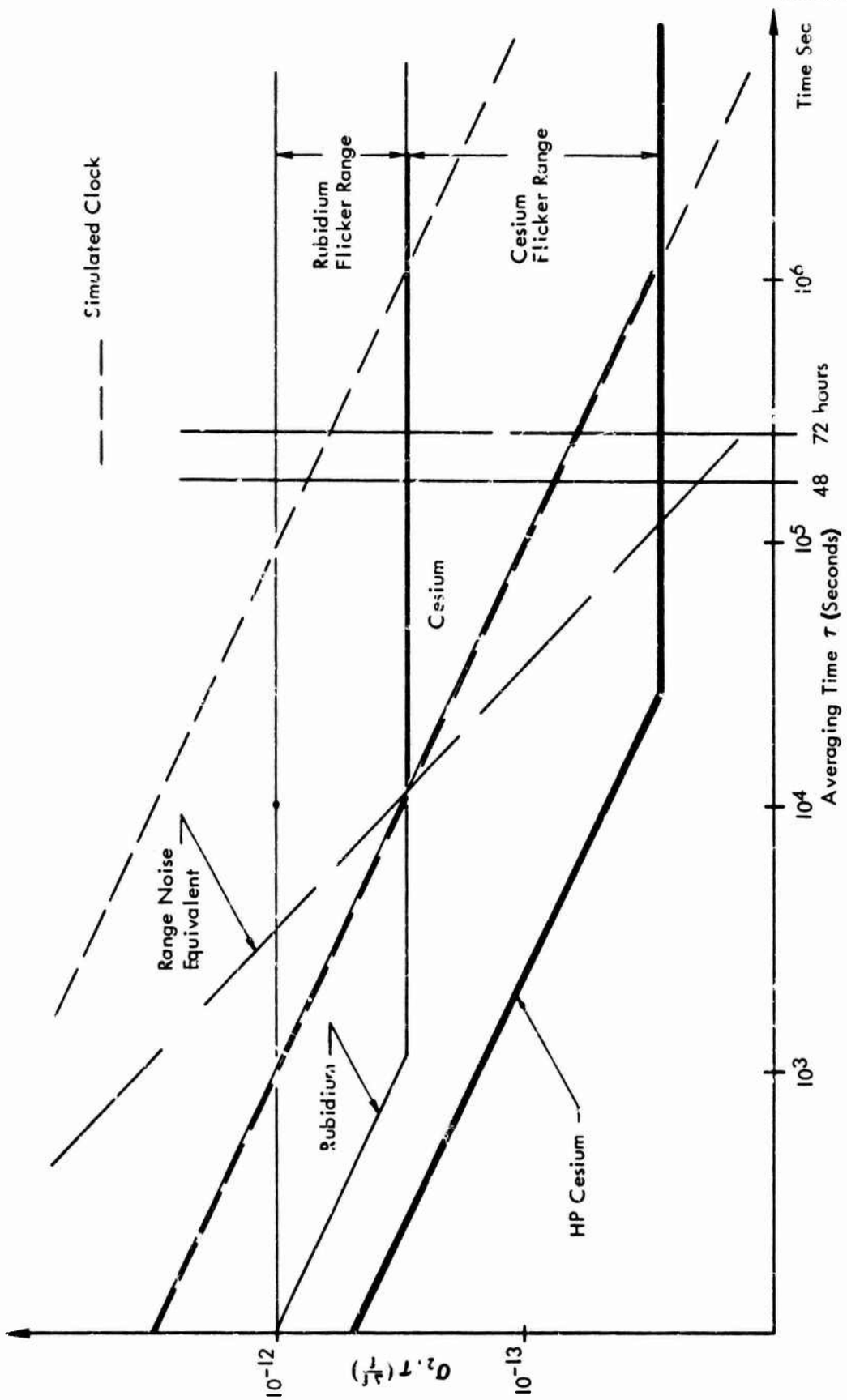


Figure 12 Allan Variance of Real and Simulated Clocks

5.3.3 TRACE Simulations

5.3.3.1 Data Generation Runs

Timing errors may be simulated in TRACE for both station clocks (DT_s) and vehicle clocks (DT_v), and the corresponding error in range measurement computed according to:

$$DR = c_{light} (DT_s - DT_v) = \dot{R} DT_s \qquad (22)$$

Also, a random error (zero mean, Gaussian white noise) is added to the range with a given standard deviation (given in SIG and KSIG TRACE input parameters).

The total range error may thus be written as:

$$DR = R_{0s} + R_{1s}t \pm R_{2s}t^2 + N_s(t) + R_{0v} + R_{1v}t + R_{2v}t^2 + N_v(t) + N_r(t) \qquad (23)$$

where N_s and N_v correspond to station and vehicle clock noise and N_r to the random range error. Because of the correspondence between clock timing error and range error depicted by Eq. 23, N_r may be expressed in terms of timing error and its variance plotted against time. This is also shown in Figure 12 for a sigma of $5\sqrt{2}$ ft (the $\sqrt{2}$ comes about because of two clocks being used).

Range observations have been generated for four stations and four vehicles. They cover 48 hours and, for the next 24 hours single observations have been generated for five users at the following times:

User	US1	US2	US3	US4	US5
Time (h/mn)	3 00	8 59	14 58	20 57	26 56

User locations were chosen to insure identical configurations of the vehicles in the five cases and are given in appendix 1 of this report.

Three runs have been made to generate three different observation tapes with the following values for the vehicles and station clocks integrated white noise coefficient C_4:

Tape #1	$C_4 = 3.10^{-11}$	
Tape #2	$C_4 = 3.10^{-10}$	
Tape #3	$C_4 = \ .0^{-20}$	(This low value was chosen rather than a hard zero to preserve the similitude between the random sequences)

Basically, Tape #3 corresponds to an ideal case with no clock error, as would be assumed in a covariance analysis. Tape #1 simulates the case of cesium clocks and Tape #2 approximates the effects of rubidium clocks.

5.3.3.2 Orbit Determination Runs

TRACE does not have the capability of directly solving for the clock parameters (i.e., C_0, C_1 and C_2). However, it can solve for range bias, drift and drift rate, and, as seen from Eq (22), one can establish a direct correspondance between clock and range parameters. Indeed, neglecting R (as much smaller than the speed of light c), it follows:

	Station Parameters	Vehicle Parameters
Bias	RBIA $\sim$ c C_{os}	VSB $\sim$ -cC_{ov}
Drift	RBD $\sim$ c C_{1S}	VSBD $\sim$ cC_{1v}
Drift rate	RBDD $\sim$ c C_{2s}	VSDD $\sim$ -cC_{2v}

Orbit determination runs were performed using Tapes #1, 2 and 3 previously obtained and solving for the following parameters:

a) RBIA and RBD for all stations except VTS

b) Vehicles

b.1	Orbit parameters	AF, AG, N, L, CHI, PSI
b.2	Perturbation parameters	CPAW
b.3	Equivalent clock parameters	VSB and VSBD

c) For all users

c.1	Position	LAT, LONG, and ALT
c.2	Equivalent clock parameters	RBIA

The a priori sigma's corresponding to those parameters are defined in Reference (4).

5.3.3.3 General Modus Operandi

All station, vehicle and user parameters were solved for simultaneously in a batch least squares mode. Since user observations amount to only 5 x 4 = 20 (against about 1000 made by the stations during the 48 hours period) they don't affect noticeably the solution for the orbit, station and vehicle parameters (as confirmed separately by a control run), and this global method being more efficient and easy, it has been used systematically for the three orbit determination runs.

The solution for the user parameters is however, obviously based upon the estimated values of vehicle parameters since it has to extrapolate clock errors and vehicle position according to the estimated values for clock bias and drift, and of orbital elements. This estimate of clock bias was made by least squares fitting of a straight line to the simulated accumulative clock phase data. Fitting to the phase data as opposed to the frequency data as done in the previous section (5.1) leads to an intrinsic timing error in the clock model. (NOTE: The recommendeded technique is to least squares fit to frequency data, integrate the curve and evaluate the constant at time t_c (the end of the calibration period). In this way, the time offset T_{ij} of the clock model is essentially without error at the start of the prediction interval. Curve fitting to the phase data as was done in this instance leads to an approximation to the time offset which led to nearly a 10 nanosecond time error at the start of the prediction period.) This is quite representative of the actual situation where the users only data will be range observations plus the orbital elements and updated clock coefficients (VSB, VSBD, etc.) furnished through the vehicles. By using data progressively polluted by clock noise, it is thus possible to directly follow the effect of this noise upon the user errors.

It should be stressed at this point that, while a covariance analysis run will use a priori estimates of range errors assuming basically a Gaussian stationary noise, the present Orbit Determination runs deal with real range noise which is non-stationary as has been previously seen.*

5.3.4 Results of the Runs

Three Orbit Determination runs have been [illegible]using observation Tapes #3, 1 and 2, for demonstrating the effect of clock noise upon the different parameters of interest.

5.2.4.1 Station and Vehicle Clock Parameters

Accurate prediction of the clock state relies on the accurate determination of the station and vehicle clock parameters. The effect of increasing clock noise is shown in Figure 13 for range bias (equivalent to clock bias C_o) and on Figure 14 for range drift (equivalent to clock drift C_1).

The "standard deviation" line corresponds to the 5 ft sigma range noise, and it is the value which would normally be obtained in a covariance analysis run. It can be seen that the values for the parameter estimates are in good agreement with the predicted sigma when no clock noise is present. The effect of clock noise is then clearly visible as it increases the error in the estimates by about an order of magnitude from one run to the next.

Some further insight may be gained by observing how TRACE was actually fitting the data. On Figure 15 the actual clock error (as obtained from the Data Generation runs print out) has been plotted versus the observation time along with the linear behavior estimated by the program. Note that clock noise generates a deviation from the input clock frequency offset of $\Delta f = 1 \times 10^{-12}$.

* This statement may seem to be contradicted by Figure 12 where the variance of the ordinary range noise appears to follow a 1/t law. This is due to the fact that Figure 2 refers to frequency noise. The phase or range noise variance may be obtained by multiplying each curve by the time t, which re-establishes a constant sigma for direct range noise.

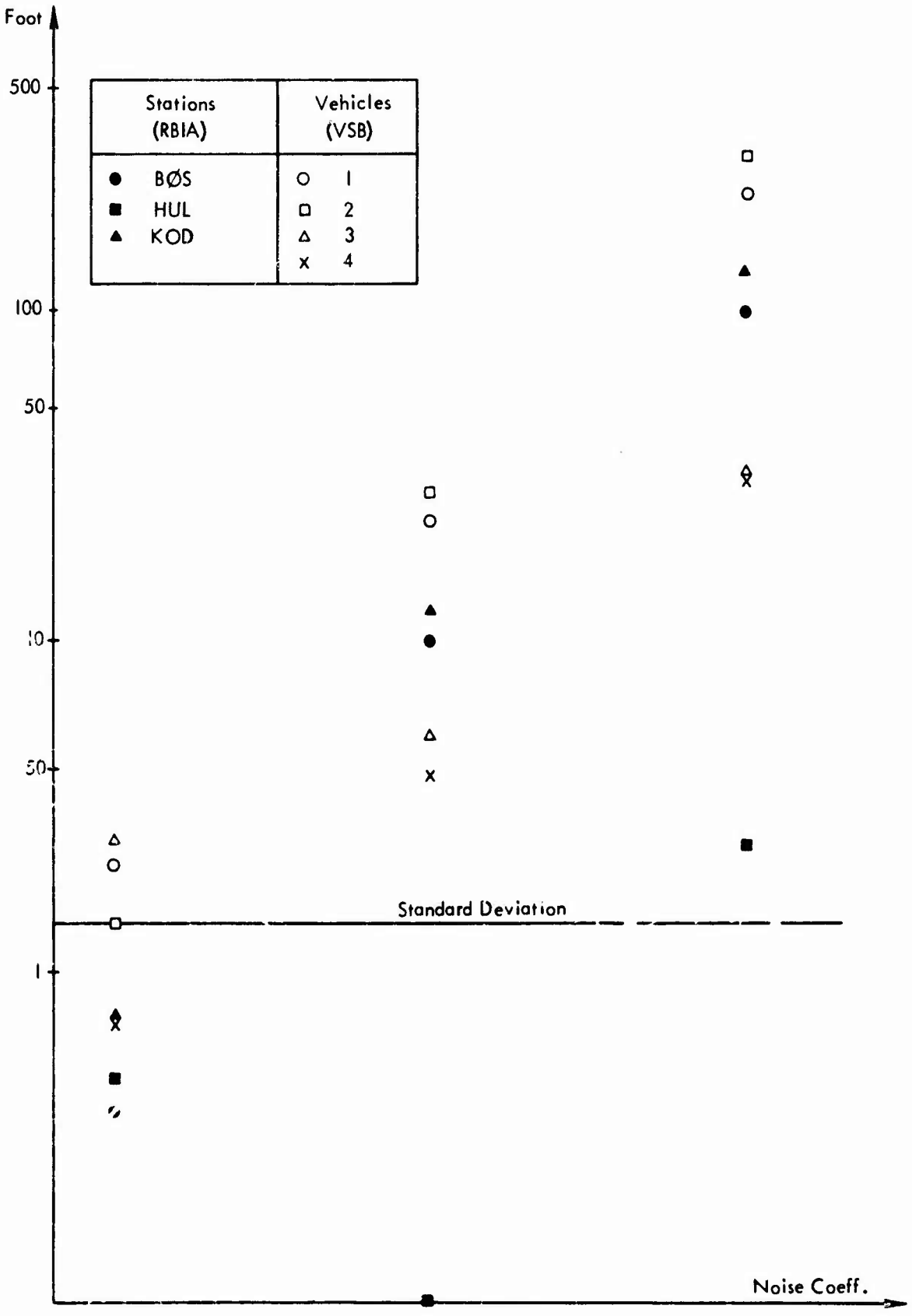


Figure 13 Error on Range Bias

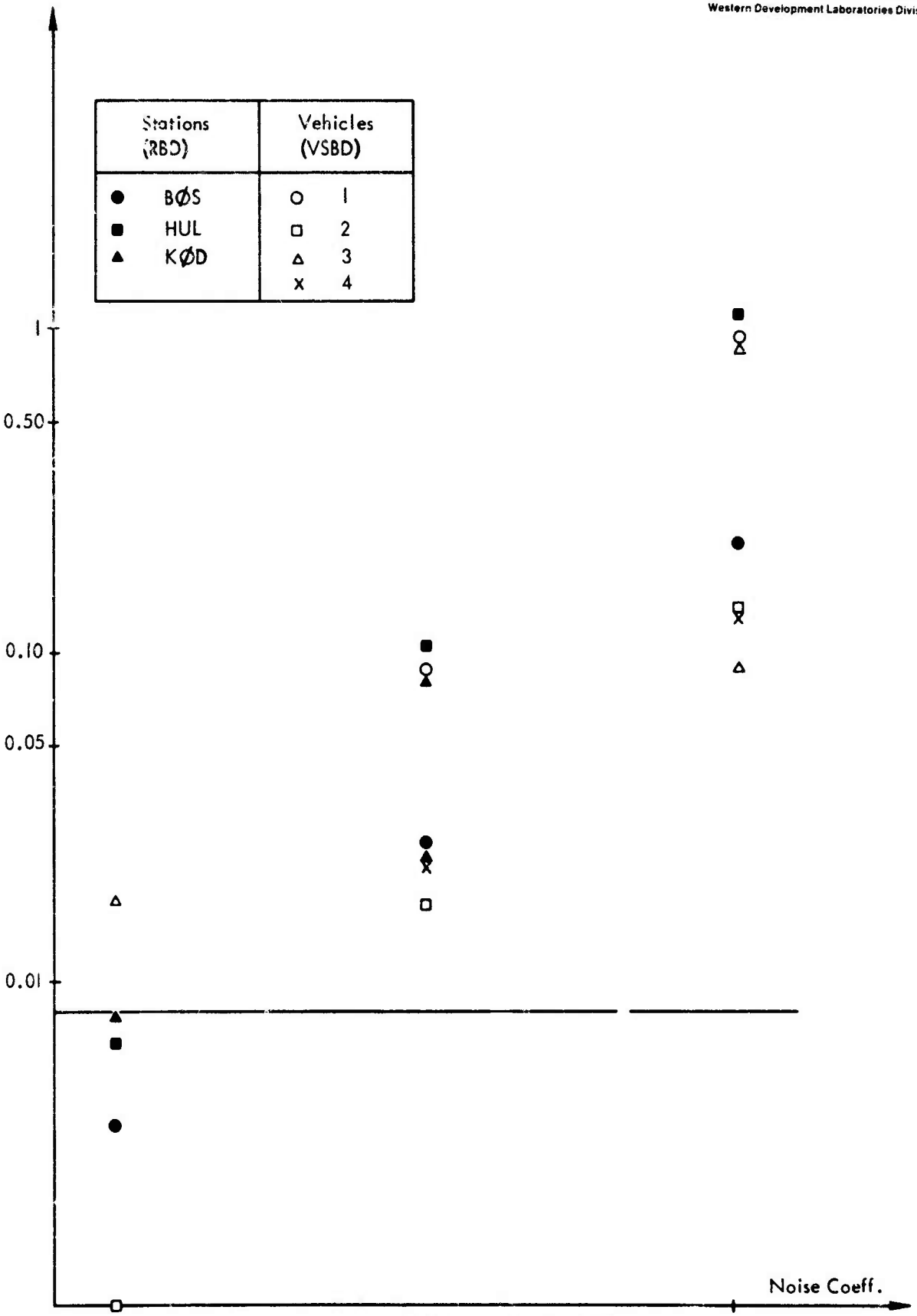


Figure 14 Error on Range Drift

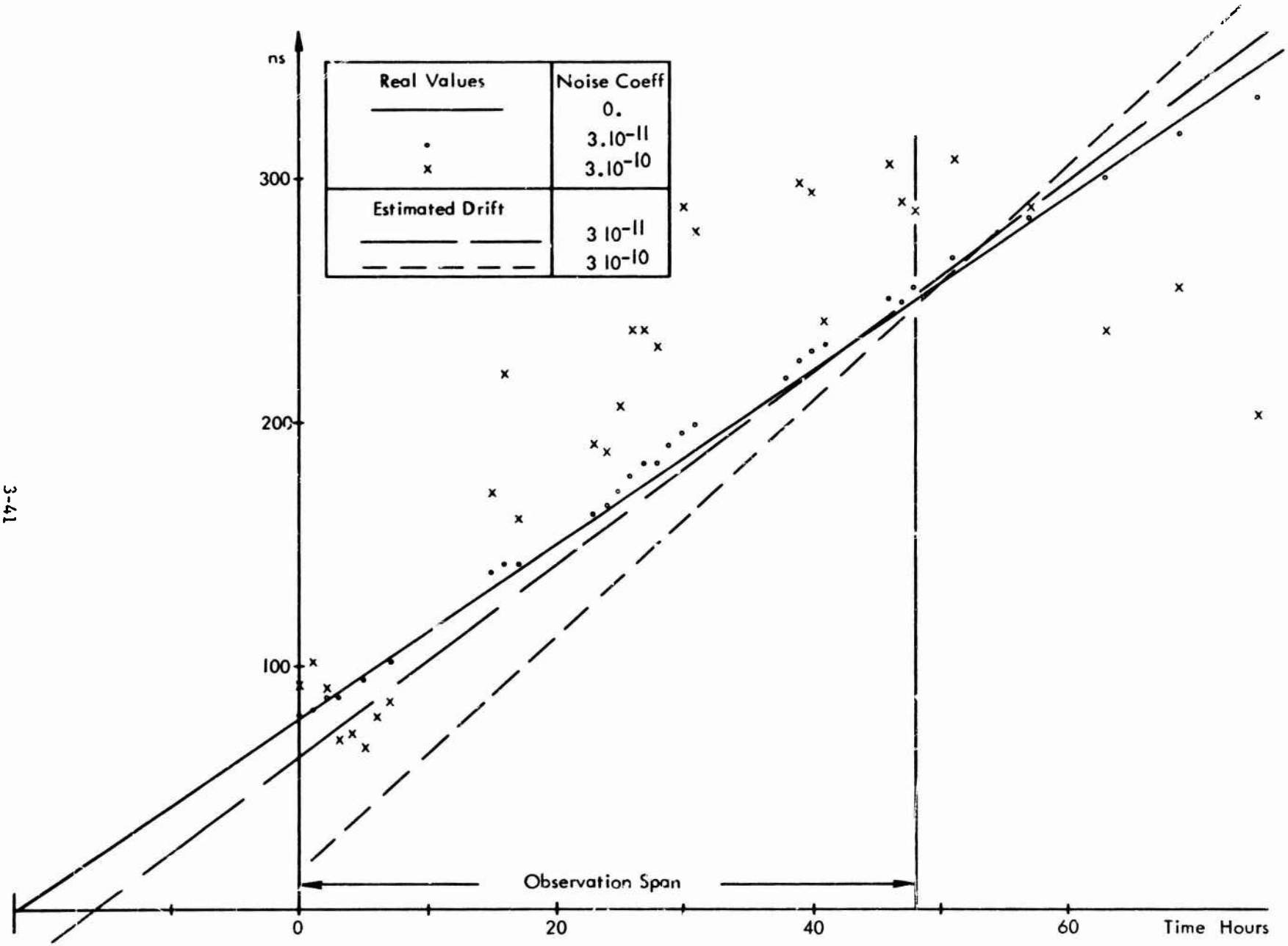


Figure 15 Clock Timing Error - Vehicle #1

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

5.2.2 User Parameters

Longitude and latitude errors are plotted on Figure 16 and 17 for the five users according to their position in time, to exemplify the degradation of the results with time. Again, the errors are within the expected sigma when no clock noise is applied, but are rapidly growing with C_4. As emphasized by the log-log scale, all the errors seem to be following at $t^{\frac{1}{2}}$ law.

Altitude and bias errors are shown on Figures 18 and 19. Here again, clock noise has, generally speaking, the same expected effects, i.e., errors are proportional to C_4 and $t^{\frac{1}{2}}$. However, it has to be understood that the observations generated represent only one of the many possible random sequences and particular results may exhibit significant deviations. Also, in the case of $C_4 = 3.10^{-10}$, it has been found that the algorithm required one more iteration to achieve convergence on the user parameters.

5.3 Concluding Remarks

This study has demonstrated the capability of simulating clock errors within the TRACE environment and shown the quite significant impact of these errors on the determination of user position. Though it has not been a statisitical study per se, the quantitative results obtained here may be considered as representative of the phenomenon, and are opening the way for more detailed and precise studies, both regarding the clock model and the data processing.

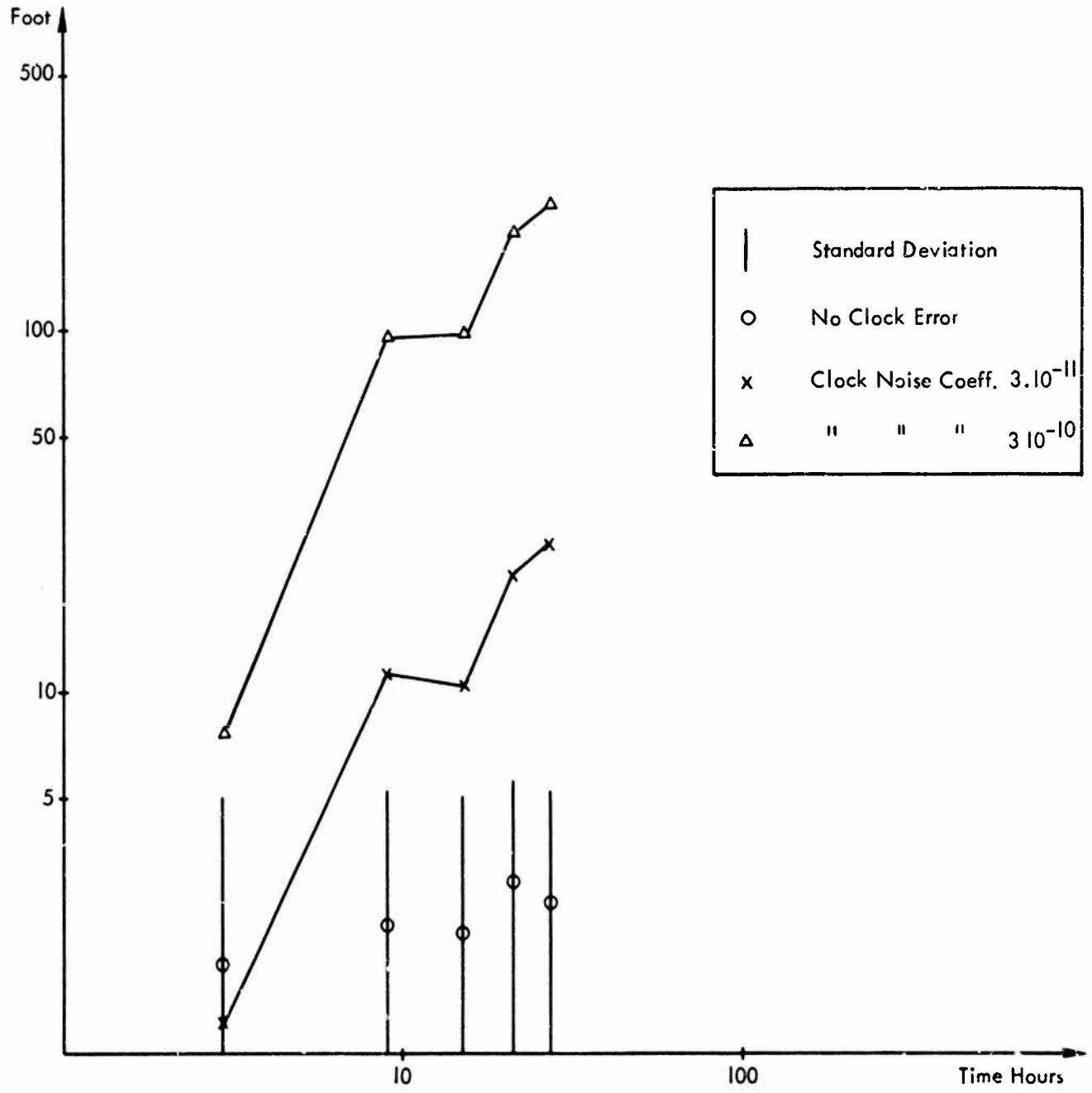


Figure 16 Longitude Error

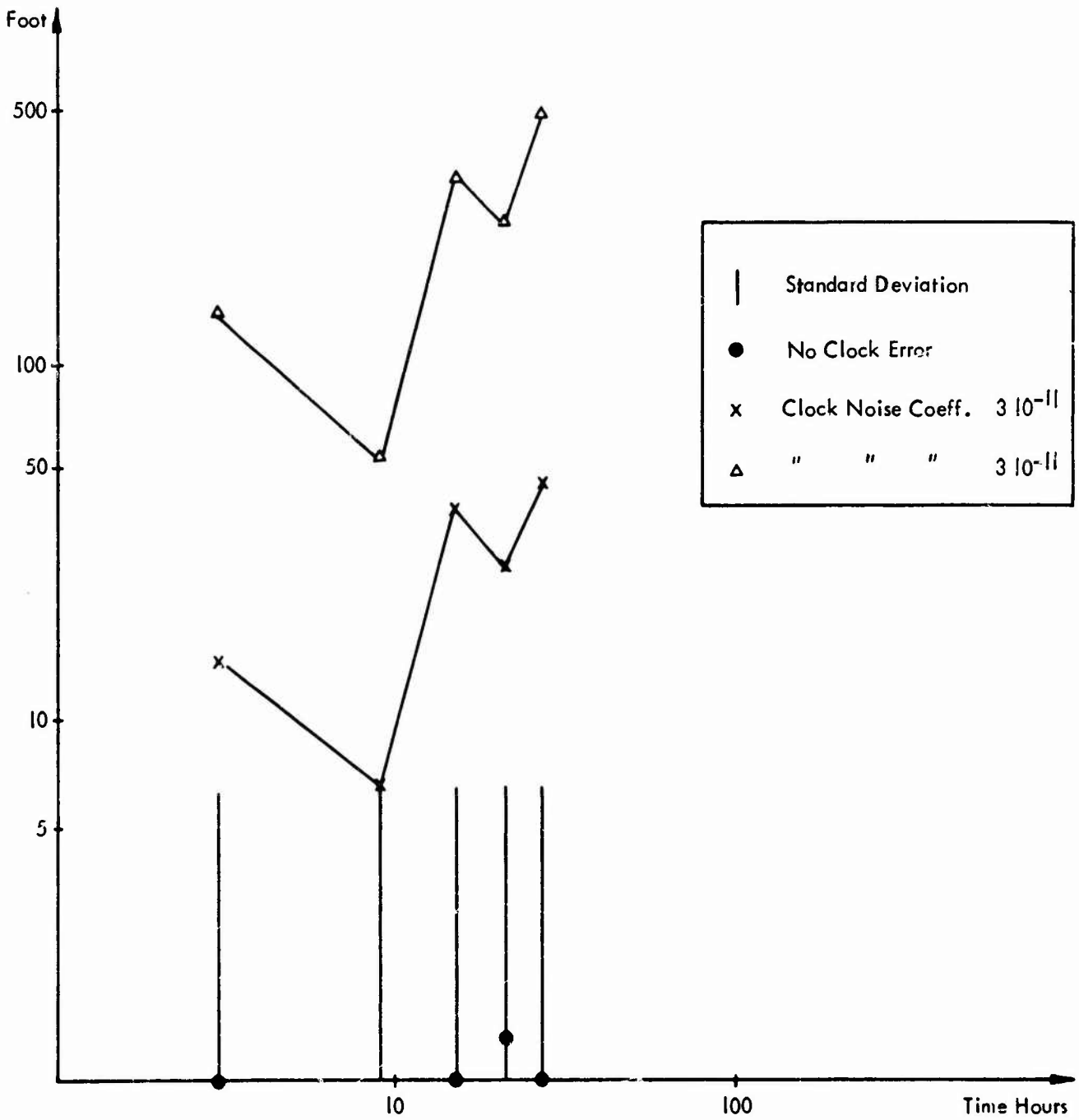


Figure 17 Latitude Error

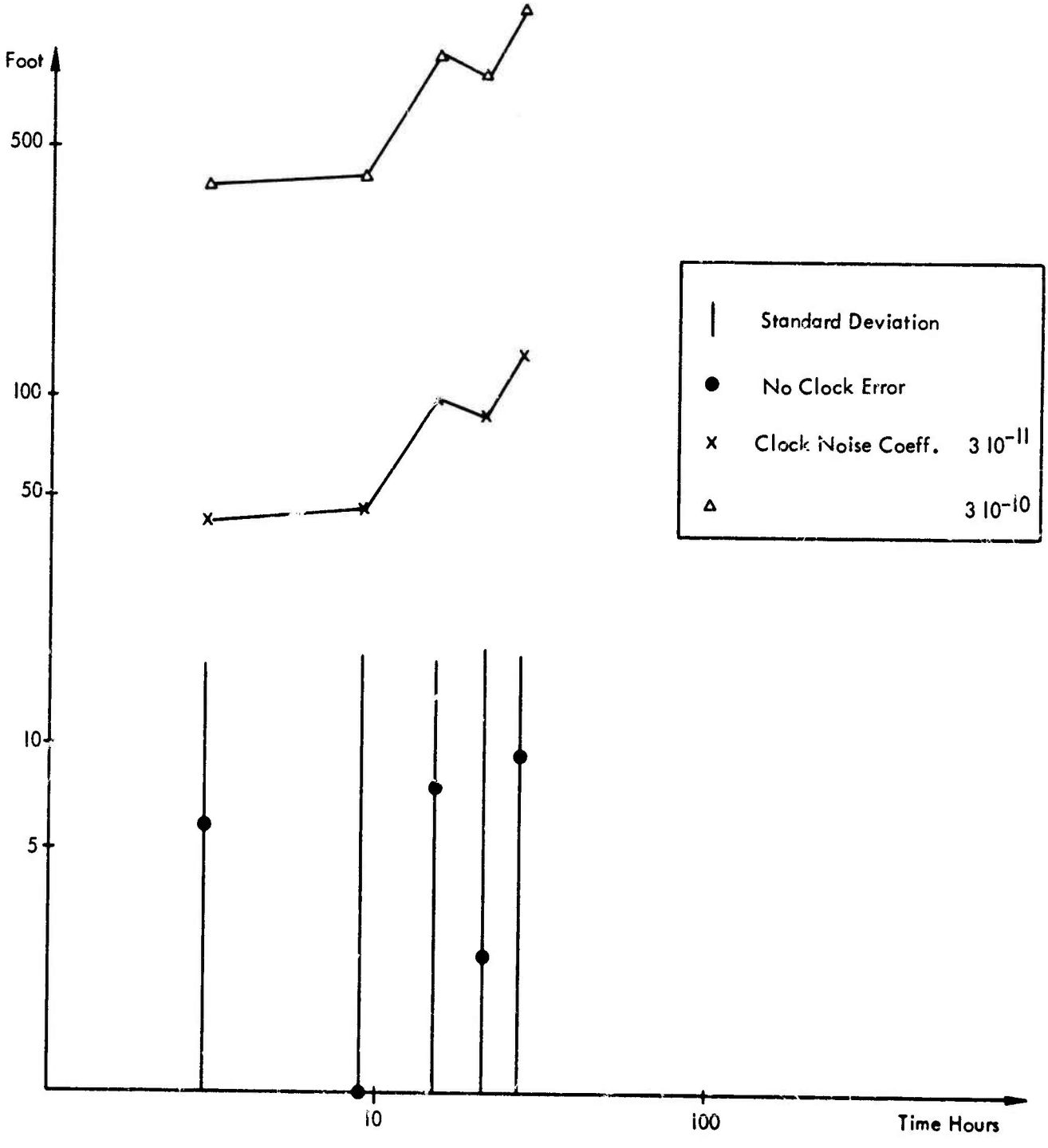


Figure 18 Altitude Error

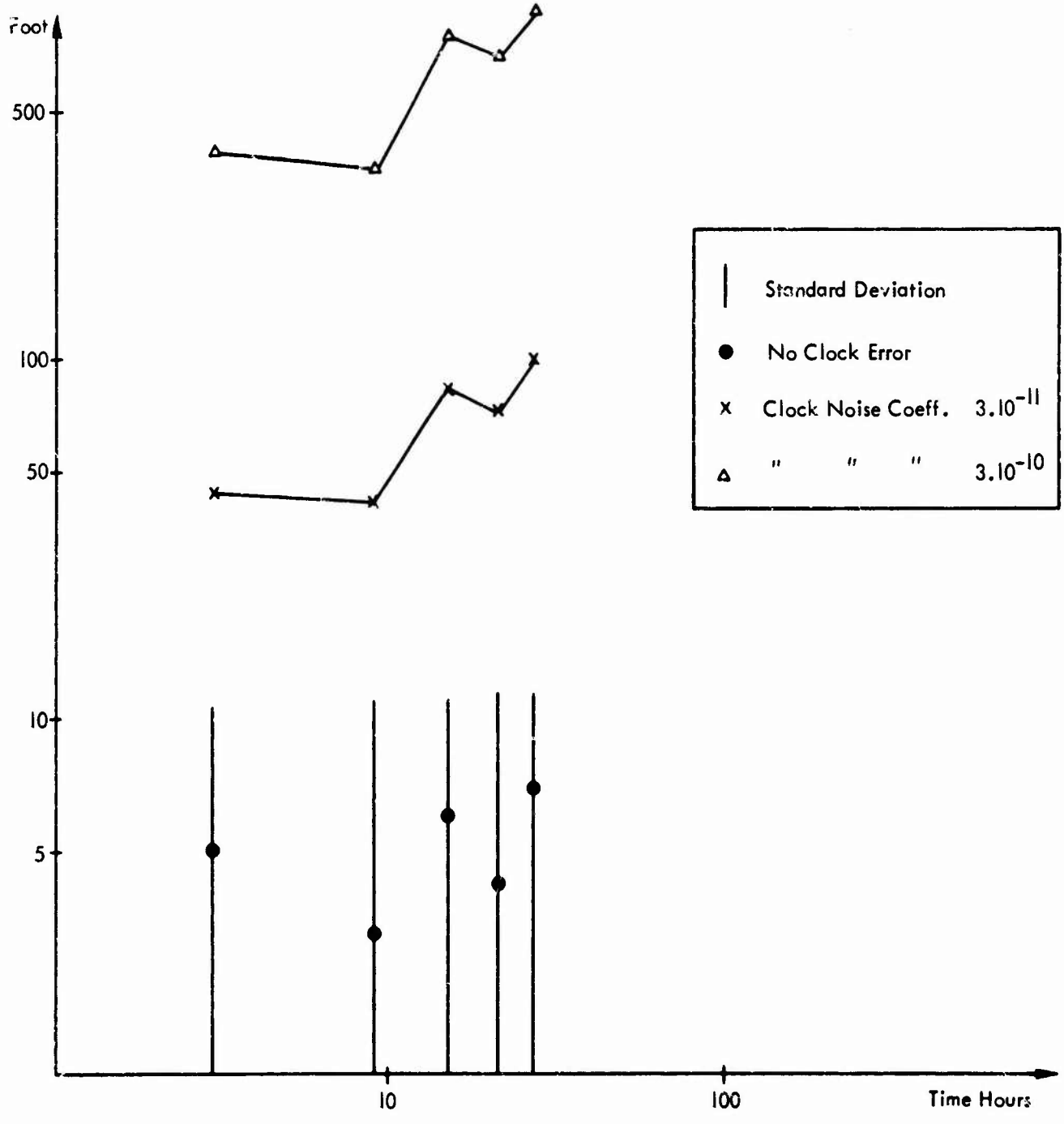


Figure 19 Bias Error

6.0 EXPECTED PERFORMANCE AND CONCLUSIONS

In this final section, we will calculate the expected rms timing uncertainty on error (from the expressions given in Section 3) for an assumed system of a cesium standard for the master station and a rubidium standard aboard the spacecraft.

In order to calculate the mean square error from prediction we must choose a prediction algorithm. In these calculations, it was assumed that the deterministic coefficients of the clock were already determined and that the exponential weighting algorithm was used on the residuals of the data to improve the estimate of rate. This error was given in equation (11) of section 3 and is repeated here.

$$\langle \epsilon^2(\tau) \rangle = \frac{h_o(m+1)}{2m+1}\,\tau + \frac{h_{-1}\tau^2}{m^2(2m+1)} \sum_{k=1}^{\infty} \left(\frac{m}{m+1}\right)^k k^2 \ln(k) \qquad (11)$$

We have chosen a value of m = 0.6 and a prediction interval maximum of 8.64×10^4 seconds (24 hours). Given the two clocks are operating independently, then their variances will add. Likewise, the coefficients h_o and h_{-1} will add linearly in the above expression. Table 4 is the expected rms timing error (or uncertainty) from prediction using equation 6 for a system using a cesium super tube (HP option 004) and a typical rubidium. Values of h_o and h_{-1} were taken from Table 1.

In conclusion, the timing error given in Table 4 represents the results of near optimum handling of the data. The difference between using an optimum filter and the exponential filter is very slight. The important observation is that the clock model as given and the data processing is the best that can be done in terms of prediction in the presence of clock noise. To maintain the timing uncertainty to the rms values given by equation 11 and tabulated for the special case in Table 4, it is necessary to not only evaluate the deterministic coefficients of Equation 11 by least squares fitting to frequency data, but to also weight the residuals and update the model based on these weighted observations.

TABLE 4
SUMMARY OF TIMING ERROR

Prediction Time (seconds) (Since last Calibration)	$\langle \epsilon_{ij}^2(\tau) \rangle^{1/2}$ (ns)
1	0.014
10	0.044
10^2	0.16
10^3	0.92
10^4	8.3
10^5	81.60

where $h_o = 2.5 \times 10^{-22}$

$2\ln(2)h_{-1} = 3.609 \times 10^{-25}$ and h_{-1} and h_o are the contributions of the cessium supertube and the rubidium. h_o(rubidium) = 2×10^{-22}, h_o(cessium) = 5×10^{-23}, $2\ln(2)h_{-1}$(rubidium) = 6×10^{-13}, $2\ln(2)h_{-1}$(cessium) = 3×10^{14}.

Lastly, it was our opinion that precise modeling of the clock environment in space would not prove fruitful for the necessary expenditure of effort. This question, however, still requires further study and some application of the technique of prediction to date obtained in a simulated environment. The most obvious conclusion is that efforts expended to maintain the clocks environment as constant as possible will yield the best performance. The techniques described in this paper have all been developed and used at the National Bureau of Standards in Bolder, Colorado to predict time dispersion in their atomic time scale. Their experience in this regard overshadow the simple excersize performed here to predict time dispersion from laboratory measurements. This effort represented a very small sample of data and was presented only to demonstrate the techniques of prediction and characterization of clock noise. Because of the limited data, the measurements do not necessarily adequately describe the possible performance of the Efratom clock.

The recommendation of this paper is that the techniques presented need further exercise in their application to GPS with data obtained for long periods of time and under simulated environmental conditions.

Appendix - Assumed Conditions for TRACE Simulation

Clock errors were simulated for both stations and vehicles, but not for users. However, a 5 ft sigma range noise was added in all cases. The common configuration for all runs was as follows:

a) Orbit Configuration (Sigma)

2 x 2 system with the following elements for Vehicle #1:

Semi-major axis, a = 87,145,102 ft RT. Ascen., $= 195^{o}$

Eccentricity, e = .0001 Argument of Perigee, = 0.

Inclination, i = 63 m = 41^{o} CPAW = 10^{-9}

The other vehicles differed in mean anomaly and right ascension of node as shown below:

Vehicle #	2	3	4
Rt. Ascen.	195^{o}	75^{o}	75^{o}
Mn. Anom.	81^{o}	64^{o}	124^{o}

b) Orbit perturbations

As defined in Reference (4).

c) Stations and users location

STATION	LATITUDE	LONGITUDE
VTS (Master)	34.83	239.
HUL	21.56	201.76
KOD	57.60	207.82
BOS	42.95	288.37
US1	33.	254.
US2	-33.	344.
US3	33.	74.
US4	-33.	164.
US5	33.	254.

d) Clock model	Bias	Drift	Drift rate (aging)
Station clocks	$T_i(T_o) = 0$	$R_i(T_o) = 0$	$D_i = 0$
Vehicle clocks	$T_j(T_o) = 0$	$R_j(T_o) = +10^{-12}$	$D_j = 0$

(The sign of R was alternated from one vehicle to the next)

Credits: We would like to extend a deep note of gratitude to Mr. David Allan of National Bureau of Standards, Boulder, Colorado for his time and efforts spent in explaining the techniques of clock prediction developed at NBS and presented here for application to GPS.

References:

1) 27th Annual Frequency Synposium, Ft. Monmouth, New Jersey, Performance, Modeling and Simulation of Cesium Beam Clocks, David Allan, et. al, Time and Frequency Division, NBS, Boulder, Colorado.

2) National Bureau of Standards Nomograph, Time and Frequency, Chapters 8 and 9, to be published this spring.

3) 27th Annual Frequency Symposium, Evaluation of a Rubidium Standard for Satellite Application, S. A. Nicols, et al, Naval Research Laboratory, Wash. D.C.

4) Update of Ephemeris Error Analysis Baseline by C. W. Hanson, DNSDP-CWH-036, January 7, 1974.

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

Report No. 4

DIRECT ACQUISITION ANALYSIS

CONTENTS

SECTION		PAGE
1.0	Scope	4-1
2.0	Effect of User Clock Drift	4-1
3.0	Effect of Range and Velocity Estimation Errors	4-3
4.0	Total Time and Frequency Uncertainty	4-3
5.0	Acquisition Time	4-4
6.0	Example Case	4-7
7.0	Conclusions	4-11

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Effect of Initial Frequency Offset, δ and Drift Rate α	4-8
2	Effect of Range and Velocity Estimate Error ΔR, ΔV	4-9
3	Effect of Dwell Time and Acquisition Bandwidth, T_D and B	4-10

DIRECT ACQUISITION ANALYSIS

1.0 SCOPE

This trade study deals with the user who has a stable clock which is synchronized to the satellite clock in a benign environment. Then, the user turns off his equipment (to conserve power) at t = 0, except for the clock. At a later time, t, he then turns on his equipment in a hostile (jamming) environment and reacquires the P-signal.

The factors which affect the time to reacquire the P-signal are:

a. accuracy with which the user clock is synchronized to the satellite clock at t = 0.

b. unknown drift of the user clock over the period t.

c. the accuracy with which the user can calculate the range and velocity to the satellite prior to reacquisition at time t, using data contained in his computer.

Section 2 analyzes the effect of user clock drift. Section 3 analyzes the effect of errors in satellite-user range and velocity prediction. Section 4 discusses time and frequency uncertainties while Section 5 relates these uncertainties to acquisition time. Section 6 provides an example and Section 7 presents some conclusions.

2.0 EFFECT OF USER CLOCK DRIFT

The following assumptions are made:

a. All satellite clocks are phase synchronized with each other and have zero drift rate.

b. User acceleration is zero during the initial synchronization and reacquisition periods.

c. The C/A signal cannot be used during the reacquisition period due to jamming.

The difference between the user clock frequency and the satellite clock frequency at time t, Δf_1, is:

$$\Delta f_1 = (f_o + \delta + at) - f_o \quad (1)$$

where

f_o = satellite clock frequency - Hz

δ = error in setting frequency of user clock at t = 0 - Hz

a = frequency drift rate of user clock - Hz/sec.

The number of radians, θ_1, the user clock gains or loses relative to the satellite clock is found by integrationg (1).

$$\theta_1 = 2\pi \int_o^t f \, dt - 2\pi \int_o^t f_o \, dt$$

$$= \theta_o + 2\pi \delta t + \pi a t^2 \quad (2)$$

where

θ_o is the phase difference error between satellite and user clocks at t = 0.

To make Δf_1 and θ_1 small, δ and θ_o must be made small. Therefore the user must accurately calculate the range and velocity to the satellite at t = 0 and transpose the received phase and frequency to that of the satellite. To do this, the user must make an accurate position determination using four satellites. Thus δ and θ_o are functions of the overall navigation system accuracy.

The oscillator drift rate, a, is minimized by use of pre-aged crystals which are temperature compensated or placed in an oven. The user must carry in his computer, either a simplified set of orbital equations for the entire satellite network or, alternatively, a set of look-up tables giving satellite-user range, range-rate to an arbitrary user location, plus some simple equations for extrapolation. If the user has moved a significant distance prior to acquisition, he must also estimate his new position relative to his position at t = 0.

At time t, the user receives a clock signal which was transmitted from the satellite at $t - \tau$ where

$\tau = \frac{R}{C}$ seconds

R = satellite-user range at $t - \tau$

C = speed of light

The satellite clock signal received at the user is shifted by an angle of $2\pi f_o \frac{R}{C}$ radians.

3.0 EFFECT OF RANGE AND VELOCITY ESTIMATION ERRORS

To make the satellite-user clock phase error as small as possible, the user estimates the satellite-user range R at t - τ. If the user software predicts R(t), then R(t - τ) can be derived as follows:

$$R(t) = R(t - \tau) + V\tau$$

$$\tau = \frac{R(t - \tau)}{C}$$

where V = the satellite-user velocity (range-rate). Combining the above,

$$R(t - \tau) = \frac{R(t)}{1 + \frac{V}{C}} \tag{3}$$

Thus the user must also estimate the range-rate or velocity, V. Let the difference between the estimates value of R and the true value of R be ΔR.

The phase error introduced by uncertainty in range prediction, θ_2, is

$$\theta_2 = 2\pi (f_o + \delta) \left(\frac{R + \Delta R}{C} \right) - 2\pi f_o \frac{R}{C}$$

$$= 2\pi f_o \frac{\Delta R}{C} + 2\pi \delta \left(\frac{R + \Delta R}{C} \right) \tag{4}$$

Notice that the user thinks the satellite frequency is $f_o + \delta$ and uses this to calculate the phase shift.

The frequency (doppler) error introduced by uncertainty in velocity prediction, Δf_2, is obtained by differentiating (4).

$$\Delta f_2 = \frac{1}{2\pi} \frac{d\theta_2}{dt}$$

$$= f_o \frac{\Delta V}{C} + \delta \frac{V + \Delta V}{C} \tag{5}$$

4.0 TOTAL TIME AND FREQUENCY UNCERTAINTY

The total phase error, θ, between satellite and user clock at time t is obtained from equation (2) and (4). However, all of the error terms are uncorrelated with each other (except for the last term in (4) which turns out to be much smaller than the others and can be

neglected). That is, θ_o, δ, a, and ΔR are considered as independent random variables with Gaussian distributions. Then

$$\theta_{RMS} = \sqrt{\theta_o^2 + (2\pi\ \delta t)^2 + (\pi\ at^2)^2 + (2\pi\ f_o\ \frac{\Delta R}{C})^2} \tag{6}$$

where the second term in equation (4) is dropped. (See Appendix A.)

Similarly, from (1) and (5) we get the total difference between satellite and user clock frequencies, Δf.

$$\Delta f_{RMS} = \sqrt{\delta^2 + (at)^2 + (f_o\ \frac{\Delta V}{C})^2} \tag{7}$$

where the second term in (5) is dropped. (See Appendix A.)

To calculate P-signal acquisition time, it is convenient to express (6) as a time error, ΔT, using the relations $\theta = 2\pi f_o\ \Delta T$, and $\theta_o = 2\pi\ f_o\ \Delta T_o$.

$$\Delta T_{RMS} = \sqrt{(\Delta T_c)^2 + \left(\frac{\delta}{f_o}\right)^2 t^2 + \frac{1}{4}\ \left(\frac{a}{f_o}\right)^2 t^4 + \left(\frac{\Delta R}{C}\right)^2} \tag{8}$$

Finally, rewrite (7)

$$\frac{\Delta f_{RMS}}{f_o} = \sqrt{\left(\frac{\delta}{f_o}\right)^2 + \left(\frac{a}{f_o}\right)^2 t^2 + \left(\frac{\Delta V}{C}\right)^2} \tag{9}$$

Equations (8) and (9) define the time and frequency uncertainty existing at the user at time t. The time to directly reacquire the P-signal is a function of these parameters.

5.0 ACQUISITION TIME

The acquisition time, T_A, to acquire the P-signal, is a function of $\Delta T'$ and $\Delta f'$ as follows:

$$T_A = T_D\ \left(\Delta T'\right)\ f_c\ \ \text{INT}\left[\left(\frac{\Delta f'}{f_o}\right)\ \left(\frac{f_{RB}}{B}\right)\right] \tag{10}$$

where

$\Delta T'$ = total time interval searched - sec

T_D = dwell time per code chip - sec

f_c = code chipping rate - Hz

$(\Delta T')f_c$ = number of code chips searched

B = acquisition bandwidth of receiver - Hz

f_{RF} = carrier frequency - Hz

$\left(\frac{\Delta f'}{f_o}\right) f_{RF}$ = total carrier frequency searched - Hz

$\mathrm{INT}\left[\left(\frac{\Delta f'}{f_o}\right)\left(\frac{f_{RF}}{B}\right)\right]$ = number of frequency bands searched

where INT $[\cdot]$ is the smallest integer larger than the quantity in the brackets.

Equation (10) is based on a sequential receiver which can detect and evaluate only one chip position in one frequency band at a time. Parallel frequency channels (i.e., filter bank) could be implemented to speed up acquisition time considerably. (Ref. - Appendix H of MRL Signal Definition Study.)

Equation (10) gives the acquisition time as a function of total time and frequency intervals searched while equations (8) and (9) give RMS values of ΔT and Δf. The optimum search procedure is to start at the most probable code chip position and frequency band, and then test positions in a two-dimensional spiral shaped time-frequency search pattern which progressively moves out to less probable time-frequency positions. Assuming that ΔT and Δf are statistically independent (worst-case since equations (8) and (9) have common error sources), we then can attach a probability of acquisition with T_A as a function of the RMS values of ΔT and Δf given by (8) and (9) as shown in Table 1.

TABLE 1
PROBABILITY OF ACQUISITION

$\frac{\Delta T'}{\Delta T \text{ RMS}}$	$\frac{\Delta f'}{\Delta f \text{ RMS}}$	JOINT PROBABILITY OF ACQUISITION
2.0	2.0	0.47
4.0	4.0	0.91
6.0	6.0	0.995

The maximum acquisition time was based on searching all time-frequency cells within $\Delta T'$ and $\Delta f'$. The probability of acquisition was derived by calculating the following joint probability.

$$\text{Prob acquisition} = \text{Prob}\left[|\Delta T| < |\Delta T'|\right] \times \text{Prob}\left[|\Delta f| < |\Delta f|\right] \qquad (11)$$

where

ΔT and Δf in equation (11) are assumed to be Gaussian random variables with zero mean and RMS values given by equations (8) and (9). The ratio, K, of $\Delta T'$ and $\Delta f'$ to ΔT_{rms} and Δf_{rms} yields the values for the probability on the right hand side of equation (11).

Substituting $\Delta T' = K\Delta T_{rms}$ and $\Delta f' = K\Delta f_{rms}$ into (10)

$$T_A = T_D\left(K\Delta T_{rms}\right) f_c \ \text{INT}\left[\left(\frac{K\Delta f_{rms}}{f_o}\right)\left(\frac{f_{RF}}{B}\right)\right] \qquad (12)$$

For t small, and $\Delta T_o \ll \frac{\Delta R}{C}$, and $\frac{\delta}{f_o} \ll \frac{\Delta V}{C}$, we can write equations (8) and (9):

$$\Delta T_{rms} \approx \frac{\Delta R}{C}$$

$$\Delta f_{rms} \approx \frac{\Delta V}{C}$$

Substituting in equation (12):

$$T_A = T_D \ K\left(\frac{\Delta R}{C}\right) f_c \ \text{INT}\left[\left(\frac{K\Delta V}{c}\right)\left(\frac{f_{RF}}{B}\right)\right] \qquad (13)$$

Since ΔT and Δf are partially correlated, it might be possible to reduce the number of time-frequency positions in the acquisition search, thereby reducing the acquisition time.

6.0 EXAMPLE CASES

Let

$\Delta T_o = 10^{-8}$ sec.

$C = 10^9$ ft/sec.

$f_c = 10^7$ chips/sec.

$f_{RF} = 1.6 \times 10^9$ Hz

$K = 4$

Using equations (8) (9) and (12) plots of signal acquisition time, T_A, versus time from initial user clock synchronization were computed and are shown in Figures 1, 2 and 3. (Note that T_A is the signal acquisition time, not the time to first navigation fix, which is longer than T_A.)

Figure 1 shows the effect of user clock offset error, δ, at $t = 0$; and clock frequency drift rate error, a. Figure 2 shows the effect of satellite-to-user range and velocity estimate errors, ΔR and ΔV, by the user at time = t. Figure 3 shows the effect of dwell time, T_D, and acquisition bandwidth, B.

When acquiring in a jammed environment, it is necessary to make T_D long and B small at the expense of making the acquisition time longer. The relationship between these parameters and reliable acquisition in a jamming environment is complex and will not be investigated under this task.

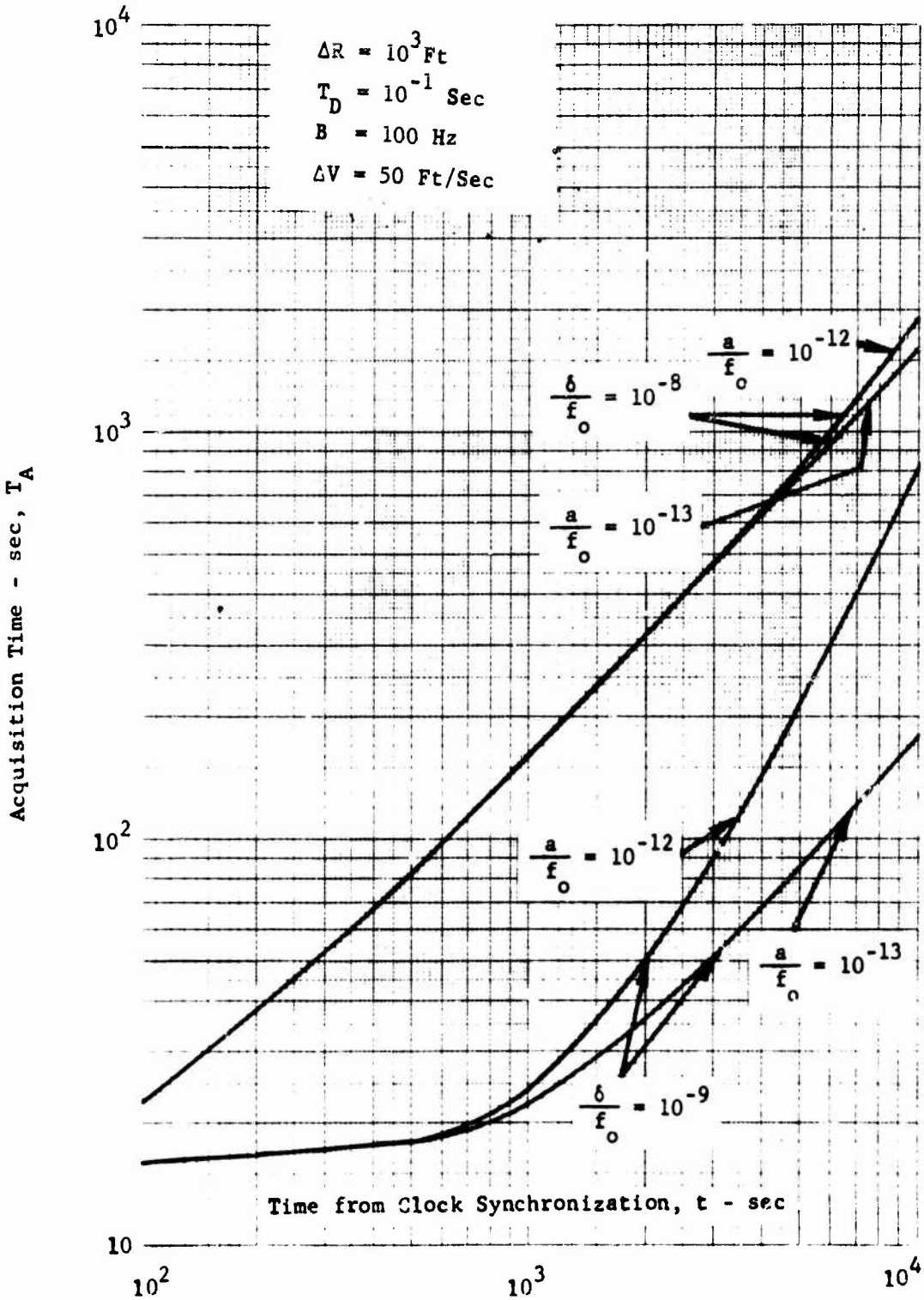


Figure 1 Effect of Initial Frequency Offset, δ, and Drift Rate, a

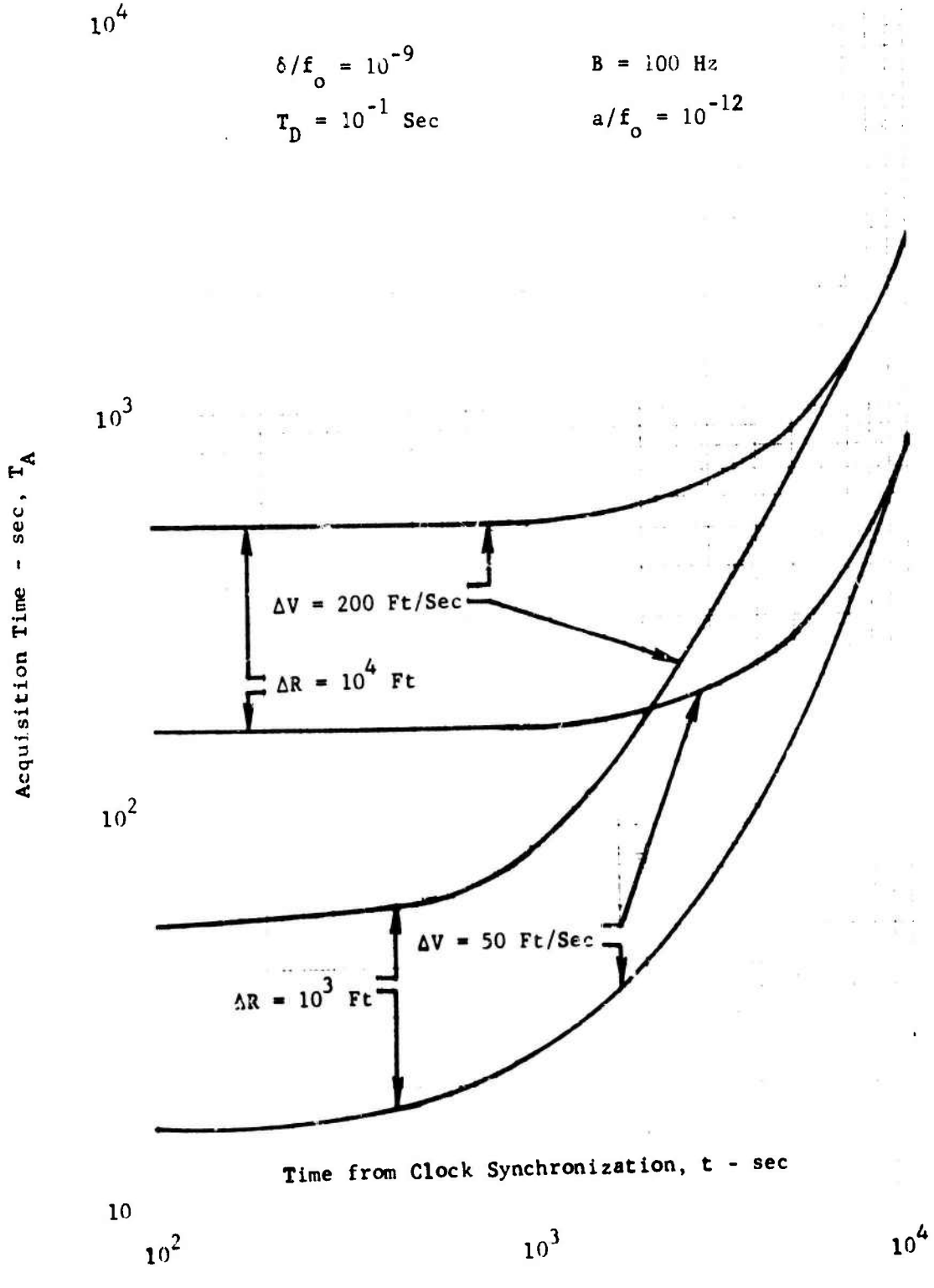


Figure 2 Effect of Range and Velocity Estimate Errors ΔR, ΔV

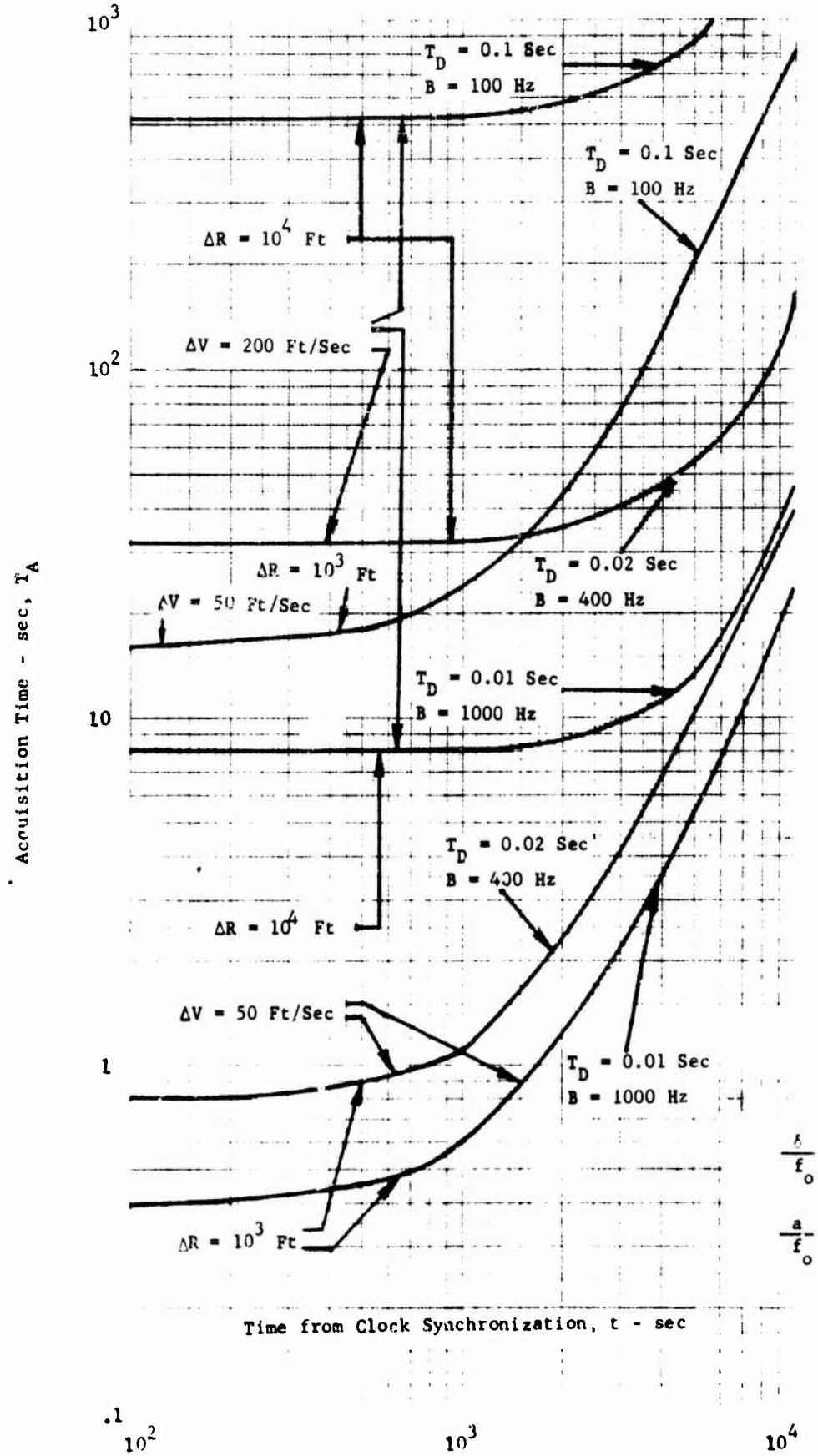


Figure 3 Effect of Dwell Time and Acquisition Bandwidth, T_D and B

7.0 CONCLUSIONS

Direct acquisition of the P-signal without the aid of the C/A signal is possible. Reasonable signal acquisition times (less than 300 seconds) are attainable provided the user has an accurate knowledge of the phase and frequency of its clock relative to the satellite clock, and a good estimate of the satellite-to-user range and range-rate. To do this, it is necessary for the user to synchronize his clock to the navigation satellite system at some time prior to operational use. A 3-hour interval appears to be a maximum for time between synchronization and P-code direct acquisition for most cases of interest.

To be more specific, direct acquisition of the P-signal can be attained within 200 seconds (with probability of 0.9) after an interruption of receiver signal reception of 10^4 seconds (2.8 hours) provided (see Figure 1 in Main text):

a. Initial user clock phase synchronization to within 10 ns (rms).

b. Initial user clock frequency synchronization to within 10^{-9} (rms).

c. User clock frequency drift rate predicted to within 10^{-13} per sec (10^{-8} per day) (rms).

d. User-to-satellite range estimated to accuracy of 1000 ft (rms) at time of direct acquisition.

e. User-to-satellite range-rate estimated to accuracy of 50 ft/sec (rms) at time of direct acquisition.

f. User receiver dwell time of 0.1 sec/code chip during code search.

g. User receiver acquisition bandwidth of 100 Hz.

With the parameters for dwell time and acquisition bandwidth specified above, it should be possible to acquire the P-signal with a jamming-to-signal ratio in the order of 45 dB. Further work is required to determine acquisition properties for specific user equipments in jamming environments. It is planned that almanac data will be periodically transmitted to the user and stored in the user computer to aid in satellite acquisition. The accuracy of this almanac should be studied to determine what values of \R and \V are feasible.

Appendix A

In equation (4), the ratio of the second term to the first term is:

$$\frac{\delta(R + \Delta R)}{f_o(\Delta R)} = \left(\frac{\delta}{f_o}\right)\left(\frac{1 + \frac{\Delta R}{R}}{\frac{\Delta R}{R}}\right)$$

Typically, $\frac{\delta}{f_o} \leq 10^{-9}$ and $\frac{\Delta R}{R} > 10^{-6}$, which makes the above ratio $<10^{-3}$. Therefore, the second term in equation (4) can be neglected.

In equation (5), the ratio of the second term to the first term is

$$\frac{\delta(V + \Delta V)}{f_o(\Delta V)} = \left(\frac{\delta}{f_o}\right)\left(\frac{1 + \frac{\Delta V}{V}}{\frac{\Delta V}{V}}\right)$$

Typically $\frac{\delta}{f_o} \leq 10^{-9}$ and $\frac{\Delta V}{V} > 10^{-3}$, which makes the above ratio $< 10^{-6}$. Therefore the second term in equation (5) can be neglected.

Report No. 5

IONOSPHERIC MODEL ANALYSIS

CONTENTS

SECTION		PAGE
1.0	Introduction	5-1
2.0	Statement of the Problem	5-1
3.0	User and Control Segment Requirements	5-1
	3.1 Users	5-2
	3.2 Control Segment	5-2
4.0	Brief Summary - Ionosphere Range Error	5-3
5.0	Brief Technical Background	5-3
6.0	Ionosphere Range Error	5-6
7.0	Determination of Ionosphere Range Error	5-13
8.0	The Problem of Ionosphere Prediction	5-14
	8.1 Magnitude of the Range Error	5-14
	8.2 Prediction Models	5-15
	8.3 Raw Data for Model Update	5-15
	8.4 Mathematical Prediction Models	5-19
	8.5 Complete Stanford Model	5-20
	8.6 Appropriate Trigonometric Model	5-22
9.0	A Two-Frequency Correction Technique	5-24
	9.1 The Basic Algorithm	5-24
	9.2 Non-Simultaneous Measurements	5-27
10.0	L_1/L_2 Range Error Statistics	5-28
	10.1 Statement of the Problem	5-28
	10.2 Conclusions	5-29
	10.3 Derivation of the Range Error Equation	5-30
	10.4 First Method, Indirect Solution	5-33
	10.5 Second Method, Indirect Solution	5-34
	10.6 The RMS Range Error Equation	5-34
	10.7 TRW Values for RMS Range Error; Empirical Approximation	5-34
11.0	Optimum Power Division Between L_1 and L_2 Signals	5-36
	11.1 Statement of the Problem	5-36
	11.2 Summary	5-36
	11.3 Derivation of Optimum L_1 and L_2 Power Ratio	5-37

CONTENTS (Cont.)

SECTION		PAGE
12.0	Pulse Distortion Due to the Ionosphere	5-44
13.0	Recommendations for Phase I	5-46
14.0	Selected Bibliography	5-47

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Ionosphere Electron Content Curves During One Day Longitude 291°	5-8
2	Ionosphere Electron Content Curves During One Day Longitude 111°	5-9
3	Ionosphere Electron Content For Two Days Two Geographical Locations	5-10
4	Ionosphere Electron Content For Two Days Single Location	5-11
5	Relative Range Error for Non-Optimum Power Ratios	5-41
6	Distorted Pulse Waveforms	5-42
7	Autocorrelator Output Functions	5-43

TABLES

TABLE		PAGE
1	Range Error Uncertainty Due to Scale Height Uncertainty	5-13
2	Percent Ionosphere Coverage vs Elevation Angle	5-18
3	RMS Range Error vs C/No	5-34
4	Relative Range Error for Non-Optimum Power Ratios	5-40

1.0 INTRODUCTION

This document summarizes the work performed to date by WDL-SDD concerning ionospheric and tropospheric errors in the proposed Global Positioning System (GPS).

2.0 STATEMENT OF THE PROBLEM

The GPS is a high-precision navigation system which permits users to receive coded signals from several earth-orbiting satellites, and through a suitable processing of modulation information contained in the several signals, determine his position and velocity, in three dimensions. The baseline system currently being studied will utilize RF carriers in the lower microwave region, ie, nominally 1575 MHz and 1230 MHz.

The received signals, upon processing, permit the user to determine the apparent range, or pseudo-range, to each of several (viz, four) satellites. Upon referring the pseudo-range measurements to a common time, and upon determining the satellite position at that time (using data transmitted from satellites) the user computes his position.

The problem of computing an accurate "fix" is compounded by certain atmospheric radio propagation anomalies, known as the "tropospheric range error" and the "ionospheric range error". The former is largely frequency-independent, and normally amounts to 2-50 feet; this error is determined entirely by meteorological conditions in the near vicinity of the GPS user. The latter error can range between 2-500 feet; this error is strongly frequency dependent, and is also a function of the radio path elevation angle and the ionospheric condition along the user-satellite path.

Users requiring the most precise fix capable of the GPS must make corrections for both tropospheric and ionospheric range errors. Various methods have been proposed and studied; the end objective of all these methods is either to compute from a prediction model, or to make measurements on or related to, the pertinent errors. The tropospheric range error is discussed in a separate document (Error Budget Analysis); this report is concerned only with the ionosphere.

3.0 USER AND CONTROL SEGMENT REQUIREMENTS

A radio ranging system, such as is planned for GPS users, performs an indirect measurement to determine range between user and satellite. The measurement consists of two steps, ie:

- Time delay between user and satellite is actually measured.
- The time delay is multiplied by a number which represents the velocity of light along the ray; the resultant product is interpreted as the range (distance) to the satellite.

Such an indirect technique is far more convenient for GPS than a direct measurement, and since it is possible to measure relative time delays with extremely high precision, the indirect technique will yield an accurate range measurement if, and only if, the average group velocity of the radio signal along the ray is accurately known. Unfortunately the ionosphere forces upon us a range correction which can be less than one foot (vertical path, low Iv) to as much as 400 feet or more (low elevation angle, high Iv); it is assumed here the RF frequency is in the vicinity of 1.2-1.6 GHz.

3.1 Users

User fix accuracy requirements may be divided into several categories; for the purpose of assessing atmospheric range errors these are conveniently divided as follows:

1. Low Accuracy User. These are users to whom a fix error of several tens of feet, or even 200-400 feet, is of no consequence. Users of this category can ignore the atmospheric effects.

2. Medium Accuracy Users. This category contains users to whom errors of 100-400 feet are unacceptable, but also to whom correcting atmospheric errors any closer than 50-60 feet may very well not warrant the additional equipment cost and complexity.

3. High Accuracy Users. These users possess fix accuracy requirements which force them to correct the atmospheric effects, even under worst-case conditions, so as to obtain a residual error not exceeding a few feet, certainly less than ten feet.

To users in category 1 the atmospheric effects are of no consequence. To those of the second category some degree of both tropospheric and ionospheric correction is required, but simplicity of the correction technique is also of paramount importance. Users of the third category require the best-available correction technique; to these users accuracy of the fix is weighed more heavily than equipment simplicity, although the latter is certainly not ignorable.

3.2 Control Segment

A major function of the GPS Control Segment is to perform high-precision range measurements on the satellites, thereby obtaining raw data for computation and prediction-ahead, of the orbits. It is a system design objective that monitor stations shall perform the range measurements with a residual error as close to one foot as can reasonably be achieved. Therefore, the control segment pseudo-range measurements necessarily will use the best-available techniques to correct for tropospheric and ionospheric range errors, on a realtime basis.

4.0 BRIEF SUMMARY - IONOSPHERIC RANGE ERROR

a. The L-band range error varies markedly with time and signal path elevation angle, from a few feet to as much as several hundred feet.

b. The range error can be computed approximately for a given path, if one is provided with a prediction model for both the vertical integrated electron content and for the effective scale height of the ionosphere.

c. Present technology does not provide an accurate prediction model, even for a limited geographical region.

d. A two-frequency technique is available, permitting users to make an almost perfect correction for L-band range errors. High-precision users should use this technique until such time that a practical, accurate prediction model can be formulated.

e. During Phase I of its operation, the GPS control segment is ideally suited to gather raw range error data, and to employ it for evaluating prediction models of varying complexities.

5.0 BRIEF TECHNICAL BACKGROUND

This memorandum is concerned with errors arising from radio propagation anomalies in the earth's ionosphere.

The ionospheric is a high-altitude region of the earth's atmosphere, lying between roughly 200 and 2000 miles altitude. In this region a fraction of the atmospheric gas molecules is ionized by certain ultraviolet radiation wavelengths from the sun, thus causing the ionosphere to become a plasma.

By solving Maxwell's equations of electromagnetic theory in a region of space containing a plasma, it is possible to demonstrate that the group and phase velocities of a radio wave are altered from the free-space value, due to the action of the free electrons comprising the plasma; the group velocity is lowered below the free-space value. It is also possible to demonstrate that the group-velocity reduction factor is an extremely complicated function of (1) the degree of ionization, (2), the particular frequency of the propagating wave, and (3) the magnitude and direction of earth's magnetic field at the ionospheric point in question.

The general effect of the ionosphere on signals above 1000 MHz is to cause an additional group time delay, due to wave slowing, thereby introducing a pseudo-range error of very nearly one foot additional range for every nanosecond of wave slowing. The dominant ionospheric range error term, in the microwave region, is inversely proportional to the square of the radio frequency; a more exact analysis reveals additional terms varying as reciprocal frequency cubed, and reciprocal

frequency fourth-power, but these terms can be shown to contribute a range error never exceeding a very few centimeters, ie, well under one foot.

For the L_1 and L_2 GPS frequencies, the ionospheric range error is seldom less than 2-3 feet (vertical path, pre-dawn, very light ionization), or more than 300 feet (low elevation angle, close to noon, intense ionization).

If the path elevation angle is not less than about 5 or 6 degrees, and if the RF frequency exceeds 1000 MHz, tropospheric and ionospheric range error effects may be treated separately, ie, coupling between the two is negligible. On the other hand a coupling error, which can amount to several feet, appears at extremely low angles, ie, very close to local horizontal.

The analytical work indicates that by far the most important ionospheric effect, pertinent to GPS ranging measurements, is that the group velocity of a microwave radio signal, passing through the ionosphere, is less than the velocity of light in a vacuum. The velocity discrepancy is a complicated function of the detailed degree of ionization along the path between a given satellite and a user; the range error resulting from this velocity discrepancy is a function of the length of the path through the ionosphere, and hence of the elevation angle of that path, and, to an excellent approximation, the range error varies as $1/f^2$. Other types of ionospheric errors have indeed been identified and sized; the group velocity effect is the only one examined which appears to produce range errors as large as one foot, (for frequencies exceeding 1 GHz.

For example, W. Waters of WDL, and A. da Rosa at Stanford University have independently computed the range error resulting from the fact that, because of ionospheric refraction, the user-satellite path is actually curvilinear, not straight; the geometrical length of a curved path between two fixed points (the user and the satellite) is obviously greater than the length of a straight line between the same two points. Both approximated the ionosphere as a uniform slab of thickness d, center height h above earth's surface, containing an assumed integrated electron content Iv vertically. (See discussion below). Da Rosa performed numerical integration of the radio ray equation, using several frequencies (100-1500 MHz), whereas Waters derived an analytical approximation from the basic ray equation in the refracting ionosphere.

Both found the path-lengthening range error to amount to only a few centimeters at 1200 MHz; da Rosa's numerically-computed points accurately fit an empirical relation which shows range error varying as $f^{-3.74}$; Waters' analytic expression varys as f^{-4}, and shows this range error varying as the square of the integrated election content. Finally, in a private communication to Waters, Dr. Rodney Bent indicated that he found a path-lengthening error of only about 2 cm, at the L_1 frequency. Consequently, no further discussion of this error is given in this document.

Waters' analytical expression is as follows:

$$\Delta R/\text{ray bending} = \frac{2.68*10^{-4}\, I_v^2\, (1 + d/2H)}{df^4} \left(\frac{\cot^2 E}{\sin E}\right) \text{feet}, \qquad (1)$$

with

I_v = vertical electron content (defined below), hexams
d = ionosphere slab thickness, Km
N = ionosphere effective height, Km
f = RF frequency, GH_z
E = elevation angle, degrees (at least 10°)

Let us now pause to introduce the concept of integrated electron content. Suppose a radio link is established between a satellite and a user; the radio path is almost a straight line. Now imagine a long cylindrical column, having a cross-sectional area of one square meter (ie, a cylinder about 44 1/2 inches diameter), extended all the way from user to satellite. Move along the cylinder, from user to satellite, and with a suitable instrument measure the total number of free electrons contained within the cylinder. These free electrons are the result of ionization of the earth's atmospheric molecules by certain ultraviolet radiation wavelengths from the sun.

The total count of free electrons in such a one-square-meter column is known as the integrated electron content, and is designated by the letter I. Values for I are usually referred to a vertical column, and are therefore symbolized I_v. Numerical values for I_v, pertaining to a particular place on the earth's surface, can vary over almost a 1000 to 1 range. Low values of Iv are obtained shortly after midnight of a day during which the characteristic solar radiation was unusually low; such a value is of the order of 10^{15} free electrons in a one-square-meter column. High values of I_v are obtained at, or soon after, noon, with the sun almost directly overhead, and with a relatively active sun; such a high value can exceed 10^{18} free electrons per square meter, with an extreme value less than 10^{19}.

It is convenient to express the vertical integrated electron content, I_v, in units of 10^{16} electrons per square meter. Such a measure is known as a HEXAM, abbreviated hex. Thus, a relatively ionized ionospheric column is equivalent to Iv equal to a few tenths of a

HEXAM, whereas a heavily ionized column contains several hundred HEXAMs.

The value of I_v over a particular place varies with:

1. The location of the user (or monitor) station
2. The local time of day (ie, position of the sun in the sky)
3. The time of year
4. The solar activity (related to, but not the same as, sunspot activity

6.0 IONOSPHERIC RANGE ERROR

The ionospheric range error is considerably more complicated to determine in the GPS than is the tropospheric error, for several reasons:

1. The error is very much a function of carrier frequency. As was mentioned in section 4, for frequencies above 1 GHz it can be shown that this error varies inversely as frequency-squared, if smaller terms which contribute a few centimeters or less are omitted.
2. At any given time the structure of the ionosphere is complicated function of position over the earth's surface.
3. The detailed structure of the ionosphere varies with time. Of primary importance are a 24-hour general trend, a 365-day trend, and a (generally) 11-year trend as solar activity varies cyclically. However, short-term variations representing significant departures from the trends frequently appear, and are far from ignorable in the GPS.
4. The ionospheric range error itself may be several times larger than the tropospheric error.
5. The ionospheric error is the resultant of solar radiation in a high-altitude layer of the ionosphere. For this reason measurements made at, or from, earth-surface locations do not provide adequate information for computing the range error of a signal completely penetrating the ionosphere.

A straightforward analysis reveals that the range error for signals in the 1200-1600 MHz range is roughly one foot for every hex of electron content; thus, a 30-hex path yields roughly a 30-foot range error (depending upon the specific frequency).

In order to illustrate the ionospheric complexity, Figures 1 through 4 show typical electron content variations. Figures 1 and 2 illustrate the 24-hour variation; in each figure a set of curves is shown for the

electron content of a vertical path, for various magnetic latitudes, and for two different geolongitudes (110° west and 291° west), both for the same day. Figures 3 and 4 illustrate how the vertical electron content varies with solar activity; curves are given for two different days, for three specific locations, in each case through a 22-hour period.

The electron content along a path other than vertical is greater than along the vertical, simply because the path is longer. It can be shown that, for very low elevation angles, the electron content is as much as three times the vertical value. (For example, compare values of range error ΔR, computed from Eq (2), for $E = 0°$ and $90°$.)

Figure 1 shows a minimum content of about 1 hex (0500 hours, -30° magnetic latitude), whereas Figure 4 shows a maximum of about 75 hex). If a GPS pseudo-range is measured along a vertical path, at the time and location corresponding to the minimum of Figure 1, the ionospheric range error for that path would be roughly one foot. On the other hand, corresponding to the peak of Figure 4, the range error for a low elevation angle, below about 5°, would be roughly 225 feet.

Although the solid curve of Figure 4 is labelled "high solar activity", this curve by no means represents an extreme. Vertical contents of 160 hex have been measured on many occasions over Hawaii, and a value of 240 hex has been measured occasionally. Such an extreme value implies an L-band range error of over 700 feet for low elevation angles.

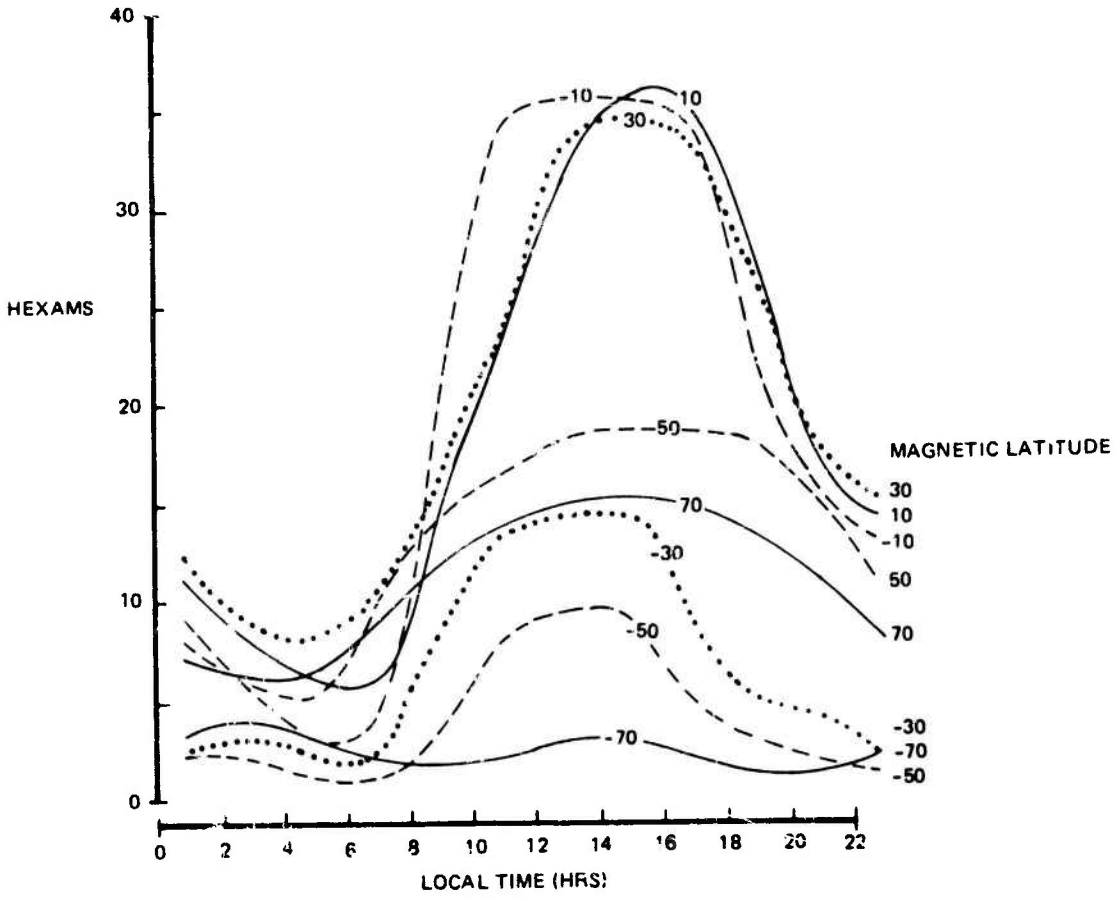


Figure 1 Ionosphere Electron Content Curves During One Day Longitude 291°

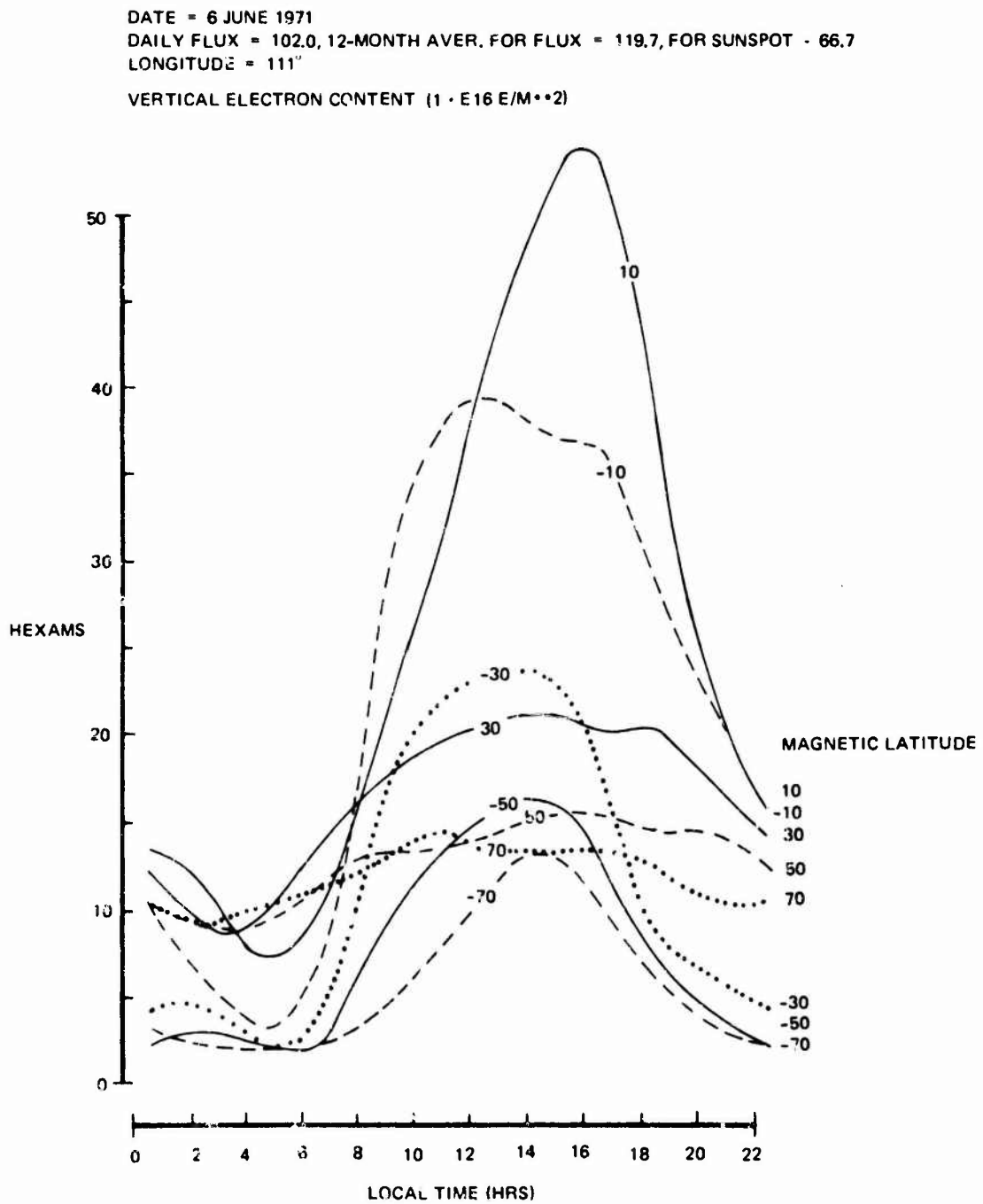


Figure 2 Ionosphere Electrons Content Curves During One Day Longitude 111°

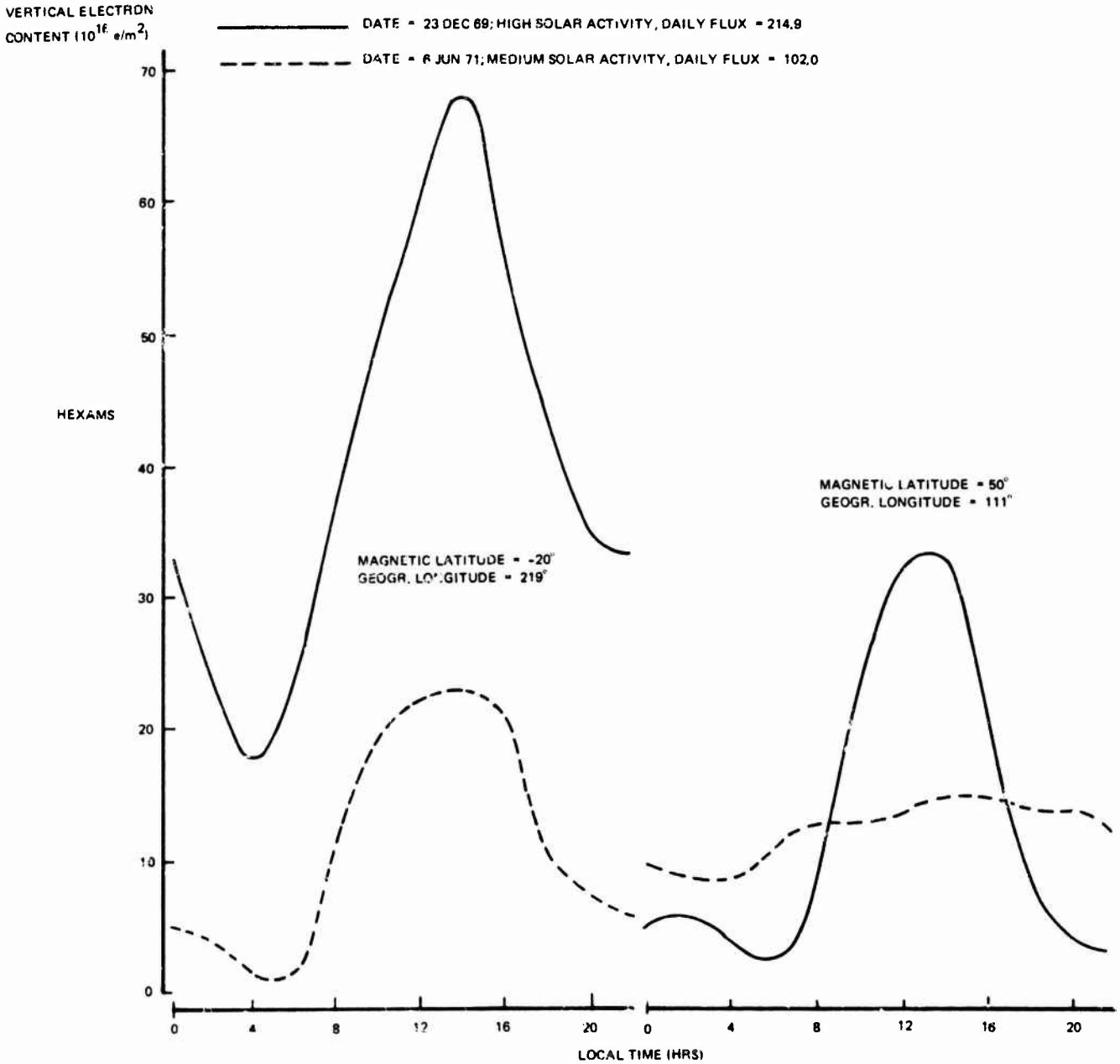


Figure 3 Ionosphere Electron Content For Two Days Two Geographical Locations

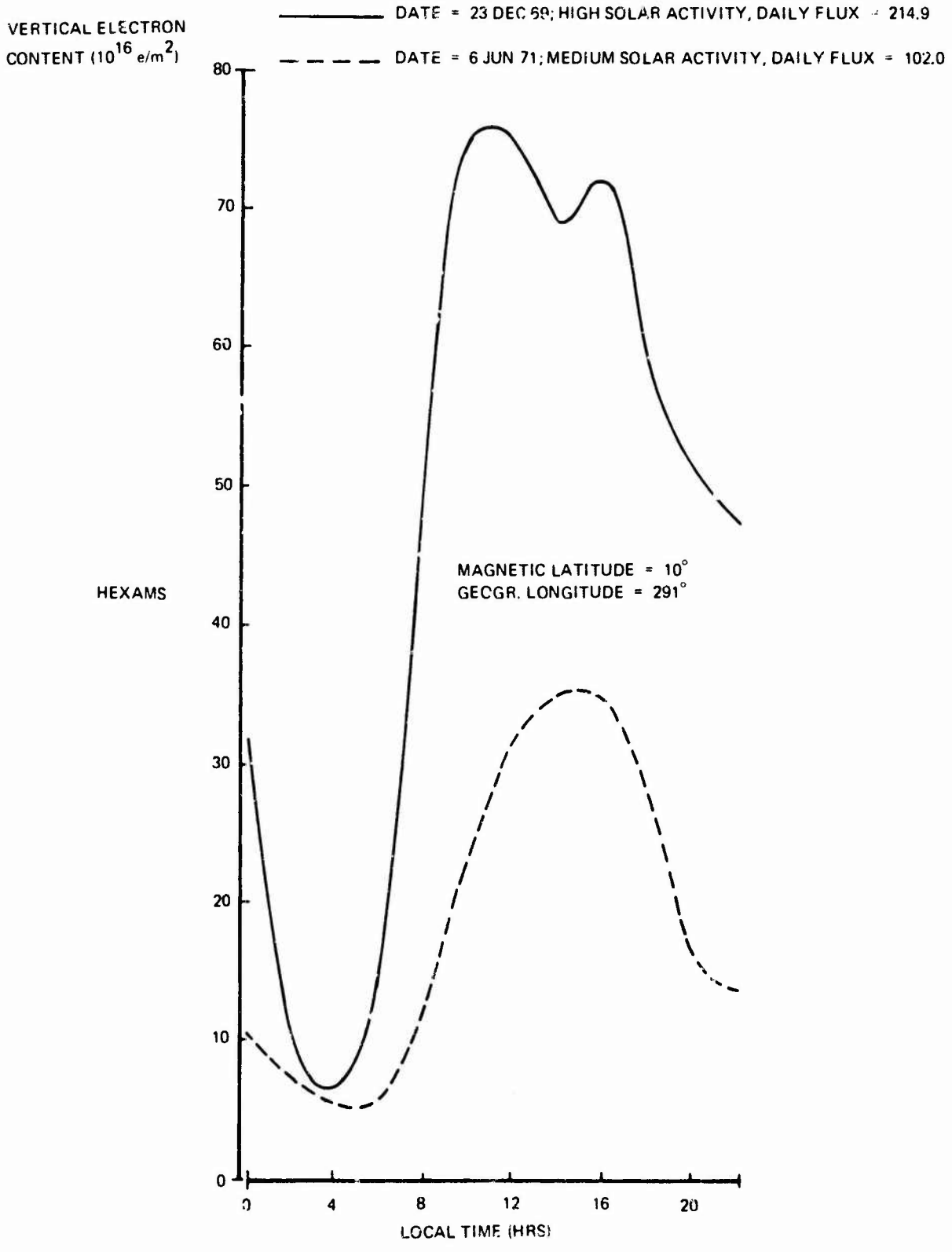


Figure 4 Ionosphere Electron Content For Two Days Single Location

The ionospheric range error due to group velocity slowing can be shown to be represented by the following approximate formula:

$$\Delta R = 1.32(1 + H/R_o)\, I_v / f^2 \left[\sin^2 E + 2H/R_o + H^2/R_o^2\right]^{1/2}, \qquad (2)$$

in which

R = range error, feet

H = effective height of the ionosphere (300-400 Km) (350 Km typical)

R_o = earth's equatorial radius (6378 Km)

E = elevation angle

f = radio frequency (GHz)

I_v = vertical electron content (Hexams)

A user or monitor station wishing to determine the ionospheric range error must, by some strategm, evaluate Eq (2) numerically, for the particular path to a satellite, at the particular time (to within 20 seconds) that a pseudo-range measurement is made. L-band frequencies for the GPS are L_1 = 1575 MHz, L_2 = 1232.5 MHz. Therefore, in Eq (2), f and Ro possess accurately known values.

The elevation angle E strongly affects Eq (2); therefore, this angle must be computed, from knowledge of the user's approximate position and the satellite position.

The ionosphere's effective height varies from place to place and time to time. For low elevation angles, below 20°, Eq (2) is strongly affected by the value of H; since the range error is largest at the smallest elevation angles, it is important to have an accurate value for H. Unfortunately, this is a very difficult parameter to measure accurately; a 10 percent uncertainty in H is typical.

In order to evaluate the effect of an uncertainty in H, differentiate Eq (2) partially with respect to H. The resultant may be cast into a form which yields the uncertainty of the computed range error, ΔR, for a given fractional uncertainty ΔH in H (ie, $\Delta H/H$) as a function of Iv and of elevation angle. Table 1 lists the end-product of such a computation, for an electron content/frequency combination which yields a 40-foot vertical range error, and for a 10 percent (typical) uncertainty in scale height H.

TABLE 1

RANGE ERROR UNCERTAINTY DUE TO SCALE HEIGHT UNCERTAINTY

R = 40 foot vertical
H = approximately 350 kM
ΔH = approximately ±35 kM

Elevation Angle, Degrees	Range Error Uncertainty, feet
0	±1.73
5	±1.71
10	±1.41
15	±1.09
20	±0.80

This table illustrates that a 10 percent uncertainty in H is tolerable (range error uncertainty less than one foot) for an angle of at least 20°, but only for a relatively "mild" ionosphere (40-foot vertical range error). Intense ionization, occurring about noon during a sunspot peak, will increase the range error uncertainties by a factor of 3 to 6.

7.0 DETERMINATION OF IONOSPHERIC RANGE ERROR

Five major options are available for handling the ionosphere range error, ie:

1. Ignore the error; this option is viable for low-precision users.

2. Compute the error from Eq (2), after making a determination of all the parameters in the equation. Possible techniques for determining Iv and H are discussed below.

3. Compute the error from Eq (2), using fixed values for all parameters except elevation angle in the equation.

4. Compute the error from Eq (2), using fixed values for all parameters, including elevation angle; this is equivalent to assuming a <u>fixed</u> ionospheric range error.

5. Exploit the fact that the range error varies as $1/f^2$, make pseudo-range measurements using two frequencies (L_1 and L_2), and combine the results such as to eliminate the ionospheric range error.

If the range error is completely ignored, the psuedo-range will always exceed the geometrical range, ie, the range error is always positive. A better, fixed approximation will be made if the range error is always assumed to be some fixed number of feet, N. The following argument is used to estimate a meaningful value for N:

If ranging is allowed over a full hemisphere, and if a given direction (elevation angle) to a satellite is as probable as any other direction, it is easy to show that the median elevation angle is 30°, ie, half the paths involve elevation angles greater than the median, and half smaller.

By referring to Figure 1, it is clear that an approximation to the vertical electrons content is a fixed quantity plus a 24-hour sinusoidal variation. Under this approximation the median is half the peak, that is, averaged over a 24-hour period the median electron content is half the peak value.

From other ionospheric data, not presented herein, it can be estimated that the peak value of Iv for each day of the year varies by about 3 to 1, so that the median daily peak is about twice the lowest peak. From this it is further estimated that the median world-wide daily peak is about 50 hex, half of which is 25 hex.

Therefore, a median value of range error can be computed from Eq (2), using Iv = 25 hex, E = 30°, H = 350 kM, and f = 1.575 GHz. This median value is 28 feet. It is suggested that all low-accuracy users should subtract this value from <u>every</u> pseudo-range measurement; half the measurements will be over corrected, and half uncorrected (some by much more than 28 feet).

8.0 THE PROBLEM OF IONOSPHERIC PREDICTION

8.1 <u>Magnitudes of the Range Error</u>

The most troublesome quantity in Eq (2) is I_v, the vertical electron content. The range error is directly proportional to I_v, hence the fractional error in ΔR is proportional to the fractional error in I_v.

The GPS is concerned with absolute, rather than fractional uncertainties in range error. When the range error is small (vertical path, light ionization) a large fractional error is tolerable, but when the range error is several hundred feet, a fractional error of only 1 percent results in an absolute uncertainty of several feet. Therefore, if range error is to be computed using Eq (2), accurate values of I_v are mandatory, particularly when Iv is relatively high.

Consider an RF frequency of 1.575 GHz (L1), vertical path (E = 90°), ionospherically quiet time (Iv = 1 hex), height = 400 km. Then the ionospheric term in Eq (2) amounts to about 0.6 feet, which is negligible for GPS purposes.

On the other hand, consider a low elevation angle (E=16°), 1.2325 GHz (L_2), intense ionization (Iv = 260 hex), and height = 400 Km. Then the ionospheric term amounts to approximately 570 feet.

Obviously such a range error is far too large to ignore. Also, obviously the correction itself must be accurately made, say to within about 1 percent (which would leave an unresolved range error of about 6 feet).

The problem for GPS, then, is how to provide the correction for the ionospheric term in Eq (2).

Strictly speaking, the ionospheric range error is proportional to the integrated electron content along the actual user-satellite path. Eq (2) has been derived on the assumption that, over the region of the ionosphere "visible" to a user the ionosphere is horizontally uniform, that is, gradients of I_v with respect to latitude and longitude are ignored. This is valid over a mid-latitude region such as CONUS, but the assumption is less valid in the polar and near-equatorial regions, where large gradients frequently occur.

8.2 Prediction Models

Computation of the range error from Eq (2) implies that values for Iv and H are available, for the particular time and geographical location at which a pseudo-range is measured. Numerous methods have been proposed for supplying this data to users and monitor stations, among which may be listed:

1. A fixed mathematical prediction model which permits calculation of I_v, given the user's local time, approximate geographical position, and day of the year.

2. A mathematical prediction model which largely contains fixed parameters, but which may be updated through the transmittal of a few coefficients from the ground segment, to the satellites, and hence to users on the P or C/A data frames.

3. A prediction model for which all coefficients are transmitted via the P or C/A data frames.

4. A prediction model requiring no coefficient transmission, but which does require the user to refer to a comprehensive lookup table, such as CCIR world-wide ionospheric data.

8.3 Raw Data for Model Update

Accumulation of the raw data constituting any prediction model is another problem, for the solution of which numerous techniques have been suggested, such as:

1. Processing of data taken from ground-based ionosonde measurements, which operate at low HF frequencies (up to a few MHz), and which basically measure ionospheric reflections

along vertical paths. This type of data is scattered - large areas of the world are not routinely covered - and the data is not available on a realtime basis.

2. Processing of data taken from long-range ionosondes, such as the 3-30 MHz unit proposed by Raytheon. This apparatus is said to be accurate to about ±10 percent out to 2000 km (1200 miles) from the instrument, ±20 percent out to 4000 km.

   Low frequency ionosondes, whatever their embodiment, provide what are called "bottom side soundings", ie, data on the lower region of the ionosphere, only up to the altitude having maximum free electron density; this region contributes, typically, only about one-third of the overall ionospheric range error. Bottom side sounding provides no direct information concerning the upper part of the ionosphere.

3. Two-frequency ionospheric range error data accumulated by GPS monitor stations (the two-frequency technique will be discussed presently).

4. Faraday rotation measurements of satellite downlink signals.

Historically, much of ionospheric research and data gathering has been concerned with the reflections of relatively low-frequency (up to a few MHz or tens of MHz) signals. For this purpose bottom-side sounding is ideally suited. It is only relatively recently that attention has been given to the upper ionosphere, and to the overall effect on UHF and microwave signals completely penetrating the ionosphere. Thus, much of the available data is incomplete for GPS purposes, or it is available only in an awkward form, requiring substantial additional modelling and data processing to produce L-band range error information.

The most reliable data on the overall ionospheric range error is obtained from Faraday rotation measurements on signals radiated from a small number of satellites, to an also small number of ground stations. This data has revealed several facts important to GPS:

1. The top structure of the ionosphere, above the altitude of maximum electron density, is not well correlated with the bottom structure. The top structure typically produces about 2/3 of the total range error, but bottom-side sounding provides data only on the lower region, contributing about 1/3 of the range error. Since top and bottom sides are not well correlated, bottom-side sounding cannot be expected to produce an accurate measure of the total range error of rays completely penetrating the ionosphere.

2. There are geographical regions, largely clustered near latitudes 17° N and S, and in the polar regions, where the vertical electron content varies rather rapidly for relatively small changes in position, ie, large horizontal

gradients of I_v appear. Furthermore, the highest values of I_v occur in these regions.

3. There are regions, notably in the vicinity of the magnetic poles - or more generally, in the geographical polar regions - where the vertical electron content can exhibit large short-term (minutes of time) rates of change, up to 2 hex per minute. Very little data is available on this effect. The data does indicate that the fluctuations can be large only when I_v itself is relatively high; the fluctuations appear to be unpredictable, except that they are most likely during early morning hours.

The only raw measurement data in existence pertaining to range error of signals completely penetrating the ionosphere consists of Faraday rotation measurements, taken at a few ground stations in CONUS, Alaska and Hawaii, and a few, short-term data sets gathered at other stations scattered around the world. Now, Faraday rotation measurements are extremely useful for learning about Faraday rotation, but these measurements are less than ideal (although they provide the best data available) for determining the range error of signals completely penetrating the ionosphere.

It is possible to show that, when an electromagnetic plane wave propagates in a plasma (the ionosphere) which is immersed in a steady magnetic field (the earth's magnetic field in space), such that the wave propagation direction possesses a component coinciding with the magnetic flux lines, the direction of polarization of the wave gradually rotates (Faraday rotation) as the wave propagates. On the other hand, if the wave propagation is entirely perpendicular to the magnetic flux, no rotation occurs.

It is further possible to show that the net wave rotation of a wave completely penetrating the ionosphere is proportional to:

- The integrated electron content along the ray path
- The integrated component of magnetic flux tangential to the wave direction

The electron content is the quantity pertaining to range error; the integrated tangential magnetic flux does not pertain to range error, and must be "removed" from Faraday rotation data before that data can be interpreted in terms of range error.

It is at, and near, earth's geomagnetic equator (where the earth's magnetic field is entirely horizontal) that the integrated electron content is anomalously high, and may exhibit large horizontal gradients. But at the geomagnetic equator a signal reaching the earth from a high altitude satellite has propagated almost entirely perpendicular to the magnetic flux, and has thereby not experienced appreciable Faraday rotation. Therefore, Faraday rotation measurements, as a technique for determining range error, are at their

worst in just those regions of the world where the most accurate data is desirable.

Dr. da Rosa, at Stanford, has estimated that, over CONUS, Faraday rotation data provides integrated electron content data which is probably in error by at least 10 percent, and that what little data has been obtained at lower latitudes, near the geomagnetic equator, provides electron content figures which may be in error by 40-60 percent. Since Faraday rotation measurements constitute the only known technique (at least, as far as the author can determine) for determing integrated electron content of rays completely penetrating the ionosphere, and since such measurements are largely confined to CONUS, it follows that:

- No data bank exists pertaining to world-wide electron content, over extended time periods.
- Such data as does exist, even for CONUS, is of questionable absolute accuracy, as a measure of electron content.

An important consideration concerning surface-located ionospheric sounders is the percent of earth coverage afforded by each station. This quantity is readily calculated if the ionospheric effective height is uniform (say 350 km). Based upon this assumption, the coverage from a single station, as a function of minimum elevation angle, is presented in Table 2. Note that, if no gaps of coverage occur, each station covers only $\pi/4$ of the area of the spherical cap corresponding to the ionosphere subpoint. Table II also includes the number of ground stations for full coverage of:

- The entire earth (ie, the entire ionosphere)
- The CONUS area, 68° W to 124° W, 25° N to 49° N, with additional angular limits to permit line-of-sight to a satellite right down to the horizon; this region is 50°-142° W, 07-67° N, and is called "extended CONUS".

TABLE 2

PERCENT IONOSPHERIC COVERAGE VS ELEVATION ANGLE
H = 350 km

ELEVATION ANGLE DEGREES	PERCENT COVERAGE PER STATION	NUMBER OF STATIONS; WORLDWIDE	FULL COVERAGE EXTENDED CONUS
0°	2.6%	49	5
10°	0.857%	149	15
20°	0.376%	338	35

This table shows that, if the entire ionosphere is to be monitored, 49 strategically-located stations, each taking data right down to the horizon (where the tropospheric error is extremely insidious) would be required. For extended CONUS five (5) such stations suffice, given suitable locations.

The requisite number of monitors increases rapidly as the minimum elevation mask increases (to alleviate large uncertainties in the tropospheric range error). In fact, three times as many stations are required for a 10° mask, and about seven times as many are required for a 20° mask. Clearly, it is a sizeable problem to monitor even the extended CONUS ionosphere, considering the cost of the monitor equipment and realtime transmittal of the data back to a central processing facility where the master ionosphere model is maintained.

Four (4) monitor stations will be utilized to maintain the satellite orbit computations, at least in GPS Phase I, and possibly also in Phases II and III. Obviously, such a small number of ground stations would provide only very sketchy data on the world-wide ionosphere, and in fact would provide only about 1/40 coverage, assuming a 10° mask, or 1/84 coverage with a 20° mask. In this event the prediction model itself would have to fill the gaps, thereby introducing indeterminent uncertainties in prediction accuracy.

8.4 Mathematical Prediction Models

Numerous techniques are conceivable for the mathematical form of an ionosphere prediction model. Because of the inherent 24-hour cyclical nature of the ionization variation over any specific location, it is natural to expect a Fourier series with a basic period of 24 hours to yield meaningful results.

Even a casual inspection of Figures 1 through 4 shows that the I_V variations, although cyclical, are not sinusoidal; therefore, several harmonics must be included in a Fourier series. Furthermore, the variation with respect to latitude also requires several terms for its representation. Finally whatever coefficients represent I_V as a function of latitude and time of day necessarily change from day to day.

Over CONUS the ionosphere is known to vary relatively smoothly, and seldom exhibits large horizontal gradients, either with respect to latitude or longitude. This suggests that a model containing only a few coefficients could represent the ionosphere reasonably accurately, ie, such as to meet requirements of medium-accuracy users, provided the coefficients are updated no less often than daily. Such a model could be maintained at the GPS master control station, with raw data being supplied by monitor stations. Coefficients would then be computed from the master model for transmittal to satellites and thence to users.

Considerable work has been performed by various workers in attempts to develop ionospheric range error models. Almost all of the work known to WDL has been limited to the CONUS or near-CONUS region of the world

- very little work has been expended upon a world-wide model. The results of this effort can be summarized as follows:

1. Models containing several hundred coefficients of a Fourier series, updated daily with a single coefficient representing solar radiation activity, statistically will yield the vertical range error at 1.575 MHz to within ±10 feet (more correctly, 3 m) about 70 percent of the time; the error will exceed ±33 ft (ie, 10 m) about 5 percent of the time. These accuracy figures apply only at a few CONUS locations, and are believed to hold within about 1000 km (600 statute miles) from each of these locations.

2. It appears that no one has developed an ionospheric prediction model which has been evaluated at <u>many</u> locations over CONUS.

3. It also appears that little, if any, work has been performed on a highly simplified model, containing only a relatively small number of coefficients. No data is available to indicate how fast the large-scale model deteriorates if the smaller coefficients are omitted.

What is desired is a simple mathematical formula, containing a few empirically-adjusted constants (ie, coefficients), from which a user could compute the vertical electron content at his location (latitude and longitude) and local time; ideally the computation would result in an ionosphere range error uncertainty of less than 1 or 2 feet, and a given set of coefficients would be valid for at least 24 hours into the future.

Most previous work (Stanford Model, Bent Model) consists of curve fitting to past data, and <u>not</u> prediction of next-day behavior. What is needed is a technique which will combine range error data measured today, yesterday, the day before, etc, and from this data produce a prediction of next-day range error. The work at Stanford, for example, has consisted of fitting an elaborate Fourier series (507 terms) to Faraday-rotation data taken (at Palo Alto) at 5-minute intervals over about 3 years. Having fitted the Fourer series to the data, computations were then made of the difference between the Fourier series and the actual data. On the other hand, no attempt was made to assess the accuracy of the Fourier series for predicting the ionosphere over Palo Alto beyond the time span of the actual data. (However, Stanford's raw data constitutes an excellent data bank for some future worker engaged in evaluating prediction models of varying complexity!)

8.5 <u>Complete Stanford Model</u>

The range error using the complete Stanford model representing the ionospheric is executed in five steps, outlined below. This technique requires a <u>daily</u> update of a single coefficient S, which represents a moving average of solar flux activity. A value of S must be sent to

users via the satellites, and a particular value of S will hold for 24 hours.

The mathematical algorithm is as follows:

Step 1. Compute a reference value of the prognosticated vertical electron content Io from the Fourier series given below:

$$I_o = \sum_{i=o}^{12} \sum_{j=o}^{12} \sum_{k=o}^{2} C_{ijk} F_i(2\pi t/24) F_j(2\pi d/365) S^k \qquad (3)$$

in which

$$F_i(x) = \cos(ix) \quad \text{if} \quad 0 \leq i \leq 6$$
$$= \sin(1-6)x \quad \text{if} \quad 7 \leq i \leq 12$$

The series represented by Eq (3) contains a total of 507 fixed coefficients, values of which have been obtained by da Rosa.

Step 2. Compute a geographical factor g for a given user location, specified by geographical latitude and geomagnetic declination, and also including day of the year d, solar activity coefficient S and local time t. The factor g is computed from the Fourier series.

$$g = \sum_{i=o}^{3} \sum_{j=o}^{3} g_{ij} G_i(\lambda, \delta) H_j(d, s) \qquad (4)$$

in which

$$G_o = 1$$
$$G_1 = \sin \lambda$$
$$G_2 = \cos \lambda$$
$$G_3 = \sin \delta$$
$$H_o = 1$$
$$H_1 = S$$
$$H_2 = S \cos (2\pi d/365$$
$$H_3 = S \sin (2\pi d/365)$$

The coefficients g (ij) change with each whole hour of local time.

Step 3. Compute the prognosticated vertical electron content I_v from the product,

$$I_v = gI_o;$$ divide by 10^{-16} to convert to HEXAMS.

Step 4. Compute elevation angle E of a straight line from user to satellite.

Step 5. Compute the ionospheric range error ΔR from Eq (2).

The technique outlined above is known as the complete Stanford model; it is complete in the sense that a specific mathematical algorithm is provided for computing the ionospheric range error on any day, at any time anywhere in the continental US, taking account of local (user) time, latitude, and (geomagnetic) declination. The model requires knowledge of about 700 fixed constants, plus a single coefficient which changes from day to day, and which must be supplied via the satellite.

Errors pertinent to Stanford University, to Ft. Collins, Colo, and to Urbana, Illinois, and within about 1000 km therefrom, are as follows: (NOTE: herein "error" is defined as the uncorrected range difference between an overhead satellite and one having an elevation angle of 15°)

Locations	% of time errors ±1 meter	% of time errors ±3 meters	% of time errors ±10 meters
Stanford	69%	5%	-
Ft. Collins	62%	6%	0.2%
Urbana	63%	5%	0.04%

No study effort has been expended to determine how much the Stanford model can be simplified, and the error statistics resulting therefrom.

The method is applicable at best to CONUS; extension to a world-wide coverage implies the utilization of a substantial number of ionospheric sounding stations which neither exist nor are currently planned by any known agency, foreign or domestic.

8.6 Approximate Trigonometric Model

An examination of the coefficients for the Stanford ionosphere model leads to the following conclusions:

1. Coefficients for K = 1 and K = 2 are generally two to five orders of magnitude smaller than those for K = 0. This immediately suggests that a reasonably good approximation to

the Stanford model should be possible by ignoring all terms for which K = 1 or 2.

2. The coefficients for J = 1, 2, 3 are generally three or more orders of magnitude smaller than those for J = 0. This suggests that the g factor can be approximated as

$$g = g_{00} + g_{10} \cos\lambda + g_{20} \sin\lambda + g_{30} \sin\delta.$$

Furthermore, the coefficient g_{30} is generally much smaller (10 percent or less) than the other three coefficients in the above approximation to g; hence it appears reasonable to neglect g_{30}.

3. According to a statistical analysis made by Dr. da Rosa at Stanford, the Stanford model results in residual errors of ±3 meters or more (about 10 feet) for some 5 percent or 6 percent of the time. Now if the maximum range error predicted, at low elevation angles, is about 300 feet, then the model predicts this error to within about ±3 percent for 94 or 95 percent of the time. Therefore, it is meaningful to ingore all terms in the Stanford model small than about ±3 percent of the largest term.

4. Coefficients of the 3rd, 4th, 5th and 6th harmonic terms are never larger than about 7 percent of the largest term, C_{000}. Secondly, in the list of coefficients for I = 0, 1, 2, 7, and 8, K = 0, there are a number of values smaller than 7 percent of C_{000} (I = 0, J = 0, K = 0, $C_{000} = 4.642 * 10^{12}$).

These considerations have led to a gross simplification of the Stanford model, in which <u>all</u> coefficients smaller than 7 percent of C_{000} are omitted; there are 27 remaining coefficients. Deletion of all terms having K = 1 or 2 in Eq (3), and retention only of g_{00}, g_{10}, and g_{20} in Eq (4), results in an approximate formulation of I_0 and of g which does not contain the solar flux parameter S, and which therefore requires no input data from the satellites. In summary:

<u>Step 1</u>. Compute I from the relation

$$I_o = \sum_{i=0,1,2,7,8} \sum_{j=0}^{12} C_{ijo} F_i(2\pi t/24) F_j(2\pi d/365)$$

<u>Step 2</u>. Compute g from

$$g = g_{00} + g_{10} \cos\lambda + g_{20} \sin\lambda$$

where λ is the latitude of the user.

Steps 3 and 4 are as described in section 8.5, above.

The accuracy of this approximation is not known, but it is certainly not expected to be as good as the complete Stanford model. The number of coefficients is reasonable, and no inputs are required from satellites (better accuracy should be obtainable if all the coefficients are updated perhaps quarterly or yearly, but not more often).

9.0 A TWO-FREQUENCY CORRECTION TECHNIQUE

9.1 The Basic Algorithm

It was stated in section 4 that the ionospheric range error (above 1 GHz), pertaining to a given path at a given time, varies as inverse frequency squared. This suggests that if a satellite-user pseudo-range is measured using two well-separated carrier frequencies, it should be possible to combine the results of the measurements in such a way as to eliminate the range error, and thereby obtain true range between satellite and user. Not only is this true, but it is also true that the data-processing algorithm which completely eliminates the f^{-2} range error will also strongly suppress higher frequency dependence; this point order range error terms, such as a term having f^{-4}. in the UHF region).

Let us therefore develop the data processing algorithm which eliminates the relatively large f^{-2} term, but simultaneously including two higher-order terms varying as f^{-3} and f^{-4}.

Since we are concerned at this point with the development of a suitable data processing algorithm, and not with an error analysis, it is assumed for simplicity that individual pseudo-range measurements are perfect, ie, not confounded by roundoff or finite signal-to-noise-ratio error. An error analysis is given in section 10.

The two carrier frequencies are designated as f_1 and f_2, with $f_1 > f_2$; it is assumed these frequencies are precisely known. Then the two pseudo-range measurements yield the following results:

$$R_1 = R_T + A/f_1^2 + B/f_1^3 + C/f_1^4 \tag{5}$$

and

$$R_2 = R_T + A/f_2^2 + B/f_2^3 + C/f_2^4, \tag{6}$$

in which

R_1, R_2 are the two measured pseudo-ranges,

R_T is the true range,

A, B, C are parameters which depend upon the satellite-user path and the instantaneous integrated electron content of the ionosphere.

The object now is to combine Eqs. (5) and (6) so as to obtain the true range R_T, and eliminate the $1/f^2$ term. This is accomplished by:

1. multiplying Eq (6) by $(f_2/f_1)^2$
2. subtracting the result from Eq (5)
3. solving for R_T; that is, upon letting $f_2/f_1 = p$

$$R_1 = R_T + A/f_1^2 + B/f_1^3 + C/f_1^4 \tag{5}$$

$$p^2 R_2 = p^2 R_T + p^2 A/f_2^2 + p^2 B/f_2^3 + p^2 C/f_2^4$$

$$p^2 R_2 = p^2 R_T + A/f_1^2 + p^2 B/f_2^3 + p^2 C/f_2^4 \tag{7}$$

Subtraction of Eq (7) from Eq (5) yields

$$R_1 - p^2 R_2 = (1 - p^2) R_T + B\left(1/f_1^3 - p^2/f_2^3\right) + C\left(1/f_1^4 - p^2/f_2^4\right);$$

this is solved for R_T upon transposing the B and C terms, and dividing by $1 - p^2$, ie,

$$R_T = \left(R_1 - p^2 R_2\right)/(1 - p^2) + \left(B/f_2^3\right)(p^2/1 + p) - \left(C/f_2^4\right)p^2. \tag{8}$$

because $p^2/f_2^2 = 1/f_1^2$.

Therefore, the measured pseudo-ranges R_1 and R_2 should be combined with the aid of Eq (9).

On the other hand, if B and/or C are not zero, and if Eq (9) is used to compute the true range, that computed range will be in error because of the terms in B and C.

It is possible to show that coefficient B is proportional to the cube of earth's magnetic field strength averaged along the satellite-user path; in addition, it is also possible to show that the range error term B/f_2^3 does not exceed a few centimeters if f_2 is at least 1 GHz.

Now, in Eq (9), the term involving B contains the factor B/f_2^3 (the "magnetic" range error at the lower frequency f_2), but this factor is reduced, or suppressed, by $p^2/(1+P)$, which is less than unity. In fact, for f_1 = 1575 MHz, then $p = 0.781$, $p^2 = 0.610$, $p^2/(1+P) = 0.610/1.781 = 0.342$.

Similarly, the term involving C in Eq (9) is also small - a few inches or less - for frequencies above 1 GHz; in Eq (8) C/f_2^4 is the range error at the lower frequency, but this error is suppressed by the factor p^2 (= 0.61 in the case above).

In summary, the two-frequency data-processing formula represented by Eq (9) completely eliminates the relatively large, troublesome f^{-2} term, and suppresses high-order range error terms in frequency.

It is extremely important to note that the term containing C in Eqns (5) and (6) increases rapidly at frequencies below 1 GHz. For example, if C/f^4 equals 1" at 1230 MHz, then in the UHF range, at 1/4 of f_2 (ie, at 307.5 MHz) this term contributes a range error which is 4^4 = 256 times larger, or 21.3 feet. At such a low frequency the f^{-2} term can exceed 1.5 <u>miles</u> when the ionization is intense.

The outstanding advantage of this two-frequency technique is that it directly and simply provides for elimination of what is by far the dominant ionosphere range error terms, ie, the f^{-2} term, at the exact moment, along the exact path, that the correction is needed. The method neither over-corrects nor under-corrects, and it does not rely upon a prediction model of dubious accuracy and complexity.

The outstanding disadvantage is that the method requires measurements using two separate carrier frequencies, a fact which unfortunately increases receiving hardware and software complexity.

9.2 Non-Simultaneous Measurements

If the two range measurements, at f_1 and f_2, are not made precisely simultaneously the results of the previous analysis must be modified in four (4) important ways. These are:

1. If R_T is the true range when the first measurement, say at f_1, is made, this range will have changed by an amount Δ_{RT} when the second measurement is made. However, if the relative velocity and acceleration between user and satellite are known, Δ_{RT} is a predictable quantity.

2. In the ionospheric term the elevation angle E will in general change by an amount $\Delta E = (dE/dt)\ \Delta t$, where Δt is the time increment between first and second measurements, and Δt is assumed short enough to permit a first-order approximation. The angle E and the rate (dE/dt) are calculable, given the (approximate) positions and motions of satellite and user.

3. In the interval Δt the integrated electron content Iv along a fixed path will in general, experience a change $\Delta I_V = (\partial I_V/\partial t)\Delta t$.

4. Because of the change in elevation angle by the amount E, the second measurement will be made along a different path, having in general a different I_V, than the first path. Thus, the integrated electron content for the second path differs from that of the first by an increment $\Delta I_V = (\partial I_V/\partial E)\Delta E$.

The first effect, ΔR_T, is assumed to calculable by the user receiving equipment.

The changes in the ionospheric term have been estimated under approximately worst-case conditions, subject to the following assumptions:

1. Measurements are being made along a path for which I_V is increasing from virtually 0 hex (pre dawn) to a peak of 240 hex in 4 hours. Thus, dI_V/dt = 1 hex/minute.

2. Orbital data pertinent to the Vandenberg monitor station were used to obtain the maximum value of dE/dt (= 5 x 10^{-3} deg/sec) and the elevation angle at which this maximum occurs (E=45 degrees).

The details of this part of the anlaysis are not included here; the final results are:

1. The range error increment resulting from effect 2, time rate of change of elevation angle, never exceeds -0.03 foot/second. In one second the error is -0.03 foot, and in ten seconds -0.3 foot.

2. The range error increment resulting from effect 3, rate of change of I_V with respect to time, never exceeds +0.05 foot/second. In one second the error is +0.05 foot, and in ten seconds +0.5 foot.

3. The range error increment resulting from effect 4, rate of change of I_V with respect to E, is extremely small, amount to $\pm 1.4 \times 10^{-4}$ foot/second, depending upon whether I_V increases or decreases with increasing longitude. In one second this error amounts to ±0.00014 foot, and in ten seconds ±0.0014 foot.

From these estimates it is concluded that as far as the ionospheric correction term is concerned, if a pair of measurements on two frequencies f_1 and f_2 are performed within 20 seconds of one another, the measurement may be considered to be simultaneous. In other words, a time lapse of 20 seconds does not introduce an additional range uncertainty of more than one (1) foot. It is emphasized that this conclusion pertains only to measurements between a given user and a single satellite, ie, along essentially the same RF path. In view of the approximate nature of the modelling employed to reach this conclusion, to be on the safe side the time increment Δt between f_1 and f_2 measurements should be held as far below 20 seconds as can be accomplished without undue user equipment cost and complexity. In particular, the conclusion is valid only for a stationary or low-velocity user. A high-speed user analysis has not been performed; but it is only reasonable that as user speed increases the whole time scale of the measurement process should shrink, ionospheric range error variations not withstanding.

10.0 L_1/L_2 RANGE ERROR STATISTICS

10.1 Statement of the Problem

It has been established in previous sections that, through the use of two carrier frequencies f_1 and f_2 the troublesome ionospheric range error can be essentially removed from user-satellite pseudo-range measurements. This section derives and discusses formulas from which range error statistics can be calculated, given pseudo-range measurements on two separate frequencies, and given the statistics of measurement error pertaining to a single measurement.

Several assumptions are made in the derivation, namely;

1. The range measurement error pertaining to a single measurement is assumed to obey a Gaussian distribution, that is, the probability P (x) that a measured range R experiences an error X is given by the relation

$$P(x) = h \exp(-h^2 x^2) \Big/ \sqrt{\pi} \qquad (10)$$

Standard references on statistical analysis can be consulated (such as J. B. Scarborough, "Numerical Mathematical Analysis", Johns Hopkins Press, 1958, sections 122 and 133) to verify that the root-mean-square error Xrms is related to the index of precision, ie,

$$X_{rms} = 1/\sqrt{2}\,h \tag{11}$$

2. Although it is understood that, in the proposed GPS user receiver a small time lapse occurs between range measurements on two separate frequencies, and that both the satellite and the user generally will experience relative motion and acceleration during this time lapse, it is assumed that motion/ acceleration effects have been "corrected out" of the measurements. It is <u>not</u> the purpose of this section to discuss errors associated with the velocity and acceleration corrections.

3. It is assumed the two measurements are close enough together in time that the user-satellite ray path through the ionosphere and the troposphere is constant. In section 9 it was stated that two measurements are essentially simultaneous, as far as the ionosphere and troposphere are concerned, if they are made within 20 seconds of time.

4. It is assumed that the tropospheric range error R_{tropo}, along a given path at a given time, is independent of carrier frequency. Because the tropospheric range error is assumed not to vary with carrier frequency it follows that range measurements performed using two frequencies provides no information concerning the troposphere. This error must therefore, be "removed" through the use of a prediction model. This section is <u>not</u> concerned with residual errors of such a model.

5. It is assumed that the ionospheric range error, along a given path at a given time (within 20 seconds) varies as the inverse square of the carrier frequency, $R_{ionos} = K/f^2$. This assumption - basically that the ionospheric error is an exactly deterministic function of frequency is the key to the two-frequency method of correcting pseudo-range measurements for the ionospheric group delay effects.

10.2 <u>Conclusions</u>

The following two statements will be proved or justified subsequently:

1. The root-mean-square range error of the two frequency technique, as a consequent of Gaussian errors pertaining to

the individual measurements, is given by the following formula:

$$\sigma(R) = \left[\sigma_1^2(R) + p^4\sigma_2^2(R)\right]^{1/2} / (1 - p^2) \qquad (12)$$

in which

$\sigma_1(R)$ is the RMS measurement error pertaining to the higher frequency f_1 (ie, L_1)

$\sigma_2(R)$ is the RMS measurement error pertaining to the lower frequency f_2 (ie, L_2)

$p = f_2/f_1 < 1.$

2. Eq (12) is valid regardless of the algorithm used to process the measured ranges, R_1 and R_2.

The RMS measurement errors σ_1 and σ_2 in general will not be equal, primarily because the carrier-to-noise ratios C_1/N_o and C_2/N_o will not in general be equal at the two frequencies. It is recognized the effective values of these errors can be reduced by making multiple measurements and averaging. It is also recognized that σ_1 and σ_2 actually are the resultants of three predominant components, ie, a noise error (due to finite C/N), quantization error (due to the digital nature of the measurement) and instrumentation errors. Here it is assumed only that σ_1 and σ_2 are the consequences of Gaussian error statistics; TRW has provided detailed data on the magnitudes of these RMS errors, for an assumed receiver design and for various C/N values (DNSDP-AG-230, 7 January 1974, A. Garabedian to M. Deckett, with corrections as supplied by H. Perasso of WDL).

10.3 Derivation of the Range Error Equation

Suppose pseudo-range between a user and a satellite is measured with two frequencies f_1 and f_2, with $f_1 > f_2$. After correction for relative motion and acceleration, and after averaging over a number of measurements to reduce the RMS measurement error, the results of the measurements, ie, the "measured and corrected" pseudo-ranges R_1^1 and R_2^1 can be represented in the following equations:

$$R_1^1 = R_T + K/f_1^2 + \Delta R_{tropo} + \epsilon_1 \qquad (13)$$

$$R_2^1 = R_T + K/f_2^2 + \Delta R_{tropo} + \epsilon_2, \qquad (14)$$

R_T is the true, or geometrical range, K is true value of the governing ionospheric constant, ΔR_{tropo} is the tropospheric range error, the same for both measurements. ϵ_1 is the residual error pertinent to the pseudo-range measurement using f_1, and ϵ_2 similarly corresponds to f_2.

The object of the measurements is to obtain the best possible value for R_T, pertaining along the particular user-satellite path, at the particular time (ie, corrected reference time) of the measurements.

A user has no need for the ionospheric constant K; the user's job is to combine Equations (13) and (14) such as to eliminate the ionospheric range error terms. On the other hand, GPS monitor stations will, in addition to obtaining R_T, also wish to obtain K, for the purpose of supplying input data to an ionosphere prediction model. A user will not require this raw-data coefficient.

For a single set of measurements the errors ϵ_1 and ϵ_2 are uncorrelated, although both are assumed subject to Gaussian distributions.

It was stated in section 2 that the tropospheric term ΔR_{tropo} is to be removed with the aid of a prediction model. It is immaterial at what stage of the data processing this correction is made, ie, Eqns (13) and (14) may be combined as-is to obtain $R_T + \Delta R_{tropo}$, following which ΔR_{tropo} may be substracted, or ΔR_{tropo} may be subtracted from both R_1^1 and R_2^1, and then these results combined to obtain R_T. With no loss of generality it is therefore, assumed that ΔR_{tropo} is removed first, yielding

$$R_1 = R_1^1 - \Delta R_{tropo} = R_T + K/f_1^2 + \epsilon_1 \qquad (15)$$

$$R_2 = R_2^1 - \Delta R_{tropo} = R_T + K/f_2^2 + \epsilon_2 ; \qquad (16)$$

now combine these two equations to obtain a value for R_T.

10.4 First Method, Direct Solution

In order to obtain R_T directly from Eqns (15) and (16), and eliminate the terms in K, proceed as follows:

1. Multiply Eq (16) by $(f_2/f_1)^2$; let $f_2/f_1 = p$. This yields

$$p^2 R_2 = p^2 R_T + p^2 K/f_2^2 + p^2 \epsilon_2 ;$$

$$p^2 R_2 = p^2 R_T + K/f_1^2 + p^2 \epsilon_2 \quad , \qquad (17)$$

since

$$p^2 K/f_2^2 = K/f_1^2 .$$

2. Subtract Eq (17) from Eq (15), yielding

$$R_1 - p^2 R_2 = R_T(1 - p^2) + \epsilon_1 - p^2 \epsilon_2 \text{ ; solve for } R_T\text{:}$$

$$R_T = (R_1 - p^2 R_2)/(1 - p^2) + (p^2 \epsilon_2 - \epsilon_1)/(1 - p^2) . \qquad (18)$$

In this equation the first term on the right side represents the apparent, or computed value of the range; it is this number, and this number only, which will be computed from the measured ranges R_1 and R_2 (ie, after correcting the actual measurements for the troposphere).

In Eq (18) the term $(p^2\epsilon_2 - \epsilon_1)/1-p^2$, represents the (unknown) error in the knowledge of R . Since, during a single pair of measurements there is no way to determine the quantities ϵ_1 and ϵ_2 (else they would be deterministically correctable, hence would not represent errors), the computed value of range must differ from the true range by the error term in Eq (18).

10.5 Second Method, Indirect Solution

In this method Eq (15) is first subtracted from Eq (16) to provide a range difference quantity $\Delta = R_2 - R_1$; next the difference equation is solved for K; finally the computed K is substituted back into Eq (15) (the measurement made at frequency f_1), and that equation solved for R_T. Thus

$$R_2 - R_1 = \Delta = K\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right) + \epsilon_2 - \epsilon_1 ; \tag{19}$$

Solving for K yields

$$K = \Delta/q + (\epsilon_2 = \epsilon_1)/q \; ; \; q = 1/f_2^2 - 1/f_1^2 \; . \tag{20}$$

Then, upon substituting for K in Eq (15), using Eq (20), and solving for R_T there is obtained

$$R_T = \left[R_1 - p^2\Delta\right]/(1 - p^2) + (p^2\epsilon_2 - \epsilon_1)/(1 - p^2) \; . \tag{21}$$

In Eq (21) the measured quantities are R_1 and Δ, p is the (accurately known) frequency ratio f_2/f_1, and the computed apparent range is $R_1 - p^2\Delta/(1-p^2)$. Once again, in exactly the same form as in Eq (18), the error term is $(p^2\epsilon_2 - \epsilon_1/(1-p^2)$.

Consequently, the residual error is exactly the same, regardless of whether the range is computed (directly) from R_1 and R_2, using Eq (16), or from R_1 and Δ, using Eq (21). Given the two measured ranges R_1 and R2, each containing an unknown error, one cannot improve upon the accuracy of the computed range R_T by selecting one computational algorithm over another. The algorithm is selected for computational convenience, and not because one is more accurate than another.

10.6 The RMS Range Error Equation

The measurement error term in Eqns (18) and (21) is simply $E = (p^2\epsilon_2 - \epsilon_1)/(1-p^2)$. If, as has been asserted, ϵ_1 and ϵ_2 individually follow Gaussian statistics, then so does ϵ, which is a weighted linear combination of ϵ_1 and ϵ_2 (see Scarborough, loc. cit., section 122). Following Scarborough's analysis it is therefore straight-forward to prove that, if σ_1 is the RMS value of range measurement error corresponding to f_1, and similarly for σ_2, corresponding to f_2, the RMS value σ of the weighted combination is given by

$$\sigma^2 = \left(\sigma_1^2 + p^4\,\sigma_2^2\right) \Big/ (1 - p^2)^2 \quad . \tag{22}$$

10.7 TRW Values for RMS Range Errors; Empirical Approximations

TRW has supplied values for σ_1 and/or σ_2, for various values of C/No; a distinction is made for the P-code and the C/A-code. The data are listed in Table 3, below. (See DNSDP-AG-230, pages 8 and 9; corrections to the listed data have been supplied by H. Perasso of WDL.) These approximations are useful for determining the optimum power division between L_1 and L_2 signals, given that the sum of the signal powers is a constant.

TABLE 3

RMS RANGE ERROR VS C/No

C/No dB	(RMS), feet P-CODE	C/A-CODE
24	9.0	171.2
26	6.2	111.2
28	4.5	72.8
30	3.3	48.7
32	2.5	33.3
34	1.9	23.3
36	1.5	16.8

It is straightforward to show that these sets of data closely follow the empirical approximations given below, with C/No given as a ratio, ie, not in decibels.

$$\sigma = 17000 \,/\, (C/No)^{0.843} \text{ feet} . \tag{23}$$

For the P code

$$\sigma = 316 \,/\, (C/No)^{0.65} \text{ feet} . \tag{24}$$

11.0 OPTIMUM POWER DIVISION BETWEEN L_1 AND L_2 SIGNALS

11.1 Statement of the Problem

This section examines the effect on the overall residual RMS range error of a two-frequency (L_1/L_2) technique for eliminating the ionospheric range error, as a function of relative power of the L_1 and L_2 signals; it is assumed that the total power of the two signals, ie, the sum of the individual signal powers, is a constant. The analysis is based upon the fundamental inputs provided from earlier sections; these are:

1. Detailed values of user receiver range error (one-sigma) as a function of signal carrier-to-noise ratio. This data is given in section 10.

2. A formula for the residual RMS range error associated with the two-frequency ionosphere correction technique. This formula is Eq (22).

11.2 Summary

The following statements are proved or justified subsequently:

1. Regarding the P-code signal only, the RMS range error is minimized if the L_1 signal power is 1.86 dB above the L_2 signal power at the user receiver RF input terminal.

2. Regarding the C/A-code signal only, the RMS range error is minimized if the L_1 signal power is 1.59 dB above the L_2 signal power at the user receiver RF input terminals.

3. For both the P-code and C/A-code the range error minimum varies only quite slowly regarding the L_1/L_2 relative power ratio. For either code a variation of the power ratio by one dB in either direction away from the optimum will increase the RMS range error by less than 2 percent, and a variation of ±2 dB will increase the RMS range error by less than 6 percent. The C/A-code signal is slightly the more sensitive to non-optimum relative power ratio.

4. If the L_1 signal is adjusted to a level 3 dB above the L_2 signal, the C/A-code RMS range error will be 3.2 percent greater, and the P-code RMS range error will be 1.4 percent greater, than if the optimum power ratio is obtained.

5. The optimum L_1/L_2 power ratios do not depend upon the absolute C/No ratio at the receiver terminals; consequently these power ratios do not depend the absolute ERP radiated by the satellite. However, the absolute RMS range error, obtained when the L_1/L_2 power ratio is adjusted to optimum, very definitely is a function of radiated ERP; this range error decreases steadily as the ERP is increased.

6. Disregarding user receiver antenna gain difference between the L_1 and L_2 frequencies, and considering only the difference in space path loss, (ie, 2.19 dB), the optimum ratios of L_1/L_2 ERP at the satellites are 4.05 dB (P-code) and 3.78 dB (C/A-code), L being the weaker signal in both cases.

11.3 Derivation of Optimum L_1/L_2 Power Ratio

It is shown in section 10 that, if σ_1 and σ_2 represent the RMS measurement errors pertaining to pseudo-range measurements at frequencies f_1 and f_2, respectively ($f_1 > f_2$), and if $p = f_1/f_2$, the residual RMS range error σ, after combining a pair of L_1/L_2 measurements to eliminate the ionospheric group delay effect, is given by the relation

$$\sigma^2 = \left(\sigma_1^2 + p^4\sigma_2^2\right)/(1-p^2)^2 \quad . \tag{22}$$

Also, in section 10 values of σ_1 and/or σ_2, as a function of C/No ratio at the user receiver input, are tabulated; both of these quantities are of the general form

$$\sigma = A/(C/No)^b \quad . \tag{25}$$

It is assumed here that some fixed amount of power P_T is available for radiation from the satellite, but that this power can be arbitrarily divided between the L_1 and L_2 signals; thus, if the L_1 signal is increased the L_2 signals must be correspondingly decreased, or conversely, such that

$$P_T = P_1 + P_2 = \text{constant}.$$

An increase in the L_1 signal from some earlier level will increase C/No at the receiver terminals; according to Eqns (23), (24), or (25) this will cause the RMS error σ_1, pertaining to the higher (L_1) frequency, to decrease. But at the same time the L_2 signal must become weaker, ie, C_2/No will decrease, and σ_2 will therefore increase.

Now, according to Eq (22), σ_1 and σ_2 are combined, but with different weighting factors, to obtain the two-frequency RMS range error. Eq (22) shows that σ_2, the lower-frequency (L_2) measurement error, is weighted considerably less heavily than is σ_1 ($p = 1232.5/1575 = 0.78254$, $p^2 = 0.61237$, $p^4 = 0.37500$). Therefore, intuition would lead to the conclusion that the L_1 signal should provide a lower RMS error than the L_2 signal, and hence that L_2 should be the weaker signal. The detailed analysis now to be presented verifies this conclusion.

and also indicates that the optimum power ratio is given by the equation

$$P_{L1} / P_{L2} = 1/p^{Q} , \tag{26a}$$

with

$$Q = 4/(2b + 1) ; \tag{26b}$$

in these expressions b = 0.65 for the P-code (Eq (23), b = 0.843 for the C/A-code (Eq (24), independent of the receiver noise No. In other words, the optimum L_1/L_2 power ratio is a function only of the frequency ratio p (because of the p^4 relative weighting factor in Eq (22)) and of the rate-of-change exponent b, in Eqs (25).

Let C_1 and C_2 represent the respective L_1 and L_2 carrier levels at the receiver input. Then, $C_1 + C_2 = C$ = constant, so that $C_2 = C - C_1$. Thus, using Eq (25) twice,

$$\sigma_1 = A / (C_1/No)^{b} , \tag{27a}$$

$$\sigma_2 = A \quad (C_2/No)^{b} = A \Big/ \left(\left[C - C_1 \right] \Big/ No \right)^{b} . \tag{27b}$$

Upon substituting Eqns (27a) and (27b) into Eq (22) there is obtained

$$\sigma^2 = \left\{ \frac{A}{(C_1/No)^{b}} \right\}^2 + p^4 \left\{ \frac{A}{\left(\frac{C - C_1}{No} \right)^{b}} \right\}^2 \tag{28}$$

$$= A^2 \, N_o^{2b} \left\{ C_1^{-2b} + p^4 \, (C - C_1)^{-2b} \right\} .$$

At a critical value of C_1, σ^2, and hence σ, will pass through a minimum. This critical value is obtained by differentiating Eq (28) partially with respect to C_1, and setting the derivative to zero. This leads to the result

$$C/C_1 = 1 + p^Q, \text{ Q given by Eq(26b).}$$

Also,

$$C_2 = C - C_1 = C\, p^Q/(1 + p^Q),$$

so th

$$C_1/C_2 = p^{-Q}.$$

But

$$C_1/C_2 = P_{L1}/P_{L2},$$

consequently Eq(26a) results.

It is now found directly from Eq (26a), for p = 0.78254, that if b = 0.65 (P-code) then P_{L1}/P_{L2} = 1.527 = +1.86 dB, and if b = 0.843 (C/A-code) then P_{L1}/P_{L2} = 0.433 = +1.59 dB.

11.4 Deviation from Optimum Power Ratio

If the L_1/L_2 power ratio is not adjusted to the optimum value given by Eq (26a) the RMS range error will not have the lowest possible value. Noting that b is different for the P-code and C/A-code signals, it is clear that a single adjustment of L_1/L_2 will not be optimum for both codes. In order to examine the consequences of non-optimum adjustment, let us introduce a parameter k such that when k = 1 the optimum ratio is obtained, and if k is above or below unity the ratio is non-optimum. Thus, let

$$C_1 = k\, C/(1 + p^Q). \tag{29}$$

and again

$$C_2 = C - C_1.$$

Then using Eq (29) in Eq (28) we obtain

$$\sigma^2 = \left[\frac{A^2 N_0^{2b} (1 + p^Q)^{2b}}{C^{2b}}\right] \left[k^{-2b} + p^4 (1 - k + p^Q)^{-2b}\right]. \tag{30}$$

For a fixed value of b, when k = 1 the minimum value of σ is obtained; this value is designated as σ_{min}. Then, from Eq (30)

$$\sigma^2_{min} = \left[\frac{A^2 (1 + p^Q)^{2b}}{(C/No)^{2b}}\right] \left[1 + p^Q\right] \quad ; \tag{31}$$

this result shows, very reasonably, that, with optimum adjustment of the L_1/L_2 power ratio, the minimum RMS range error decreases steadily as the total-carrier-power-to-noise ratio, C/No, increases.

For non-optimum adjustment substitute Eq (31) back into Eq (30), yielding

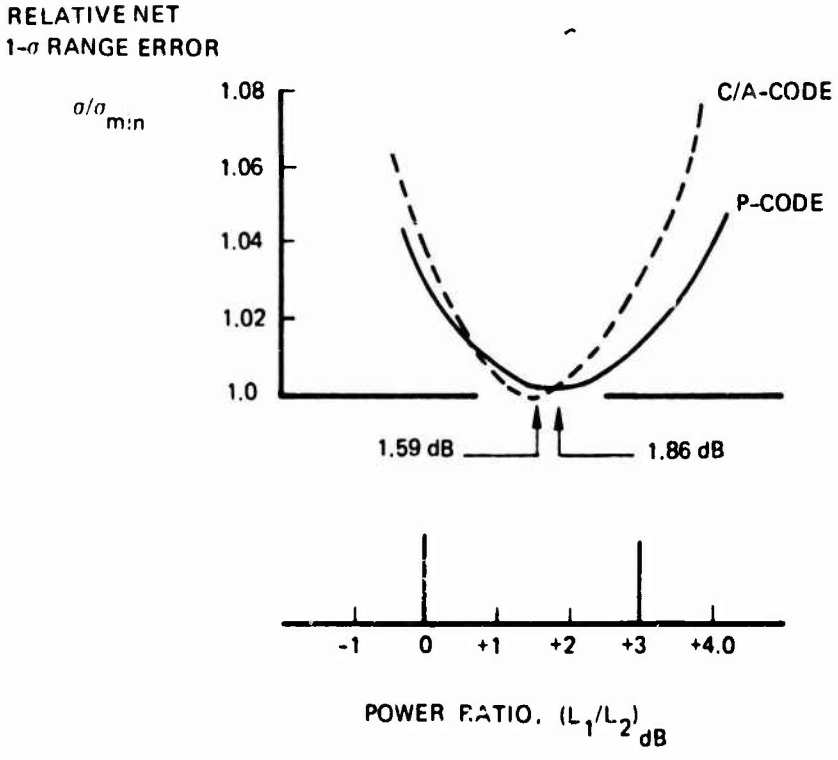
$$\sigma^2 = \sigma^2_{min} \left[k^{-2b} + P^4 (1 - k + p^Q)^{-26}\right] \Big/ \left[1 + p^Q\right] \quad . \tag{32}$$

Table 4 presents a few results from Eq (32), for k between 0.8 and 1.2, for both the P-code and C/A-code; these results are shown graphically in Figure 5. It is evident that the RMS range error, although actually passing through a minimum when the L_1/L_2 ratio is optimum, increases only rather slowly for non-optimum adjustment (See summary, paragraph 3).

TABLE 4

RELATIVE RANGE ERROR FOR NON-OPTIMUM POWER RATIOS

k	P-CODE		C/A-CODE	
	P_{L1}/P_{L2}/dB	σ/σ_{min}	P_{L1}/P_{L2}/dB	σ/σ_{min}
0.8	-0.27	1.043	-0.48	1.062
0.9	+0.78	1.011	+0.54	1.016
1.0	+1.86	1.000	+1.59	1.000
1.1	+2.98	1.012	+2.68	1.018
1.2	+4.23	1.050	+3.86	1.077



NOTE: THE ABSOLUTE 1-σ RANGE ERROR
DEPENDS UPON NET RECIEVED POWER
AND RECIEVER NOISE

Figure 5 Relative Range Error for Non-Optimum Power Ratios

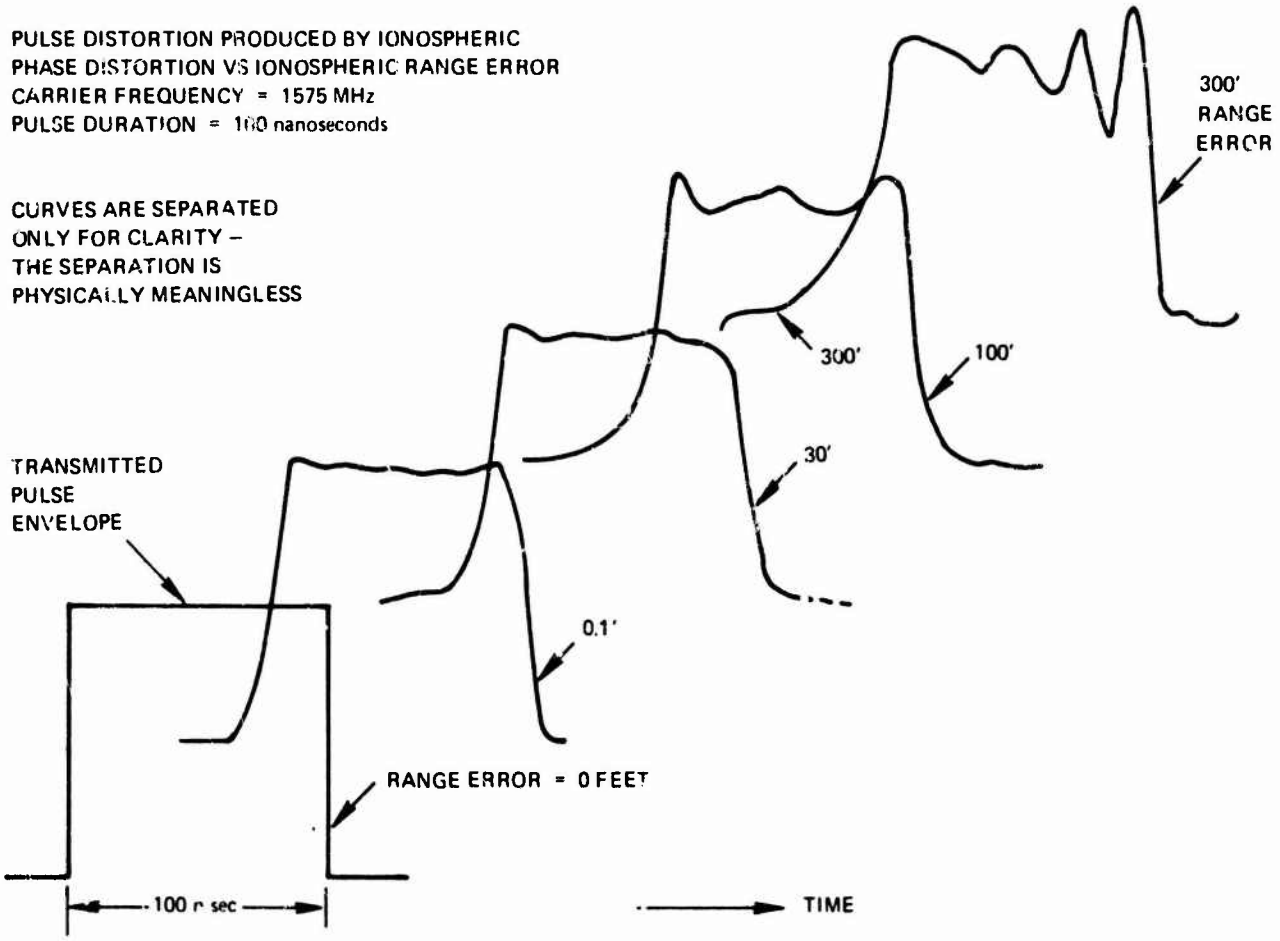


Figure 6 Distorted Pulse Waveforms

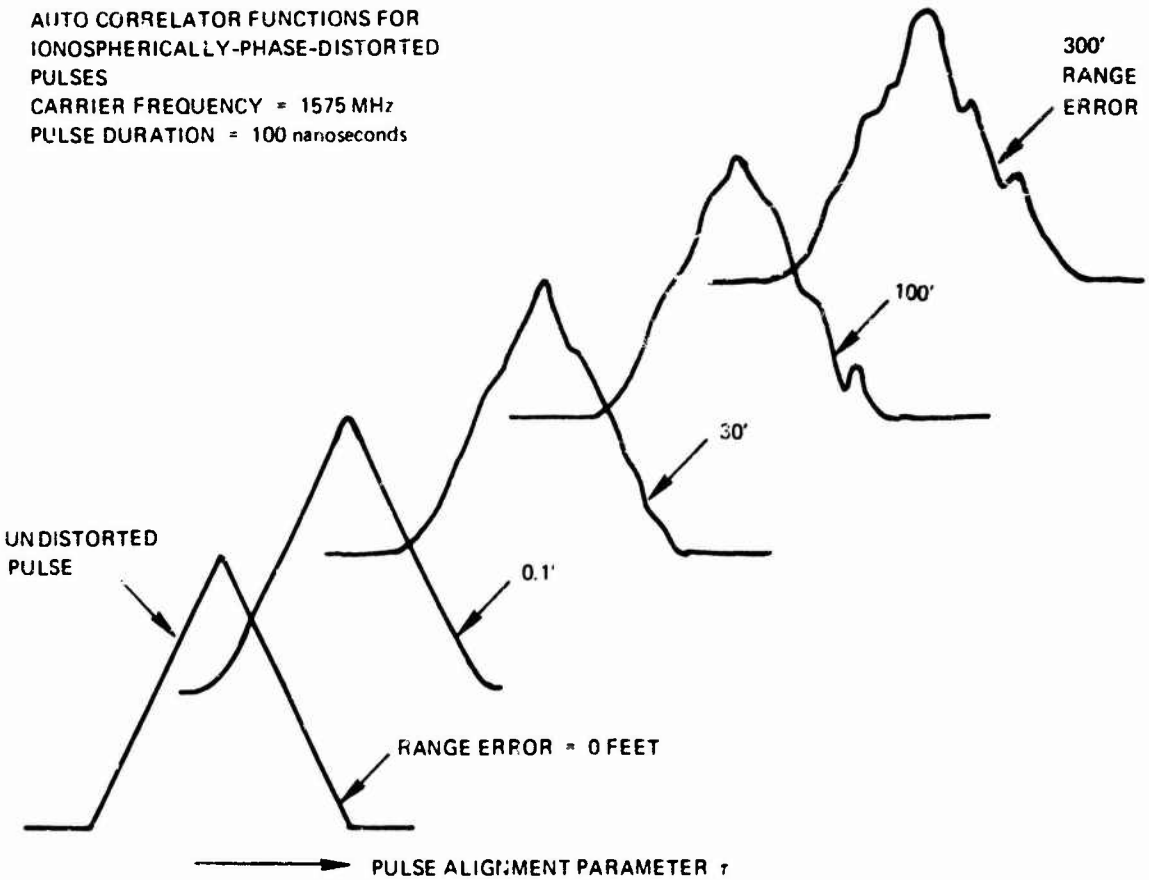


Figure 7 Autocorrelator Output Functions

12.0 PULSE DISTORTION DUE TO THE IONOSPHERE

The pulse distortion produced by the ionosphere on the 100-nanosecond, 1575 MHz pulses, has been analyzed, starting with earlier analyses performed by TRW (see DNSDP-WEW-025 for references), and adding some material discussed by J. J. Jones and W. K. S. Leong (Filter Distortion Effects on Correlation Detection and Ranging Accuracy, Tech Memo 216, WDL-IRDP-G612.HHGA, November, 1972). Limited computer analysis of the envelope distortion produced on 100-nanosecond, 1575 MHz pulses, over a rather wide range of ionospheric range error (at 1575 MHz), ie, 0.1 foot to 1000 feet, has been also been performed. From this work the following conclusions are drawn:

1. Pulse distortion is produced almost entirely because, over the relatively small bandwidth occupied by 100 nanosecond pulses centered at 1575 MHz, the ionosphere produces a non-linear frequency-sensitive phase variation (ie, phase distortion). Almost no amplitude distortion occurs.

2. The envelope distortion produced even by relatively severe ionization, ie, corresponding to a 1575 MHz range error of 100 to 300 feet, is relatively mild (see Figure 6 for computed envelopes). Some moderate rounding of the leading and trailing edges appears, together with a small degree of pulse ringing. Only for a range error well in excess of 100 feet do these distortions become severe.

3. Since auto-correlation involves integration over a considerable number of pulses, and since the integration process is relatively insensitive to pulse phase distortion (particularly insensitive to ringing), the auto-correlator output function is only slightly affected by ionosphereic range error (see Figure 7 for computed auto-correlator output functions, corresponding to the distorted pulses depicted in Figure 6).

   This conclusion is advantageous, from the point of view that successful operation of the GPS, which will utilize auto-correlation of received pulses and user-generated pulses, is virtually unaffected by even fairly severe ionospheric range error, at 1575 MHz. On the other hand, the conclusion is disadvantageous, from the point of view that virtually nothing can be "measured" concerning ionospheric range error, by manipulating the auto-correlator output.

4. The phase transfer effect of the ionosphere is fully equivalent to that of a linear, constant-amplitude filter having the following complex transfer function:

$$y(f) = \exp(j2\pi k/f)$$

with k = a constant, increasing linearly with the total integrated electron content along the satellite-user RF path, f = excitation frequency, in the vicinity of 1575 MHz.

The GPS signals occupy a narrow spectrum centered about the L1 carrier frequency, fo (= 1575 MHz). Therefore, the frequency f in the above phase function can be represented as a linear deviation from fo, that is, $f = fo + \sigma$; in this expression fo = 1575 MHz, and σ does not exceed about ±20 MHz. The phase function may be expanded into a power series in σ, yielding

$$Y(f) = \exp(j2\pi K/fo)\ (1 - \delta/f_o + \delta^2/f_o^2 + - - -).$$

The successive terms represent, respectively:

Carrier Phase Shift	$2\pi K/fo$
Group Time Delay	$2\pi K/fo^2$
Quadratic Phase Distortion	$2\pi K/fo^3$

High-order terms occur in the expansion, but are negligibly small.

Thus, the phase distortion is quadratic in σ. Now, Jones and Leong develop an elegant proof that "a purely even phase deviation produces a symmetrical correlation function" (See Jones and Leong, page 3-8, paragraph 1). Therefore, the correlation output curves shown in Figure 7 should be symmetric with respect to their peaks; the fact that these curves as computed are not quite symmetric may be because of quantization errors in computation of the curves and because of certain approximations in computation of the Fresnel integrals which unlie the correlator output function.

The Jones-Leong proof of symmetric correlator functions strongly suggests the futility of attempting to "measure" the pulse distortion by any process which involves partial pulse sampling and integration. This is particularly true when it is recognized that, in GPS user receivers, correlation occurs at IF (70 MHz), and a single "pulse" actually consists of about 7 IF cycles. (For example, C. E. Gilchriest discusses distorted pulse sampling (JPL Space Programs Summary No. 37-16, Vol. IV, 31 August 1962, pages 81-93; see especially page 89 and Figure 18).

5. If the distorted GPS pulses could be sampled point-by-point there would be some hope of "measuring" the distortion, and thereby in effect determining the ionospheric range error. On the other hand, integration over many pulses seems necessary in view of the relatively low signal-to-noise ratios characterizing the system. Integration has the effect of greatly reducing sensitivity to waveform distortion, hence makes the measuring process insensitive to ionospheric range error.

6. In a subsequent TRW analysis (P. R. Karr, Effect of Ionospheric Dispersion on UHF Pulse Transmission, Memo 7353.3-11, 19 November 1970) it is shown that IF bandwidth limiting of the DNSS ranging pulses produces more serious pulse distortion than would be expected from the ionosphere, but not enough distortion to seriously degrade the auto-correlator output. Thus, the ionospheric effect would be very difficult to separate from bandwidth-limiting effects.

13.0 RECOMMENDATIONS FOR PHASE I

It is recommended that during Phase I the ground segment should be tasked to evaluate ionospheric prediction models of varying degrees of complexity. It is only to be expected that a major tradeoff occurs among:

1. Accuracy of the prediction
2. Time duration of the prediction
3. Geographical coverage
4. The number and locations of stations acquiring raw input data
5. The advisability of using monitor-station-only (L_1/L_4) data, vs supplementary data from other types of receivers and/or other satellites (such as ATS-F, due for launch in April 1974)
6. Hardware costs
7. Data communications considerations including the transferral of raw data to a prediction model processing center, and transferral of prediction data to users
8. Software and data processing costs

During Phase I this tradeoff should be performed; the end result will be recommendations for supplying users with ionospheric data during Phases II and III.

At some risk of prejudging the success of a prediction technique, it is conjectured that, for the purpose of the highest precision navigation, no prediction will exceed the inherent accuracy of the L_1/L_2 technique; it is therefore recommended that during Phase I, and probably also thereafter, high-precision users should rely on the two-frequency correction method. On the other hand, again making a conjecture, it is probable that a practical prediction model will evolve during Phase I, adequate for the needs of medium-precision users (ie, RMS residual range errors of perhaps 30-60 feet).

The four permanent monitor stations which will be used primarily to supply ephemeris data during Phase I will provide limited, although regular and systematic, ionospheric range error data, through the use of both L and L measurements (in essence, subtract Eq (13) from Eq (14), thereby eliminating RT, and solve for the ionospheric constant K). However, because the GPS satellites from which these stations obtain L_1/L_2 measurements are constantly moving across the sky, as seen from any one of these monitors, and because both the in-view time spans and satellite look angles vs time are accurately duplicated from day to day, it is not at all certain that the L_1/L_2 data obtained from the monitors will constitute an adequate data base for updating a meaningful prediction model. Consequently it may be highly desirable, or even virtually mandatory, to provide both additional monitor station locations and monitor receiving equipment capable of acquiring ionospheric data from other than GPS satellites. A more specific recommendation, including a detailed plan for Phase I activities, will be prepared.

An ultimate objective will be a model useful for predictions at least 24 hours into the future; during Phase I an opportunity will be available to test several models with respect to accuracy, geographical coverage, prediction time span, and both formats and sizes of data flow to users. Technology does not permit such evaluations to be made at the current time; reliable raw data, gathered on a systematic basis, simply does not exist. Phase I will permit this lack to be supplied, and in the process of so doing will provide valuable data to knowledge of the ionosphere.

14.0 SELECTED BIBLIOGRAPHY

1. Radio Meteorology, Bean & Dutton, U. S. Dept. of Commerce, NBS, 1966

   Tropospheric ray eqn. derived, pgs. 84-87
   Integrals discussed, Sect. 3-3, pg 53
   More ray theory pg 357, Sect. 8.4.2 et ux
   Length of curved ray path, Sec. 8.3.2, pg. 334 et ux

2. A First-Order Prediction Model of Total Electron - Content Group Path Delay for Mid Latitude Ionosphere

   J. A. Klobachur, R. S. Allen, USAF, AFCRL-70-0403
   8 July 1970 Hanscom Field

3. Symposium on the Application of Atmospheric Studies to Satellite Transmission

4. Prediction Tech for the Effect of the Ionosphere on Pseudo-Ranging from Sync. Alt. Satellites

   V. L. Pisacane, M. M. Feen, M. Sturmanis
   Johns Hopkins SAMSO-TR-72-22
   APL/JHU TG 1197
   August 72

5. Essa Tech Report ERL-110-ITS-78 May 69
   Predicting Long-Term Operational Parameters of High Freq. Sky-Wave Telecomm System

   A. F. Barghausen
   J. W. Finney
   L. L. Proctor
   L. D. Schultz

6. Corrections for Tropospheric Range Error

   E. A. Altshuler AFCRL-71-0419
   27 July 1971 ERP #369
   Hanscom Field

7. A Model of Ionospheric Total Electron Content

   A. Ramsay
   Air Weather Service TR-234 June 1970

8. Prognostication of Ionospheric Electron Content, Vol. 1

   H. Waldman & A. V. da Rosa Aug 71
   Stanford U
   SAMSO F04701-70-C-0233
   TR SAMSO TR-71-82-Vol 1

9. Feasibility Study of Correcting for the Excess Time Delay of Transionospheric Navigational Ranging Signals

   N. N. Rao, M. Y. Youakim, K. C. Yeh
   U. of Illinois SAMSO-TR-71-163 TR #43
   14 Aug 71

10. Description and Evaluation of the Bent Ionospheric Model, Vol. 1 (1 of 3)

    SAMSO TR-72-239 Oct. 72
    R. B. Bent S. K. Llewllyn M. K. Wallock
    DBA Systems, Inc., Melbourne, Fla.
    SAMSO Contract F04701-72-C-0380

11. ESSA Technical Report ERL 107-ITS-75 May 1969

Advances in Ionospheric Mapping by Numerical Techniques

William B. Jones
Ronald P. Graham
Margo Leftin

12. Atmospheric Error in Range and Range-Rate Measurements Between a Ground Station and an Artificial Satellite

K. Takahashi
IEEE Trans on Aerospace and Electronic Systems
AES-6, No. 6 Nov 1970

13. A Survey of Ionospheric Effects Upon Earth-Space Radio Propagation

R. S. Lawrence, C. G. Little, HJA Chivers
Proc IEEE Jan 1964

This article is reproduced in its entirety, as part of this bibliography.

Appendix A

A Survey of Ionospheric Effects Upon Earth-Space Radio Propagation

A Survey of Ionospheric Effects Upon Earth-Space Radio Propagation*

R. S. LAWRENCE†, C. G. LITTLE†, MEMBER, IEEE, AND H. J. A. CHIVERS†

***Summary*—The frequency dependence is derived and the order of magnitude is presented for various ionospheric effects upon radio waves which have frequencies greater than the penetration frequency of the ionosphere. Among the phenomena considered are phase-path length change, refraction, frequency change, group-path delay, polarization rotation, and absorption. A detailed discussion is given of the mean value and variability of ionospheric absorption, refraction, scintillations, and polarization changes.**

Introduction

THIS SURVEY treats the effects of the ionosphere upon radio waves of frequency greater than the vertical-incidence ionospheric penetration frequency. Thus, our remarks apply to radio frequencies above about 10 or 20 Mc, and as high as several thousand Mc. Rationalized mks units are used throughout, and the following symbols will be employed:

c = free-space velocity of light
e = charge on an electron
$f = \omega/2\pi$ = frequency of a radio wave
H = geomagnetic field intensity
m = mass of an electron
N = number density of free electrons
ϵ_0 = electric permittivity of free space
θ = angle between wave normal and geomagnetic field
$\lambda = 2\pi c/\omega$ = free-space wavelength of a radio wave
μ = real part of the refractive index
μ_0 = magnetic permittivity of free space
ν = frequency of collision of free electrons with heavy particles
ω = angular frequency of a radio wave
$\omega_H = \mu_0 eHc/m$ = angular gyro frequency of an electron
$\omega_L = \omega_H \cos\theta$
$\omega_N = \sqrt{Ne^2/\epsilon_0 m}$ = angular plasma frequency
$\omega_T = \omega_H \sin\theta$
$X = \omega_N^2/\omega^2$
$Y = \omega_H/\omega$
$Y_L = \omega_L/\omega$
$Y_T = \omega_T/\omega$

An essentially plane radio wave may be specified in terms of five parameters: its amplitude, direction of propagation, phase, frequency [$1/(2\pi) \times$ time rate-of-change of phase], and polarization, together with their time and spatial variations. A radio wave traveling from some object in space and incident upon the earth's surface has of necessity traversed the earth's ionosphere, a region in which the refractive index may deviate significantly from unity. Particularly at frequencies below 1000 Mc the ionosphere may affect any or even all of the five basic parameters defining the radio wave. Thus the amplitude of the radio wave received at the ground may be affected by absorption, or by large-scale focusing effects or by small-scale diffraction effects; the angle of arrival may be changed by refraction or diffraction; the instantaneous frequency of the radio wave may be affected by fluctuations in the received phase; and the polarization of the radio wave may be changed by the magneto-ionic splitting and subsequent differential absorption, refraction, and phase changes. The time and space variations of these five basic parameters may be affected by ionospheric irregularities of various sizes and, for a moving source, even by a smooth, nonvarying ionosphere.

In the next two sections, we mention briefly the magnitude and frequency dependence of these effects; in later sections we consider the more important effects in greater detail.

Simplifications of the Appleton-Hartree Equation for Frequencies Large Compared with Penetration Frequency

Quasi-Longitudinal and Quasi-Transverse Propagation at VHF

At frequencies large compared with the ionospheric penetration frequency, the exact expression for the refractive index may be replaced by approximate forms as follows [1].

The first approximation, and one which is well satisfied by the normal ionosphere for frequencies greater than the critical frequency, is to equate the absorption term to zero. Under such circumstances we write

$$\mu^2 = 1 - \frac{X}{1 - \frac{1}{2}Y_T^2/(1-X) \pm \left[\frac{1}{4}Y_T^4/(1-X)^2 + Y_L^2\right]^{1/2}}. \quad (1)$$

We see that the refractive index in a given direction is double-valued, because of the + and − signs in the equation. The two values refer, respectively, to the ordinary and extraordinary waves.

At frequencies large compared with the gyro frequency, it is convenient to divide the propagation into either of two classes, namely, quasi-longitudinal and quasi-transverse propagation, depending upon which of the two terms inside the square root is the larger.

* Received June 19, 1963; revised manuscript received October 3, 1963.

† Central Radio Propagation Laboratory, National Bureau of Standards, Boulder, Colo.

For

$$\frac{Y_T^4}{(1-X)^2} \gg 4Y_L^2 \tag{2}$$

the propagation is said to be quasi-transverse (QT approximation) since Y_L may be ignored. In this case the positive sign gives a refractive index which tends toward the index for the same ionosphere in the absence of the magnetic field, while the negative sign gives a refractive index which is still dependent upon the strength of the (transverse) magnetic field.

For

$$\frac{Y_T^4}{(1-X)^2} \ll 4Y_L^2 \tag{3}$$

the propagation is said to be quasi-longitudinal (QL), since the Y_L term is dominant over the Y_T term. In this case, the refractive index for both the + and − signs is dependent upon the magnitude of the longitudinal component of the magnetic field. Note that, in the quasi-longitudinal case, the presence of the magnetic field increases the refractive index for the ordinary wave, and reduces the refractive index for the extraordinary wave, relative to the same ionosphere in the absence of the magnetic field.

At VHF, where X may be taken as very small, the conditions for QL and QT propagation become

$$\frac{4Y_L^2}{Y_T^4} \gg 1 \ (QL) \tag{4}$$

and

$$\frac{4Y_L^2}{Y_T^4} \ll 1 \ (QT). \tag{5}$$

Since $Y_L = (\omega_H/\omega)\cos\theta$ and $Y_T = (\omega_H/\omega)\sin\theta$, the transitional angle θ_c at which $4Y_L^2/Y_T^4 = 1$ is given by

$$\sin\theta_c \tan\theta_c = \frac{2\omega}{\omega_H}. \tag{6}$$

Since ω_H is of the order 10^7 radians/sec, and is therefore very much less than ω at VHF, it can be seen that $\tan\theta$ must be much greater than unity, and θ_c must be close to 90°. In other words, at VHF, propagation in all directions, except those which are almost exactly transverse to the magnetic field, will be quasi-longitudinal. The magnitude of the transitional angle is about 89° at 40 Mc. For quasi-longitudinal propagation, the characteristic polarizations are circular, and remain essentially circular as θ increases until they rather suddenly become elliptical at $\theta = \theta_c$ and then linear as the transverse direction ($\theta = 90°$) is reached.

For quasi-longitudinal propagation at VHF, (1) reduces to

$$\mu^2 \approx 1 - \frac{X}{1 \pm Y_L}. \tag{7}$$

Since both Y_L and X are very much less than unity at VHF, this can be expressed as

$$\mu \approx 1 - \tfrac{1}{2}X(1 \mp Y_L). \tag{8}$$

Approximate Expression for Ionospheric Refractive Index at VHF

Under conditions where the magneto-ionic splitting of the rays may be ignored, the refractive-index expression may be still further simplified by equating Y_L and Y_T to zero. Under such circumstances we have

$$\mu^2 = 1 - X. \tag{9}$$

At VHF, X will be small compared with unity and we may write

$$\mu \approx 1 - \tfrac{1}{2}X. \tag{10}$$

Defining $\Delta\mu$ by $\Delta\mu = (1-\mu)$ and noting that $X = \omega_N^2/\omega^2 = Ne^2/\epsilon_0 m\omega^2$, we see that

$$\Delta\mu = b\frac{N}{\omega^2} \tag{11}$$

where

$$b = \frac{e^2}{2\epsilon_0 m} = 1.6 \times 10^4 \text{ (mks)}.$$

In other words, at VHF, and ignoring the effects of the magnetic field and of absorption, the deviation of the ionospheric refractive index from unity is proportional to the electron density, and inversely proportional to the square of the operating frequency.

Change in Phase-Path Length Introduced at VHF by the Ionosphere

In order to determine the effect of the ionosphere upon the phase-path length of a VHF radio path from a satellite to the ground, we remember that the effective phase-path length P is given by

$$P = \int_0^s \mu ds \tag{12}$$

where the integral is taken along the ray path. (We have omitted a factor $\cos\alpha$, where α is the angle between the wave normal and the ray. At VHF, this factor is essentially unity.) The change Δl in the (phase)-path length introduced by the presence of the ionosphere is given by

$$\Delta l = \int_0^s (\mu - 1)dl = -\int_0^s \Delta\mu dl$$

$$= -\frac{b}{\omega^2}\int_0^s N dl \text{ meters} \tag{13}$$

and we see that Δl is proportional to the integrated electron density along the path and is inversely proportional to the square of the operating frequency. Notice

that the effect of the ionospheric electrons is to reduce the phase-path length, relative to free space.

Ionospheric Wedge Refraction at VHF

The refractive effects of the ionosphere may be considered as due to a gradient, in a direction transverse to the optical line of sight, of the radio path length from the radio source to the observer. Such a gradient could be produced by a wedge of ionization as indicated in Fig. 1. The tilting of the wave front produced by such a wedge is given by

$$\tau = \frac{d(\Delta l)}{dx} = \frac{b}{\omega^2}\frac{d\int N dl}{dx} \text{ radians.} \tag{14}$$

Note that the wedge refraction is inversely proportional to the square of the operating frequency, and is proportional to the transverse gradient of the electron content; since the refractive index in the ionosphere is less than unity, the ray is bent away from the normal upon entering.

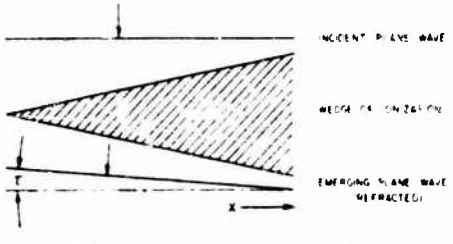


Fig. 1—The tilting of a wave front by a wedge of ionization.

Irregular Phase Changes Introduced by the Ionosphere at VHF

Irregular amplitude scintillation effects are produced whenever inhomogeneities in the ionosphere cause phase distortions comparable with one radian in (transverse) distances of the order of one Fresnel zone radius [2]–[4]. (The Fresnel zone radius on a plane wavefront at a distance R from, and traveling toward, an observer is given by $r = \sqrt{R\lambda}$. At VHF, and for normal ionospheric heights, r is of the order of 1 km.)

The phase change, δ, introduced by the ionosphere is given by

$$\delta = \frac{2\pi}{\lambda}\Delta l = \frac{-b}{c\omega}\int_0^s N dl \text{ radians} \tag{15}$$

and is inversely proportional to the operating frequency. For this reason, and because the radius of the Fresnel zone is proportional to $\lambda^{1/2}$, scintillations are most pronounced at the lower frequencies.

Frequency Changes Introduced by the Ionosphere

It has been shown above that the phase change introduced by the ionosphere is proportional to the integrated electron density along the line of sight. Any change of this integral with time (produced, for example, by the motion of the source across the sky) will result in a corresponding variation in the phase change due to the ionosphere. Such a variation with time of the received phase is, of course, a perturbation of the observed frequency.

The ionospherically induced shift of the observed frequency is therefore given by

$$\Delta f = \frac{1}{2\pi}\frac{d(\delta)}{dt} = \frac{-b}{2\pi\omega c}\frac{d\int_0^s N dl}{dt}. \tag{16}$$

and is inversely proportional to the operating frequency.

Group-Path Delay

Because the ionosphere is a dispersive medium, a pulse of radio energy centered on an angular frequency ω travels in the ionosphere with a group velocity different from the phase velocity appropriate to the angular frequency ω. The equivalent group-path length

$$P' = \int_0^s \mu' dl \tag{17}$$

where the "group refractive index" [1],

$$\mu' = d(\mu\omega)/d\omega = \mu + \omega\frac{d\mu}{d\omega}, \tag{18}$$

from which it follows that

$$P' = P + \omega\frac{dP}{d\omega}.$$

The factor $\cos\alpha$, which is negligibly different from unity, has been omitted.

At VHF, ignoring the magnetic field, we may use (10). Combining this with (18) we see that

$$\mu' \simeq \frac{1}{\mu}. \tag{19}$$

The change $\Delta l'$ in the group-path length introduced by the presence of the ionosphere is given by

$$\Delta l' = \int_0^s (\mu' - 1)dl \sim \int_0^s (1 - \mu)dl = -\Delta l \tag{20}$$

and we find that $\Delta l'$ is positive and, like Δl, is proportional to the integrated electron density along the path and is inversely proportional to the square of the operating frequency. The group-path delay is, of course, simply

$$\Delta T = \Delta l'/c. \tag{21}$$

Notice that the effect of the ionospheric electrons is to increase the group path, relative to free space.

Differential Phase-Path Length for the Ordinary and Extraordinary Waves

As already indicated, in the presence of the magnetic field the ionosphere acts as a birefringent medium and in general, splits an incident VHF radio wave into two components. These two components proceed independently through the ionosphere with different phase velocities. If both components reach the observer, there is a differential phase-path length given by

$$P_+ - P_- = \int_0^l (\mu_+ - \mu_-)dl \approx \Delta l_+ - \Delta l_- \qquad (22)$$

where the subscripts + and − denote the ordinary and extraordinary waves, respectively.

From (8), for the quasi-longitudinal case at VHF,

$$\mu_+ - \mu_- = XY_L = \frac{\omega_N^2 \cdot \omega_L}{\omega^3}. \qquad (23)$$

Hence

$$\int_0^l (\mu_+ - \mu_-)dl = \frac{2b}{\omega^3}\int_0^l N\omega_L dl \text{ meters} \qquad (24)$$

where, as before

$$b = \frac{\omega_N^2}{2N} = \frac{e^2}{2\epsilon_0 m} = 1.6 \times 10^3 \text{ (mks)}.$$

Polarization Rotation at VHF

The differential phase-path length for the two magneto-ionic components results in a rotation of the plane of polarization, as discussed in detail under "Polarization Effects," p. 21. This polarization rotation, Ω, in radians, is equal to one half of the differential phase path multiplied by $2\pi/\lambda$, *i.e.*,

$$\Omega = \frac{\pi}{\lambda}\int_0^l (\mu_+ - \mu_-)dl = \frac{b}{\omega^2}\cdot\frac{1}{c}\int_0^l N\omega_L dl \text{ radians}. \qquad (25)$$

We see that this polarization rotation is inversely proportional to the square of the operating frequency, and because of its dependence on ω_L, is directly proportional to $H\cos\theta$, the longitudinal component of the magnetic field.

Ionospheric Absorption at VHF

The absorption of radio waves in the ionosphere arises from the collisions of the electrons (which are oscillating under the influence of the incident radio wave) with the neutral and ionized constituents of the atmosphere. The absorption may be calculated from the complete Appleton-Hartree expression. For a plane wave, the electric field is

$$E = E_0 e^{-\int \kappa dl} \qquad (26)$$

where E_0 is the amplitude of the incident wave and, for quasi-longitudinal propagation [1],

$$\kappa = \frac{e^2}{2mc\epsilon_0}\cdot\frac{1}{\mu}\cdot\frac{N\nu}{\nu^2 + (\omega \pm \omega_L)^2}.$$

Here κ is the exponential decay coefficient for the *amplitude* of the radio wave in unit distance. At VHF (and indeed at HF), it is customary to assume that $\omega \pm \omega_L \gg \nu$, and if $\mu \approx 1$ the absorption in decibels is given by

$$A = -20\log_{10}\frac{E}{E_0} \propto \frac{1}{(\omega \pm \omega_L)^2}\cdot\int_0^l N\nu dl. \qquad (27)$$

At VHF, $\omega \gg \omega_L$, so the absorption of VHF radio waves in the ionosphere is therefore inversely proportional to the square of the operating frequency.

Differential Ionospheric Absorption

Eq. (27) shows that the absorption of radio waves in the ionosphere is different for the two characteristic modes. This differential absorption is given by

$$A_- - A_+ = -20\log_{10} \cdot \exp\left\{-\frac{e^2}{2mc\epsilon_0}\cdot\frac{1}{\mu}\left[\frac{1}{(\omega-\omega_L)^2} - \frac{1}{(\omega+\omega_L)^2}\right]\int N\nu dl\right\}. \qquad (28)$$

At VHF, $\omega \gg \omega_L$, and it follows from (28) that

$$A_- - A_+ \propto \frac{\omega_L}{\omega^3}. \qquad (29)$$

We see that this differential absorption, which would be accompanied by changes in the polarization of the resultant radio wave, decreases very rapidly with frequency, being inversely proportional to the cube of the observing frequency. Like the polarization rotation, its dependence upon ω_L makes the differential absorption proportional to $H\cos\theta$, the longitudinal component of the magnetic field.

Order of Magnitude of Ionospheric Effects upon Earth-Space Communication

The conclusions of the preceding section are summarized in Table I, which shows the frequency dependence of the various ionospheric propagation effects at VHF and higher frequencies. It should be noted that the expressions are based on approximations to the Appleton-Hartree equations and may not be valid in the quasi-transverse case.

Table II has been prepared to give some feeling for the magnitude of the ionospheric effects at a frequency of 100 Mc. The values are based on the total ionospheric contribution for zenithal quasi-longitudinal propagation through an ionosphere where $\int_0^\infty N dh$ is $10^{13}/\text{m}^2$ column, and $\omega_L \approx 5 \cdot 10^6$ radians/sec. The wedge refrac-

TABLE I

FREQUENCY DEPENDENCE OF IONOSPHERIC PROPAGATION EFFECTS

Phase-path length change (negative)	$\Delta l \propto 1/\omega^2$
Refraction	$\tau \propto 1/\omega^2$
Phase change (negative)	$\delta \propto 1/\omega$
Frequency change (negative)	$\Delta f \propto 1/\omega$
Group-path delay	$\Delta T \propto 1/\omega^2$
Differential phase-path length	$\Delta l_+ - \Delta l_- \propto 1/\omega^3$
Polarization rotation	$\Omega \propto 1/\omega^2$
Absorption (db)	$A \propto 1/\omega^2$
Differential absorption	$A_+ - A_- \propto 1/\omega^4$

TABLE II

ORDER OF MAGNITUDE OF IONOSPHERIC PROPAGATION EFFECTS AT 100 Mc

(See text for detailed propagation conditions)

Phase-path length change		$\Delta l = -400$ meters
Wedge refraction		$\tau = 4 \times 10^{-3}$ radians
Phase change		$\delta = -840$ radians
Frequency change		$\Delta f = -6$ cps
Group-path delay		$\Delta T = 1.3 \times 10^{-6}$ sec
Differential phase-path length		$\Delta l_+ - \Delta l_- = 6.4$ meters
Polarization rotation		$\Omega = 6.6$ radians
Absorption	A (night)	$= 0.005$ db
	A (day)	$= 0.05$ db
	A (daytime PCA*)	$= 1$ db
Differential absorption	$A_+ - A_-$ (night)	$= 5 \times 10^{-5}$ db
	$A_+ - A_-$ (day)	$= 5 \times 10^{-4}$ db
	$A_+ - A_-$ (daytime PCA)	$= 1 \times 10^{-2}$ db

* Polar-cap absorption event.

tion is based upon a transverse gradient in $\int_0 N dh$ of 10^{11} electrons/m^2 column per meter, *i.e.*, 1 per cent per 10 km; the value of Δl is based upon a satellite observed near the horizon in a typical orbit just above the ionosphere. The value of $\int_0 N dh$ of 10^{17} m^{-2} column is believed to be typical of an average ionosphere, and could vary by at least a factor of two, in each direction, *i.e.*, from 10^{16} to 10^{18} m^{-2}, depending upon the time of day, the season, the position in the sunspot cycle, and the geographical location. For ionospheric conditions other than those used in Table II, the proportional changes in $\int_0 N dl$ and ω_L should be used to estimate the ionospheric effects at 100 Mc; for VHF and UHF frequencies other than 100 Mc, the frequency dependences of Table I should then be used.

The absorption values are order-of-magnitude estimates of the one-way absorption, at vertical incidence, of a 100-Mc radio wave during the night, during the day, and during a strong daytime polar-cap absorption event.

Ionospheric Absorption at VHF

Quantitative Theory

Magneto-ionic theory shows that, for quasi-longitudinal propagation at VHF, the absorption coefficient κ, in nepers per unit length is given with good accuracy by

$$\kappa = \frac{e^2}{2mc\epsilon_0} \frac{1}{\mu} \frac{N\nu}{\nu^2 + (\omega \pm \omega_L)^2} \tag{30}$$

where ν is the frequency of collisions between an electron and heavy particles. At VHF, μ will be approximately unity in the normal ionosphere, and we may omit the $1/\mu$ factor. This omission of the refractive-index dependence of absorption is an indication of the fact that, at VHF, the ionospheric absorption occurs in a nondeviating medium. Such absorption is commonly termed "nondeviative."

The exponential decay experienced by a VHF radio wave traveling from the satellite to the ground is therefore given by

$$\int_0^s \kappa dl = \frac{e^2}{2mc\epsilon_0} \int_0^s \frac{N\nu}{\nu^2 + (\omega \pm \omega_L)^2} dl. \tag{31}$$

Numerically, the absorption in decibels is given by

$$A = 4.6 \times 10^{-5} \int_0^s \frac{N\nu dl}{\nu^2 + (\omega \pm \omega_L)^2}. \tag{32}$$

Since the absorption of VHF normally occurs in a region where ω is large compared with ν, and $\omega \gg \omega_L$, the absorption may be written as

$$A = \frac{4.6 \times 10^{-5}}{\omega^2} \int_0^s N\nu dl. \tag{33}$$

and is inversely proportional to the square of the operating frequency.

Regular D-Region Absorption

During daytime hours, much of the absorption of VHF radio waves occurs in the D layer. This layer is primarily produced by solar ultraviolet radiation, and therefore has a pronounced diurnal and seasonal variation. As shown in Table II, the regular D-region absorption at 100 Mc is less than 1 db, even after allowing for the effect of oblique incidence; hence ionospheric absorption in the normal D region is negligible for frequencies above about 100 Mc. This regular D-region absorption has been found to obey the $1/(\omega \pm \omega_L)^2$ relation down to frequencies of a few Mc. At 30 Mc, therefore, [illegible] science transmissions through the normal daytime ionosphere could be expected to undergo up to several decibels of attenuation.

Solar Flare Effects on D-Region Absorption

During the phenomenon known as a solar flare, the output of ultraviolet and X-ray photons by the sun can increase very markedly. The resultant enhancements in the ionizing flux incident upon the D region over the sunlit hemisphere can produce large increases in the D-region ionization, and hence large changes in the absorption of radio waves penetrating the D region. Fig. 2 shows the absorption of 10-Mc cosmic radio noise at College, Alaska, during the large solar flare of July 11, 1961.

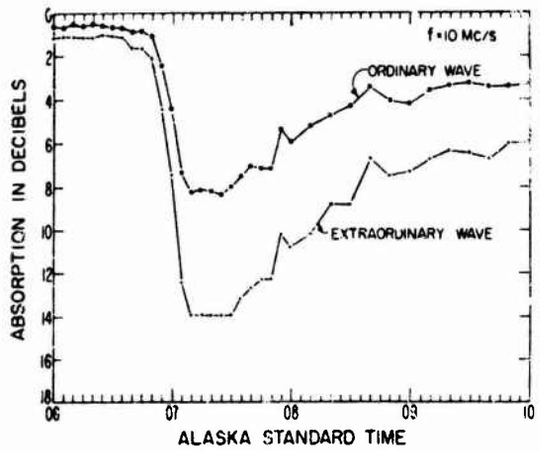


Fig. 2—Solar flare absorption event, College, Alaska, July 11, 1961.

The two curves refer to the ordinary and extraordinary waves, respectively; it will be seen that at this relatively low frequency the differential absorption effect is large. The ratio of the differential absorption to the ordinary-wave absorption is essentially that predicted for absorption in a region where $\omega \gg \nu$; we may therefore compute the 100-Mc absorption with confidence using a $1/(\omega \pm \omega_L)^2$ dependence. This gives a maximum zenithal one-way absorption of 1.1 db at 100 Mc for this flare, *i.e.*, of the order 20 times greater than the normal *D*-region daytime absorption.

Solar flare absorption events may be expected to be at a maximum at the subsolar point, and are not detected on the night side of the earth. The flares themselves occur most frequently during sunspot maximum and typically have durations of several minutes (for the smaller flares) to a few hours (for the most intense flares).

Auroral Absorption Events

Anomalously high *D*-region ionization densities may occur during an aurora, apparently due to ionization by bremsstrahlung X rays generated in the upper atmosphere by the bombarding electrons producing the the aurora [5], [6]. These aurorally associated absorption events occur primarily at the auroral zones, and at higher magnetic latitudes. These events are closely associated with luminous aurora. They are seen with decreasing frequency as one moves to lower geomagnetic latitudes.

Typical values of absorption are of the order 3 db at 30 Mc at vertical incidence, though on rare occasions the aurorally associated absorption reaches about 10 db at 30 Mc. Auroral absorption events are seen almost daily at the auroral zone during equinoctial months, and are closely related to the local geomagnetic *K* index. They may occur at essentially any hour of the day, though statistically there is a pronounced minimum in the occurrence probability for the three hours centered on 21 hours local time. The auroral absorption events typically have periods of several minutes to a few hours. The detailed correlation between absorption values is usually poor for observations made at separations of more than a few hundred kilometers.

For frequencies greater than 30 Mc, the aurorally associated absorption is believed to obey the $1/(\omega \pm \omega_L)^2$ relation quite accurately. However, there is evidence that this approximation begins to break down for auroral absorption events at about 15 Mc, and is in serious error at frequencies as low as 5 Mc, the observed absorption being considerably *less* than that predicted. This fact indicates that the absorption takes place mainly at heights where $\nu/2\pi \geq 5$ Mc.

Polar-Cap Absorption Events

On relatively rare occasions, a solar flare can result some hours later in a widespread, long-lasting absorption event which is detectable over both polar caps down to the equatorial limits of the respective auroral zones. Such events occur most frequently during the peak of the sunspot cycle, and may last for up to four days [7]. The absorption is greater during the daylight hours than at night by a factor of about five. While Table II shows a strong PCA event giving a vertical-incidence absorption of about 1 db at 100 Mc, the occasional occurrence of 100-Mc absorption up to 3 db was probably produced by the strongest PCA's seen during the recent sunspot maximum. The frequency dependence of this characteristic polar-cap absorption is approximately (but not exactly) $1/(\omega \pm \omega_L)^2$ for frequencies above 30 Mc; very little is known about the frequency dependence below 30 Mc.

The polar-cap absorption events are due to ionization in the lower *D* region produced by a flux of energetic protons (low-energy cosmic rays) of solar origin. Fig. 3 illustrates the zenithal absorption of 30-Mc cosmic noise at College, Alaska, during the outstanding polar-cap absorption event of February 23, 1956.

F-Region Absorption

Several of the observers who have reported measurements of the total attenuation of cosmic radio waves at various magnetic latitudes have attempted to separate the contributions due to the *D* and *F* regions [8]–[11].

Although the conclusions reached by these workers conflict in detail, it is apparent that, at least during years of high sunspot number, the absorption in the undisturbed *F* region can be comparable with the absorption in the undisturbed daytime *D* region. The median *F*-region absorption when transformed to 100 Mc by a $1/(\omega \pm \omega_L)^2$ scaling factor, is usually within a factor of 3 of 0.1 db at vertical incidence, and is therefore hardly of significance to space communications at frequencies above about 100 Mc. Because of the great height of the *F* region, the correction for obliquity does not exceed about a factor of two, even at large zenith angles. The *F*-region absorption is believed to be much

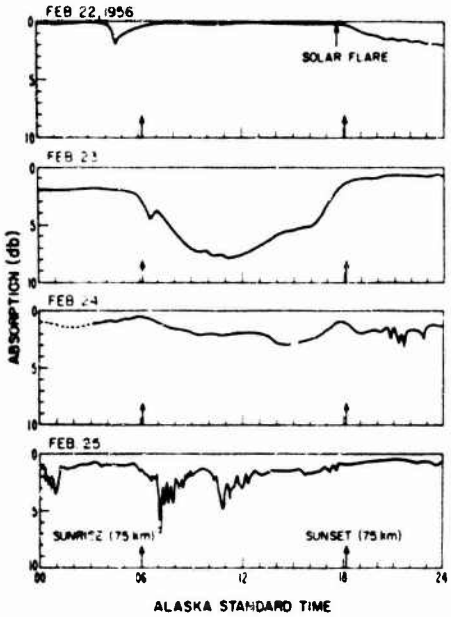


Fig. 3—30-Mc cosmic noise absorption measurements during the February, 1956 polar-cap absorption event.

less affected by solar flares, aurora, and polar-cap absorption events than the D-region absorption.

At frequencies below about 30 Mc, deviative absorption may also be observed, since at times μ may be significantly less than unity. Under such circumstances, we see from (30) that the absorption at the lower frequencies will *increase* more rapidly than is predicted by the $1/(\omega \pm \omega_L)^2$ law.

Ionospheric Refraction

Ray Tracing in Anisotropic Media

Insofar as the usual assumptions of geometrical optics are satisfied, the refractive index given by the Appleton-Hartree equation may be used for ray tracing through the ionosphere. However, when the effect of the geomagnetic field is included, the resulting anisotropic properties of the medium introduce several complications. Just as in the optics of crystals, the wave normal does not coincide with the direction of the ray. The reason for this becomes apparent if, with reference to Fig. 4, we consider an angular spectrum of monochromatic plane waves, all in phase at point O, traveling in the general direction OB. Because of the anisotropy of the refractive index, the phase fronts at some later time are given by the lines AD, BD, and CD where the points A, B, and C are at different distances from O. The various components interfere constructively only at the point D. Thus, while the wave normals lie in the general direction OB, the energy flow (and therefore the "ray") lies in the direction OD.

There is another way in which the anisotropy of the medium contributes to the complexity of ray tracing. Suppose the ray arrives at the plane boundary separating one homogeneous isotropic medium from another

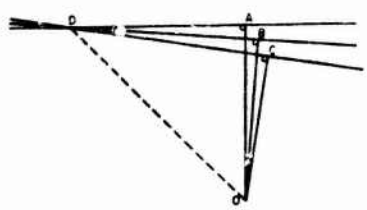


Fig. 4—A sketch illustrating the difference between the direction of the wave normal and that of the ray in an anisotropic medium.

If the media were isotropic, one should simply calculate the refractive indices of the two media and, knowing the angle of incidence (the angle between the phase front and the boundary), would use Snell's law to determine the angle of refraction. In the anisotropic case, however, the refractive index cannot be determined until the angle of refraction is known, but that angle in turn depends upon the refractive index. The values of these two related quantities must be obtained by a process of iteration, successively improving a first trial value of the refractive index.

Even in the ideal case of a plane or spherically stratified ionosphere, the anisotropy caused by the geomagnetic field may cause the ray to deviate from the plane of incidence. Thus, for a complete picture of the ray path through the ionosphere, it is necessary to use a three-dimensional coordinate system for the ray-tracing process. The various complications of ray tracing have been discussed in detail by Bremmer [12] and by Booker [13].

The Geometry of Ionospheric Refraction

Elevated Refracting Layer: As we have seen, the ionospheric refractive index is less than unity, is frequency dependent, and is anisotropic. The effective ionosphere exists only within a definite range of heights outside of which the refractive index is unity. The simple statement of Snell's law of refraction, $\mu \sin i = \text{constant}$, applies if we are willing to represent the ionosphere crudely by an elevated, plane-stratified slab. In this case, as shown in Fig. 5, the ray is deviated within the ionosphere but its direction above the ionosphere is the same as that below. There is bending within the layer but the "total bending," *i.e.*, the difference in direction between the ray at the source and at the observer, is zero. (In Fig. 5, as in those which follow, the "observer" is shown at the top of the effective troposphere or, what is equivalent, the refraction of the troposphere is neglected.)

Although the ionosphere in this case causes no total bending of the ray, the apparent and true directions of the source from the observer do not agree. The difference, β, is called the "angular error at the observer" and is equal, in this case, to $-\beta'$, the "angular error at the source." The effect of the plane-slab ionosphere is to increase the apparent elevation angle of the source.

We have already mentioned the simple form of

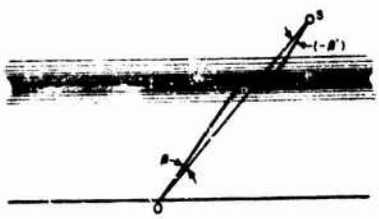


Fig. 5—Bending in a plane-stratified ionosphere.

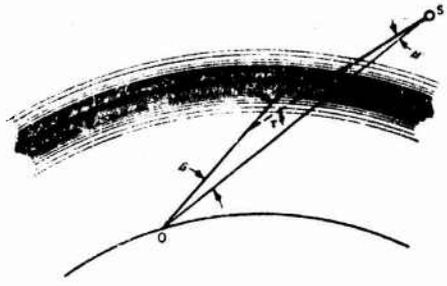


Fig. 6—Bending in a spherically-stratified ionosphere.

Snell's law which applies for the case of a plane-stratified ionosphere, it may be written

$$\mu \sin i = \sin i_0,$$

where μ is the refractive index in the ionsphere, i is the local zenith angle of the ray (we neglect the effect of the magnetic field), and i_0 is the zenith angle of the ray at the ground. From this relationship it is clear that a ray will penetrate the ionosphere if and only if $\mu \geq \sin i_0$ at all ionospheric heights. In particular, if we denote by μ_m the refractive index at the height of maximum electron density, all rays outside the cone $i_0 = \sin^{-1}\mu_m$ will be reflected back to the earth. This blocking effect of the ionosphere for zenith angles greater than the critical value is sometimes called the "iris" effect.

The effects which we have mentioned for a plane-stratified ionosphere are modified slightly by the curvature of the ionosphere (sometimes called "curvature of the earth"). Fig. 6 shows the somewhat more realistic situation of a spherically stratified ionsosphere. Here the total bending, $\tau = \beta + \beta'$, is no longer zero. The contrast between Figs. 5 and 6 shows that the total bending which occurs in a ray passing entirely through a spherically stratified ionosphere is a second-order effect caused only by the curvature of the ionosphere, since the bending of a down-coming ray on entering the medium is compensated for (actually, overcompensated for) by the bending on leaving it. With the exception of the special case of radio astronomy where the distance from the observer to the source is essentially infinite, the angular error at the observer generally differs from, and may be greater or less than, the total bending. Examples of each case will be evident later when we compare Figs. 9 and 13.

It is well known that the refraction which occurs in the troposphere increases very rapidly with zenith angle as the zenith angle approaches 90°. Ionospheric refraction behaves in a quite different manner. Because the elevation of the ionosphere is a sizeable fraction of the radius of the earth, it is impossible for a ray which originates on the surface of the earth to arrive at the ionosphere with an angle of incidence of 90°. Accordingly, ionospheric refraction increases relatively slowly as the ray approaches the observer's horizon and, in the last four or five degrees, becomes relatively independent of elevation angle.

The iris effect which we have mentioned in connection with a plane-stratified ionosphere becomes considerably more complicated. The cone of rays which pass through the edge of the iris leaves the ionosphere with a cone angle greater than that subtended by the iris. Thus, the portion of the actual sky which is visible from the ground exceeds the size of the iris, the apparent position of each point having been moved toward the zenith by ionospheric refraction. The "mapping" of this extra part of the sky into the ionospheric iris reduces the intensity of the radio wave received from a point source. This divergence effect has been treated in detail by Bremmer [12]. For an antenna with a narrow beam which lies entirely within the iris, the apparent brightness temperature of the background sky is unaffected by the iris, and thus the signal-to-noise ratio available from a point source in the antenna beam is decreased by the presence of the ionosphere.

Anisotropic Medium: The primary effect of the anisotropy which the geomagnetic field introduces into the refractivity of the ionosphere is the splitting of a radio wave into two characteristic modes called the ordinary and extraordinary waves. Their polarization is of opposite sense and, for most directions of propagation, practically circular.

The two rays are deviated in opposite directions from the plane of incidence but, for moderate or large zenith angles, the deviation from the vertical plane is ten to one hundred times smaller than the vertical deviation from the line of sight. Thus, we get a good approximation to the actual ray path if we ignore this deviation.

Figs. 7 and 8 show examples, projected on the vertical plane and not drawn to scale, of calculated ray paths emanating from a 20-Mc satellite-borne transmitter. In this case, the observer is located in the northern hemisphere (Boulder, Colo.; geomagnetic latitude 49°) and the satellite is in the northern (Fig. 7) or southern (Fig. 8) sky at a height which is well above the maximum electron density of the ionosphere.

The Appleton-Hartree equation shows that, for frequencies as high as 20 Mc, the refractive index for the ordinary ray is always closer to unity than the refractive index for the extraordinary ray. This fact has frequently led to the erroneous conclusion that the "extraordinary satellite" always appears higher in the sky than the "ordinary satellite." Fig. 7 shows that this is not the case. The proper statement is that the total bending along the extraordinary ray always exceeds

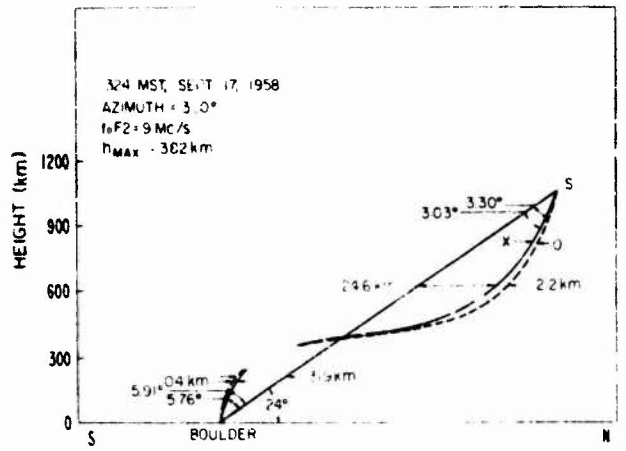


Fig. 7—An example, projected on the vertical plane and not drawn to scale, of calculated ray paths from a 20-Mc satellite-borne transmitter in the northern sky.

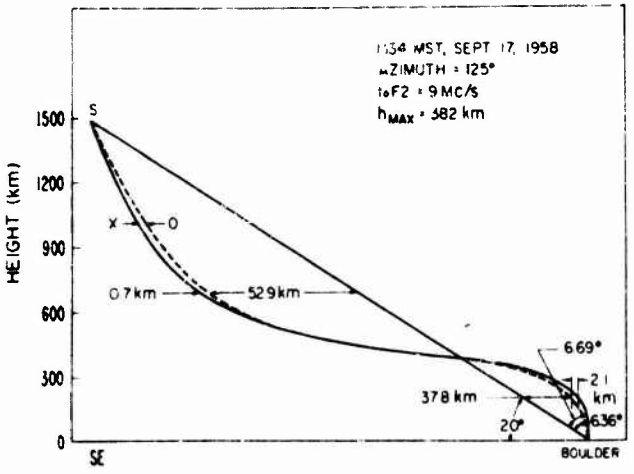


Fig. 8—An example, projected on the vertical plane and not drawn to scale, of calculated ray paths from a 20-Mc satellite-borne transmitter in the southern sky.

the total bending along the ordinary ray. In Fig. 7 the excess is 0.12°.

Approximations to the Refraction in Spherically Stratified Model Ionospheres

Complete ray tracing through an abitrary model ionosphere is a complicated process and it is fortunate that there are many practical cases for which it is not necessary. The most common useful approximations include neglecting the magnetic-field effects and making the assumption of a spherically stratified ionosphere.

Under these conditions, Bailey [14] has derived the following approximate expression for the total bending, τ, in a model of the ionosphere for which the height profile of electron density is parabolic:

$$\tau = \frac{d + d'}{3a}\left(\frac{f_0}{f}\right)^2\left(1 + \frac{h_m}{a}\right)\left(\cos^2\alpha + \frac{2h_m}{a}\right)^{-1/2}\sin\alpha. \quad (34)$$

Here, d and d' are the thicknesses, not necessarily the same, of the two half parabolas which comprise the ionospheric layer, f_0 is the critical frequency of the layer, f is the wave frequency, h_m is the height of the level of maximum electron density, a is the radius of the earth, and α is the apparent zenith angle of the ray at the ground. Since it is only an approximation, this formula should be used only for frequencies which are at least twice the MUF (maximum frequency which would be reflected from the layer at the given value of α). We now know that the upper part of the electron-density profile of the ionosphere differs markedly from a parabolic form, and that Bailey's formula tends to give refraction values which are too small. Nevertheless, it provides a convenient and useful method of estimating total bending.

Even when the magnetic field is neglected, the analytic solution becomes impractically complex for an ionospheric layer of more realistic shape, such as the Chapman model. However, Weisbrod and Anderson [15] have derived a convenient approximate procedure which is applicable to an arbitrary ionospheric profile. Their method is even more general, since they include a process for estimating the angular error of a source within the ionosphere. Weisbrod and Colin [16] show a numerical example of the method. Carru, Gendrin, and Reyssat [17] have given several examples of ray-tracing calculations of total bending and of angular error for sources at finite distances. Figs. 9 and 10 show their results for 20-Mc signals passing through an ionosphere of 12.7-Mc critical frequency, at apparent zenith angles of less than 52°. The critical apparent zenith angle for this particular ionosphere model is 54°. Counter [18] has used a ray-tracing computer program to produce a large collection of graphs which show the angular error to be expected under a variety of geometrical conditions.

Wedge Refraction

The work which we have just been discussing is limited to consideration of spherical refraction, *i.e.*, to cases where the ionosphere is spherically stratified. Although the actual ionosphere usually differs significantly from this simple model, very little serious investigation has been made of the effects of horizontal gradients and ionospheric irregularities. Smith [19] encountered the problem of such irregular refraction when he was making careful measurements of the absolute positions of radio stars. He derived (14), which gives the relationship between the amount of "wedge refraction," τ, near the zenith and the horizontal gradient of the total electron content of the ionosphere. Komesaroff [20] gives a more detailed discussion of wedge refraction at high elevation angles. Lowen [21] has derived some relatively simple integrals, involving the electron density profile and its horizontal gradients, which approximate the spherical and wedge components of radio-astronomical refraction.

8

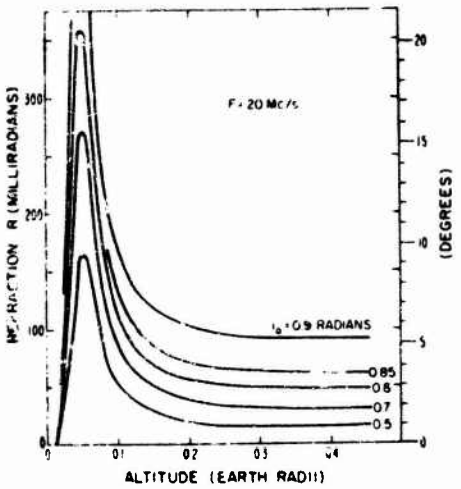


Fig. 9—The total bending (R), given as a function of source altitude, of a 20-Mc radio ray passing through an ionosphere of critical frequency 12.7 Mc. The parameter (i_0) is the apparent zenith angle at the ground. (From Carru, Gendrin, and Reyssat [17].)

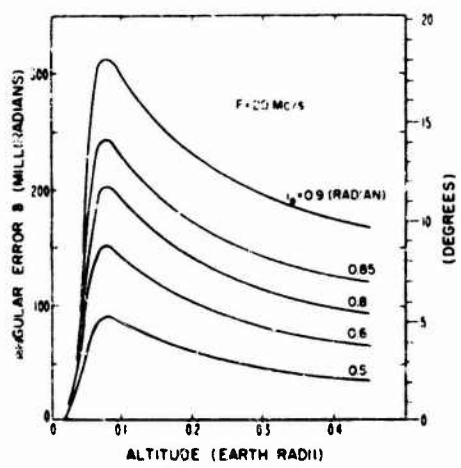


Fig. 10—The angular error (δ), at the ground, given as a function of source altitude, of a 20-Mc radio ray passing through an ionosphere of critical frequency 12.7 Mc. The parameter (i_0) is the apparent zenith angle at the ground. (From Carru, Gendrin, and Reyssat [17].)

The literature of radio astronomy contains several references (mentioned below) to observations of irregular ionospheric refraction and of the more regular form of wedge refraction which results from the normal diurnal variation in total electron content of the ionosphere. It was this latter form of wedge refraction which was studied quantitatively by Smith [19].

From measurements of variations in the apparent position of radio stars at 81.5 Mc, Smith was able to deduce that the total electron content of the ionosphere is two or three times greater than the electron content below the level of maximum density. He alluded, in passing, to irregular effects with time durations of twenty minutes or less, but he made no quantitative study of them. More recently, Komesaroff [20], observing at the much lower frequency of 19.7 Mc, has also been able to deduce total electron content from refraction measurements and, in addition, to observe the irregular component of refraction.

Vitkevich [22], Vitkevich and Kokurin [23], and Wild, Sheridan, and Neylan [24] have all reported refractive effects in the observed positions of the sun and of radio stars. These they have attributed to ionospheric irregularities having horizontal dimensions of the order of 200 km, although little can be deduced regarding their shape or orientation.

Fig. 11 shows an extreme example of the irregular motion of the radio star Cygnus-A observed at 108 Mc at Boulder, Colo. Such irregularities are observed at Boulder on about one third of all days [25].

The derivations of wedge refraction given by Smith, by Komesaroff, and by Lowen [21] are restricted to the high elevation angles which are generally of interest to radio astronomers. The rate at which wedge refraction might be expected to increase near the horizon is a more difficult problem, and few pertinent observations exist. The calculations summarized in Fig. 12 are a first step in this direction. The refraction shown in this figure is that which would be imposed upon a 100-Mc radio wave by thin horizontal prisms at various heights. The refracting surfaces of the prisms are taken to intersect at an angle of 5°. The vertex, the line of intersection of the refracting faces, is assumed to lie in a horizontal plane and to be perpendicular to the line of sight. One such prism is assumed to be in the line of sight for each elevation-angle computation. The prisms are assumed to have a plasma frequency of 12 Mc. Within 35° of the zenith the refraction varies slowly, approximately as csc H, the cosecant of the elevation angle. It then increases 1.5 to 2 times as fast as csc H, down to within 10° of the horizon. Below 10° elevation angle the effect of the prism, like all ionospheric refraction effects, varies relatively slowly down to the horizon. The zenith-angle effect shown here is for very thin prisms and so represents an extreme case. As the irregularities become thicker, approaching a spherical shape, the elevation-angle variation of their refracting effect will disappear.

Range and Doppler Effects

In discussing the refractive effects of the ionosphere we have so far been concerned only with the direction of arrival of the radio wave. The ionosphere also affects the "radar range" and the observed Doppler shift of an object which lies in or above it.

The deviation in radar range, being the change in group path length, can be evaluated for any specific case by integrating the quantity $(\mu' - 1)$ along the ray path. For a frequency of 100 Mc, Counter [18] has calculated range deviations for a large number of special cases. At frequencies above about 40 Mc, the range deviation varies as the square of the wavelength, so Counter's calculations can be applied easily. For lower frequencies, there is no substitute for applying a detailed ray-tracing treatment to the specific problem at hand.

The Doppler shift of observed frequency is, of course,

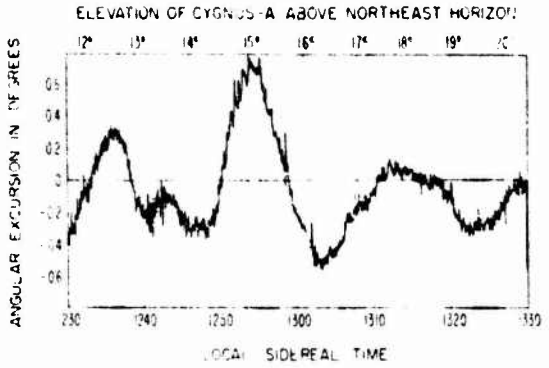


Fig. 11—A one-hour record showing the effects of large ionospheric irregularities upon the apparent position, at 108 Mc, of the radio source Cygnus-A relative to its mean apparent track.

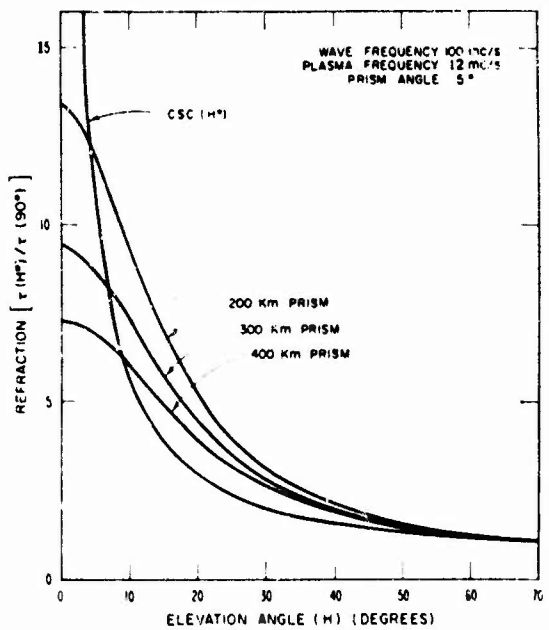


Fig. 12—A calculation of the rate at which wedge refraction, caused by thin ionospheric prisms at heights of 200–400 km, increases toward the horizon. The value of refraction at the zenith, r(90°), is 0.036°.

proportional to the rate of change of phase path. If there were no ionosphere, the Doppler shift in the frequency of a radio signal transmitted from an arbitrary height, and received on the ground, would be determined solely by the component of the velocity of the transmitter in the direction of the straight line between the transmitter and the receiver. The presence of the ionosphere modifies this Doppler shift, first because refraction causes the ray to leave the transmitter in a different direction from that of the straight-line path, and secondly because the ionosphere adjacent to the transmitter makes the phase velocity there differ from the free-space value. Hibberd [26] and Weekes [27], using slightly different lines of approach, have each derived expressions for the ionospheric effect upon the Doppler shift of the signal from an artificial earth satellite. Of necessity, the useful forms of their results have been limited to the highly simplified case of circular orbit, spherically stratified ionosphere, and no magnetic field. Ross [28] considers a variety of corrections which must be applied in order to approximate an actual situation. He gives particular attention to the "dispersive Doppler shift," i.e., to the difference in relative Doppler shift at two different radio frequencies. While the ionospheric contribution to the Doppler shift is only a second-order effect, the ionosphere is solely responsible for the dispersive Doppler shift and so this quantity which is directly observable gives a first-order measure of the rate of change of total number of electrons along the ray from the satellite to the observer.

A related technique for measuring the total electron content of the ionosphere, the exosphere, and interplanetary space has been discussed by Eshleman, *et al.* [29]. Variations of their technique involve measuring the relative phase difference of the modulation envelopes of two identically modulated carrier frequencies or measuring the relative phase of the various sidebands of a modulated carrier

Fig. 13 illustrates, with a particular example, the typical magnitude of the ionospheric effect upon Doppler shift. It is calculated for the 20-Mc signal from an artificial satellite and gives a comparison between the correct result obtained by actual ray tracing and an approximation which involves simply integrating the refractive index along the straight line joining the satellite and the observer.

The Mean Value and Variability of Ionospheric Refraction

The best readily available single indicator of the spherical component of ionospheric refraction is the critical frequency, f_0F2, of the F region. The F region varies diurnally and seasonally; it varies with latitude and with the sunspot cycle. World-wide observations of f_0F2 have been available for many years, and the general behavior of its regular variations is summarized in a set of world maps by Zacharisen [30]. In addition to the regular variations shown on the maps, the ionosphere undergoes day-to-day changes of a factor of two or more, and the world maps must not be considered an adequate substitute for actual observations.

Considering the variability of f_0F2 and the fact that fully three fourths of the ionosphere lies above the level of maximum electron density (the highest level which can be studied by conventional ground-based reflection measurements), it is difficult to arrive at a detailed description of the variability of spherical ionospheric refraction. Fig. 14 gives a *very rough estimate* of the median and the 5 and 95 percentile values of spherical ionospheric refraction of the VHF radio waves in temperate latitudes.

A casual glance at the world maps of f_0F2 reveals that the horizontal variation of even the "monthly median" ionosphere frequently exceeds one per cent in a hundred kilometers. These ubiquitous diurnal and latitudinal

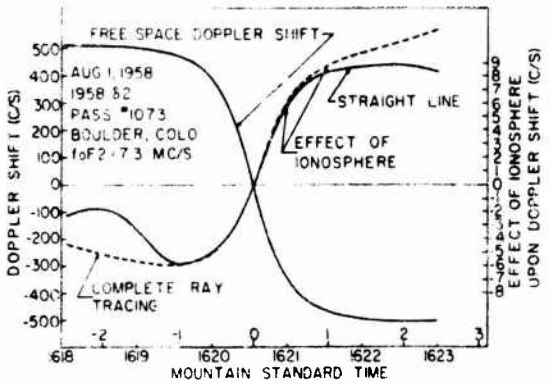


Fig. 13—The calculated effect of the ionosphere upon the Doppler shift of a 20-Mc signal from a satellite. The result of complete ray tracing is compared with the result deduced by simply integrating the refractive index along the straight line which joins the satellite and the observer. Satellite 1958 δ2 as observed at Boulder, Colo., August 1, 1958.

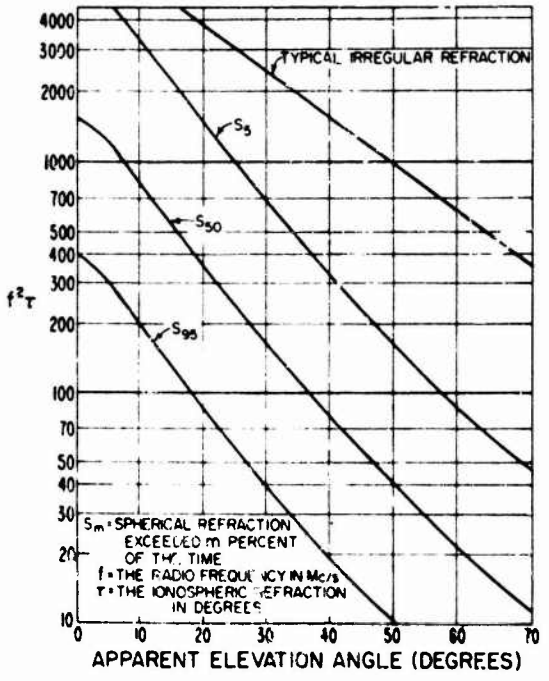


Fig. 14—Rough estimates of the magnitude of spherical and irregular ionospheric refraction at temperate latitudes.

variations account for an important component of wedge refraction. In addition, when the ionosphere is "disturbed," (as during a magnetic storm), much stronger horizontal gradients occur [31]. Furthermore, variable wedge refraction, fluctuating from one 10-minute period to the next, is frequently observed in the daytime. The magnitude of this variable wedge refraction is comparable with the more orderly varieties, and it is quite unpredictable. Fig. 14 includes a curve which indicates the importance of irregular wedge refraction. This curve has been sketched from the limited number of observations which are described in the literature. However, its elevation-angle dependence is steeper than the calculations of Fig. 12 would have predicted. We suggest that this apparent discrepancy may be the result of contamination of low-angle observations by irregular tropospheric refraction which becomes dominant at sufficiently high frequencies.

Ionospheric refraction may be predicted in a very crude way by using the value of f_0F2 from world maps, assuming a model for the upper ionosphere, and applying the approximate methods described by Komesaroff [20] or by Weisbrod and Anderson [15]. In any particular case, the result should be correct to within a factor of five. If actual observations of the ionosphere are available at the time and place of interest, the estimated refraction will usually be correct to within a factor of two.

The only way to obtain a really accurate estimate of ionospheric refraction is to make use of ionospheric dispersion by obtaining simultaneous observations at two or more frequencies. The refractivity of the ionosphere varies in proportion to the square of the wavelength of the radio wave and, to a very good approximation for frequencies above about 50 Mc, the total bending and angular errors behave in the same way. If two-frequency observations are contemplated, the lower of the two frequencies might very well lie in the range between 50 and 100 Mc. It is advantageous for the frequencies to be sufficiently low so that ionospheric effects will predominate over tropospheric effects. On the other hand, they must not be so low that the ray paths differ appreciably on the two frequencies. For instance, observations show that the magnitudes of the ionospheric effects upon the 20- and 40-Mc transmissions from artificial satellites are not in the predicted ration, nor do they always differ in the expected sense.

Ionospheric Scintillations

General Description

A radio wave passing through the ionosphere is scattered by irregularities in the electron density distribution. The scatter results in a redistribution of the amplitude of the radio wave in a manner similar to the effect of a diffraction grating, and also causes irregular fluctuations in the apparent position of a radio source viewed through the ionosphere. These processes are usually described as amplitude and angular scintillations, respectively, and were first noticed during the observations of radio stars which heralded the beginning of the radio astronomy of discrete sources [32], [33]. The radio noise from the sky was found to fluctuate in amplitude at times and this was first thought to be an indication of an inherently variable star signal. However, records obtained simultaneously from stations separated by 200 kilometers showed no similarity when the fading patterns were compared [34], [35], whereas it was found that when the receiving site separation was only about 1 km, then the fading patterns of the two records were very similar. This result showed that the scintillations were caused by irregular refraction in the ionosphere rather than variations in the source, and the

scintillation phenomenon has since proved to be useful as a tool for ionospheric research.

Both radio stars and satellites have now been used as sources for scintillation studies, and the differences in the characteristics of these sources can be very important. The radio stars give rise to plane electromagnetic waves at the ionosphere and appear to move slowly across the sky, whereas most satellites are sufficiently near the earth to produce spherical wave fronts in the ionosphere and move relatively quickly across the sky. Radio stars are wideband noise transmitters of finite angular size whereas satellites are essentially narrowband point sources of coherent waves.

Amplitude scintillations of a radio source appear as fairly rapid and irregular variations in the recorded signal strength and, in the case of the more intense sources, are usually much larger than the output noise fluctuations of the receiving system in the absence of scintillations (see Fig. 15). The magnitude of the scintillations is often comparable with the source signal strength at low VHF band frequencies, but there is much variability.

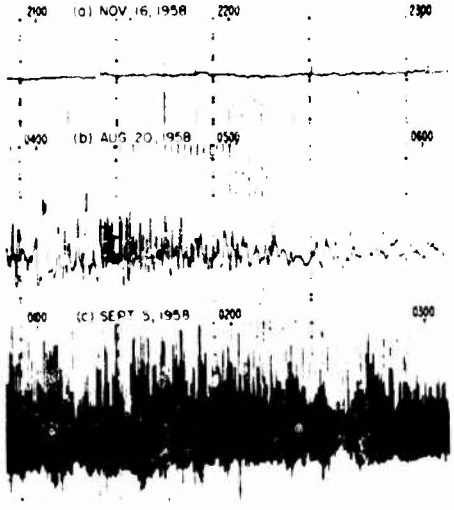


Fig. 15—Sample records from total-power scintillation equipment

In radio-star work, a recording time constant of about 0.5 sec is commonly used, and the periodicity of the scintillations observed varies widely in the range from 1 sec to several minutes. The "scintillation rate" is determined mainly by the drifting of the ionospheric irregularities in the radio star case, but when satellites are observed, the rapid motion of the satellite determines the rate, and the effect of ionospheric drifts is negligible. Typically, satellite signals produce a scintillation rate of about 10 cps and a recording time constant as small as 0.01 sec is necessary.

Phase scintillations can most easily be observed using a pair of antennas connected as an interferometer. The principal phase difference between the antennas will vary cyclically as the source moves through the interferometer lobes, and phase scintillations, when present, will be superimposed on this otherwise smooth variation. Phase scintillations may have a magnitude of 1 radian or more in the low VHF band, and this is equivalent to an apparent flicker in the direction of the source of about $\pm 1°$. Values much smaller than this are more common, and a value of a few minutes of arc could be considered as a typical direction error for the VHF band during scintillation conditions.

Diffraction at a Phase-Changing Screen

Observations of scintillations have shown that the ionosphere contains irregularities in electron density which are elongated along the earth's magnetic field, and at a given time, patches of these irregularities extend over wide geographic areas. It has been shown by various authors [36]–[39] that such a distribution of irregularities can be treated as a random phase-changing screen, although opinions differ as to the effect of the thickness of the postulated screen. If we assume a plane wave to be incident upon the ionosphere and that no absorption of the wave takes place, then immediately below the region containing the irregularities the surface of constant phase in the emergent wave will be corrugated in an irregular way. The amplitude distribution will, however, still be uniform across the wavefront since no energy has been extracted from the wave.

To investigate the process by which amplitude variations are built up from the purely phase-modulated wave, we must first consider the angular spectrum of the wave. The angular spectrum of a phase-changing screen is analogous to the frequency spectrum of a phase-modulated carrier wave, the deviation in angle of a particular component being equivalent to the shift in frequency of a particular sideband with respect to the carrier wave. For the case of a phase-modulated carrier wave it is well known that the sidebands occur not only at all the frequencies in the Fourier analysis of the phase-modulation, but also at sums and differences and multiples of these frequencies, and that the relative magnitude of these sidebands depends on the phase deviation. By analogy, it can be shown [40] for the case of diffraction, without specifying the screen in any detail, that the total spread of the angular spectrum depends on the phase deviation as well as on the lateral scale of the phase variations measured across the wavefront.

Now the phase relationship between the components of the angular spectrum is such as to produce a pure phase modulation at the screen itself. As the wavefront leaves the screen, however, the relative phases will be changed owing to the different directions of propagation within the angular spectrum, and in this way amplitude variations will be introduced into the diffraction pattern. The amplitude variations will grow in magnitude as the distance from the screen is increased until the initial phase relationship is entirely destroyed. The width of the angular spectrum increases with the degree of dis-

turbance of the screen, and the structure of the diffraction pattern becomes finer.

As long as the average phase changes introduced by the diffracting screen are less than about one radian, then the scale size of the irregularities observed on the ground is the same as the scale of the irregularities in the ionosphere. If the average phase changes are greater than one radian, then the scale of the pattern observed on the ground is smaller than the scale size in the ionosphere by a factor equal in magnitude to the average phase change in radians.

Magnitude of the Irregular Component

In order to estimate the magnitude of the irregular component in electron density which is necessary to produce scintillations, we must first compute the total change in phase-path length which is imposed on a radio wave in traversing the ionosphere, then consider the fractional variation in this path which will result in phase differences of about 1 radian in waves emerging at places separated by approximately the Fresnel-zone radius.

From (13) we see that the magnitude of the change in phase path length is given by

$$|\Delta l| = \frac{b}{\omega^2}\int_0^z N dl.$$

Taking the value 10^{17} electrons/m^2 column for $\int N dl$ and assuming a radio wave frequency of 40 Mc, we get $\Delta l \approx 2500$ m, which corresponds to about 2000 radians at 40 Mc. Now if we take $z = 500$ km to be the range of the irregularities, since $\lambda = 7.5$ m, the Fresnel-zone radius, $\sqrt{z\lambda}$, is approximately 2000 m.

For strong scintillations, we require changes in phase path of about 1 radian over the Fresnel-zone distance, so these calculations show that in order to produce scintillations at 40 Mc, the total electron content of the ionosphere must change by about 1 part in 2000 over distances of about 2000 m. Such variations can be expected to be quite common, indicating that scintillations on 40 Mc should be observed very frequently, and this is in fact the case [41].

Similar calculations for a radio frequency of 400 Mc show that electron-density variations of 1 part in 200 are required over distances of about 500 m. Since this distance is of the same order as the mean free path for electrons in the upper ionosphere, it is unlikely that there would be sufficiently intense density variations to cause scintillations at 400 Mc. Experimental observations support this conclusion; strong scintillations at 400 Mc are typically observed only during exceptionally disturbed conditions.

Theoretical Derivations

Phase scintillations: The magnitude of the phase scintillations which will be imposed on a radio wave by an assembly of spherical electron-density irregularities in the ionosphere can be shown theoretically in an approximate way [4]. The treatment which follows is similar to that given by Booker, and it is first necessary to express the variation of refractive index of the ionosphere as a function of electron-density variation. Taking the logarithmic derivative of (10),

$$\frac{\delta\mu}{\mu} = -\frac{\delta X}{2(1-\frac{1}{2}X)}.$$

Since

$$X = \frac{\omega_N^2}{\omega^2} = \frac{Ne^2}{\epsilon_0 m\omega^2}$$

and is small at VHF,

$$\frac{\delta\mu}{\mu} \approx -\frac{\delta X}{2} = -\frac{\delta N e^2}{2\epsilon_0 m\omega^2} = -\frac{\omega_N^2 \delta N}{2\omega^2 N}.$$

In terms of mean-square variations,

$$\overline{\left(\frac{\delta\mu}{\mu}\right)^2} = \frac{\omega_N^4}{4\omega^4}\cdot\overline{\left(\frac{\delta N}{N}\right)^2}.$$

For a single irregularity of effective thickness L, we can express the mean-square phase deviation which it produces as

$$\left(\frac{2\pi L}{\lambda}\right)^2\overline{\left(\frac{\delta\mu}{\mu}\right)^2}.$$

In a distance D along the radio path there will be D/L irregularities, assuming that all the irregularities have the same size. These irregularities result in a total mean-square phase deviation of

$$\overline{(\Delta\phi)^2} = \frac{D}{L}\left(\frac{2\pi L}{\lambda}\right)^2\overline{\left(\frac{\delta\mu}{\mu}\right)^2}.$$

In terms of electron density variations this can be written

$$\overline{(\Delta\phi)^2} = \frac{D}{L}\left(\frac{2\pi L}{\lambda}\right)^2\frac{\omega_N^4}{4\omega^4}\overline{\left(\frac{\delta N}{N}\right)^2}$$

which simplifies to

$$\overline{(\Delta\phi)^2} = \frac{DL\omega_N^4\lambda^2}{16\pi^2 c^4}. \qquad (35)$$

For a more exact treatment we could derive the mean-square phase deviation per unit distance, and then integrate along the radio path.

Angle-of-Arrival Fluctuations: Since phase scintillations are usually observed as a phase difference between the antennas of an interferometer system, and this in turn appears as an apparent change of direction of the radio source, we express the mean-square phase fluctuations in terms of the mean-square angular flicker.

If the line of separation of the two antennas of the interferometer is perpendicular to the direction of the

source, and if the phase fluctuations at the antennas have a correlation coefficient ρ, then the mean-square fluctuation in phase difference between the antennas is

$$\overline{\{\Delta(\phi_1 - \phi_2)\}^2} = 2(1 - \rho)\overline{(\Delta\phi)^2}.$$

For an antenna separation d, the phase difference can be expressed in terms of fluctuations in the apparent direction of a source as

$$\theta^2 = \left(\frac{\lambda}{2\pi d}\right)^2 \overline{\{\Delta(\phi_1 - \phi_2)\}^2}.$$

Hence,

$$\overline{\theta^2} = 2\left(\frac{\lambda}{2\pi d}\right)^2 (1 - \rho)\cdot\frac{D}{L}\cdot\left(\frac{2\pi L}{\lambda}\right)^2 \frac{\omega_N^4}{4\omega^4}\left(\frac{\delta N}{N}\right)^2$$

which simplifies to

$$\overline{\theta^2} = \frac{(1 - \rho)D L\omega_N^4}{2\omega^4}\left(\frac{\delta N}{N}\right)^2. \tag{36}$$

Amplitude Scintillations: We have discussed the way in which amplitude variations are built up from a wave which is modulated only in phase by the ionospheric irregularities. To establish a quantitative measure of amplitude variations we can first consider a wave of amplitude A which has small phase fluctuations imposed upon it. At a considerable distance from the screen, these phase fluctuations cause small amplitude fluctuations which have random phase relationships with respect to the unscattered portion of the wave, as seen in Fig. 16. In this vector representation of the phase and amplitude variations, it can be seen that the following relationship holds:

$$\overline{\left(\frac{\Delta A}{A}\right)^2} = \overline{(\Delta\phi)^2}. \tag{37}$$



Fig. 16—Vector representation of amplitude and phase scintillations.

If the phase fluctuations are of the order of 1 radian or more, then a large fraction of the original wave is scattered and the mean-square fractional deviation of amplitude reaches its limiting value.

In a rigorous treatment, the above relationship between amplitude and phase deviations is found to hold only when observations are made at distances exceeding that at which the effective size of the ionospheric irregularities is comparable with that of the Fresnel-zone radius. Since this distance (the "Fresnel-zone distance," L/λ^2) is several thousand kilometers at VHF, it is clear that observations are usually made at a distance where the amplitude variations are not fully developed. In this case, Hewish [2] has shown that the relevant formula is

$$\overline{\left(\frac{\Delta A}{A}\right)^2} = \frac{z^2}{z_F^2}\overline{(\Delta\phi)^2} \tag{38}$$

where z is the observing range and z_F is the Fresnel-zone distance. Therefore we get the following formula for amplitude fluctuations:

$$\overline{\left(\frac{\Delta A}{A}\right)^2} = z^2\left(\frac{\lambda^2}{L^4}\right)\frac{D}{L}\left(\frac{2\pi L}{\lambda}\right)^2\frac{\omega_N^4}{4\omega^4}\left(\frac{\delta N}{N}\right)^2 \tag{39a}$$

which simplifies to

$$\overline{\left(\frac{\Delta A}{A}\right)^2} = \frac{r_F^4 D\lambda^2\omega_N^4}{16\pi^2 L^3 c^4}\left(\frac{\delta N}{N}\right)^2 \tag{39b}$$

where r_F is the Fresnel zone radius, $\sqrt{zL}$.

Comparison of Theory with Observations

We have shown in an approximate way how both phase and amplitude scintillations can be expressed quantitatively. We can now examine the formulas to predict some ways in which scintillation phenomena depend on various parameters, and compare these predictions with experimental results.

Variation with Observing Frequency: Eq. (36) shows that the mean-square fluctuation in angle of arrival will depend on the fourth power of the observing wavelength. Values of around a few minutes of arc are commonly observed at the low-frequency end of the VHF band, and the predicted variation with wavelength agrees well with such observations [40].

The variation of the mean-square scintillation amplitude with wavelength is also as the fourth power of the wavelength, according to (39a). At about 30 Mc, the power amplitude index is commonly unity. In general, the variation with observing wavelength agrees with the theory throughout the VHF band [41].

Variation with Zenith Angle: The expected variation with zenith angle for the case of a homogenous ionosphere can be determined for both angle-of-arrival fluctuations and amplitude scintillations since, assuming a plane earth, both D and z vary linearly with the secant of the zenith angle (χ). Using (36) and (39a), respectively, we get

$$\overline{(\theta)^2} \propto \sec\chi$$

$$\overline{\left(\frac{\Delta A}{A}\right)^2} \propto \sec^2\chi$$

The variation of angle-of-arrival flicker with zenith angle has been deduced from observations made in the United States in the zenith angle range 55°–78° on 53 Mc and 108 Mc [42]. The results are in good agreement with the theory derived above in that the mean-square angular flicker was found to vary with the secant of the zenith angle on both frequencies.

Several workers have examined the variation of amplitude scintillations with zenith angle. All the results referring to a large range of zenith angles, Lawrence, *et al.* [42], Little and Maxwell [43], and Little, *et al.* [44], show that the increase with zenith angle is appreciably less than that expected using the above theory. However, the theory which has been presented involves only plane-ionosphere geometry and assumes that the irregularities are spherical in shape. In the case of a curved ionosphere, it is easy to show that, at large zenith angles, the quantities D and z which have been used no longer vary proportional to sec χ. For this reason alone, then, we would expect our simple theory to be inadequate. Little, *et al.* [44] have shown that their observations concerning the variation of amplitude index with zenith angle can be explained by modifying the above theory to include the fact that the irregularities are elongated with an axial ratio of between 8:1 and 4:1.

Variation with Ionospheric Conditions: Both the mean-square angular flicker and mean-square scintillation amplitude depend on the mean-square fractional deviation in electron density, as shown in (30) and (39b). Therefore, we would expect that scintillations would increase when observing through a more disturbed ionosphere. This feature has been demonstrated both in the variation of scintillations over the sunspot cycle [46] and in the observation of severe scintillations on high VHF band frequencies at high latitudes [44]. In addition, an extremely violent type of scintillation is observed near the geomagnetic equator [45].

Variation with Source Distance: In the case of radio-star observations, we are concerned with a plane wave incident on the ionosphere, and hence a Fresnel zone radius of $\sqrt{z\lambda}$, where z is the distance from the ground to the irregularities. In the case of a transmitter whose distance is z' from the ionospheric irregularities, it can be shown by simple trigonometry that the Fresnel-zone radius is

$$\sqrt{\left(\frac{zz'}{z+z'}\right)\lambda}.$$

Then, for a nearby transmitter, since the mean-square amplitude of scintillations is proportional to the fourth power of the Fresnel-zone radius (39b) we can write

$$\overline{\left(\frac{\Delta A}{A}\right)^2}_{z'} = K\left(\frac{zz'\lambda}{z+z'}\right)^2.$$

But we know from (38) that when $z' \to \infty$,

$$\overline{\left(\frac{\Delta A}{A}\right)^2}_{z'\to\infty} = \frac{z^2}{z_F^2}\,\overline{(\Delta\phi)^2}$$

$$\therefore \lim_{z'\to\infty} K\left[\frac{zz'\lambda}{z+z'}\right]^2 = \frac{z^2}{z_F^2}\,\overline{(\Delta\phi)^2}$$

$$\therefore K z^2\lambda^2 = \frac{z^2}{z_F^2}\,\overline{(\Delta\phi)^2}$$

$$\therefore K = \frac{\overline{(\Delta\phi)^2}}{z_F^2\lambda^2}$$

$$\therefore \overline{\left(\frac{\Delta A}{A}\right)^2}_{z'} = \frac{\left(\frac{zz'}{z+z'}\right)^2}{z_F^2}\,\overline{(\Delta\phi)^2}.$$

Hence when comparing the scintillations of a source at distance z' from the ionosphere with those from a source at infinity viewed through the same region of the ionosphere, we expect the mean-square scintillation amplitude for the nearby source to be a fraction

$$\left(\frac{z'}{z+z'}\right)^2$$

of that of the infinitely distant source. In the special case of a satellite at a distance twice that of the irregularities from the observer, this fraction is $\frac{1}{4}$. The variation of the fraction is plotted in Fig. 17 as a function of z'/z.

There are virtually no results with which to compare this theoretical dependence on source distance. One report from Alaska [47] states that a 20-Mc satellite signal showed only small scintillations when a radio-star signal was scintillating violently at 223 Mc, thus indicating that the irregularities were very near the satellite. However, the difficulty of simultaneous satellite and radio-star scintillation observations lies in the fact that a suitable star and a satellite are only in the same region of the sky for a very short time and on rare occasions. In any case, as indicated in the following sections, the observations may easily be confused by the differences in angular size and bandwidth of the sources.

Effect of Observing Bandwidth

Since radio stars are wideband sources of noise, it is interesting to consider the dependence of the observed scintillation activity on the receiver bandwidth in use. If the scintillation pattern is not identical at all frequencies in the band, there will be a reduction in the resultant scintillation index because the uncorrelated variations will in effect tend to cancel one another in the receiving system.

Weak-Scattering Case: The semi-angular width of the cone over which scattered radiation is received is λ/L,

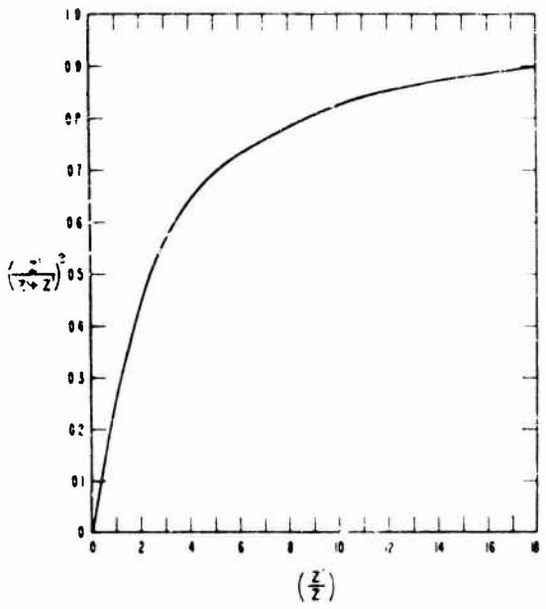


Fig. 17—Variation of mean-square scintillation amplitude with distance of source from diffracting screen.

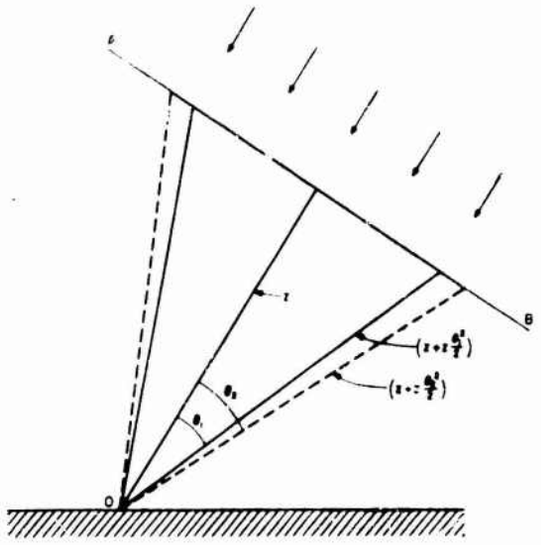


Fig. 18—Scattered cones of radiation from a point source on two nearby frequencies. AB is the wave normal for an observer at O, and z is the range of the irregularities. The cone angles are very exaggerated.

where L is the scale size[1] of the irregularities. The cone size is therefore frequency dependent and we must calculate the frequency separation which gives rise to significantly different scintillation patterns. If the record is to be insensitive to the bandwidth of observation, this frequency separation must be greater than the over-all receiver bandwidth. A condition at which correlation just begins to diminish is that the difference in path between the generators of the cones (in terms of phase), differs by more than 1 radian at the two frequencies.

Referring to Fig. 18, which shows the scattered radiation cones for two frequencies, we see that since θ_1 and θ_2 are small we can express the condition for good correlation on the two frequencies as

$$\frac{2\pi}{\lambda_2}\frac{z\theta_2^2}{2} - \frac{2\pi}{\lambda_1}\frac{z\theta_1^2}{2} < 1. \tag{40}$$

Since $\theta_1 = \lambda_1 / L$ and $\theta_2 = \lambda_2 / L$, this reduces to

$$\lambda_2 - \lambda_1 < \frac{L^2}{\pi z}.$$

Setting $\lambda_2 = \lambda_1 + \Delta\lambda_1$, we get for the fractional correlation bandwidth

$$\frac{\Delta\lambda_1}{\lambda_1} = -\frac{\Delta f_1}{f_1} < \frac{L^2 f_1}{\pi z c}. \tag{41}$$

Using (41), and assuming that the irregularities have a scale size of 5 km [43] and are at a range of 500 km, the fractional correlation bandwidth is clearly greater than unity throughout the VHF band. This means that there will be no effects due to correlation bandwidth in any conventional receiver. Even at 10 Mc, the correlation bandwidth is about 5 Mc, which is still an unrealistically wide bandwidth for the frequency considered.

Strong-Scattering Case: As stated earlier, the scale of the irregular diffraction pattern on the ground (L_0) is smaller than the ionospheric scale (L) by an amount ϕ (>1), which is the magnitude of the phase deviations in radians. Therefore, we have

$$\theta_1 = \frac{\lambda_1}{L_0} = \frac{\lambda_1\phi_1}{L}, \qquad \theta_2 = \frac{\lambda_2\phi_2}{L}.$$

and the condition for correlation bandwidth (40) reduces to

$$\lambda_2\phi_2^2 - \lambda_1\phi_1^2 < \frac{L^2}{\pi z}.$$

Since ϕ varies linearly with λ, this becomes

$$\lambda_2^3 - \lambda_1^3 < \frac{L^2}{K^2\pi z}$$

where $K = \phi_1\lambda_1 = \phi_2\lambda_2$. Setting $\lambda_2 = \lambda_1 + \Delta\lambda_1$ and ignoring second order terms we get the condition

$$\lambda_1^2\Delta\lambda_1 < \frac{L^2}{3K^2\pi z}$$

so that the fractional correlation bandwidth is

$$\frac{\Delta\lambda}{\lambda} = -\frac{\Delta f}{f} < \frac{L^2 f^2}{3\pi K^2 z c^2}. \tag{42}$$

This equation can be written in terms of the amount of phase deviation, ϕ, as

$$\Delta f < -\frac{L^2 f^2}{3\pi z c\phi^2}. \tag{43}$$

[1] "Scale size" is the dominant term in the Fourier analysis of the irregularity distribution in space and is 2π times greater than the autocorrelation distance.

For irregularities of scale size 5 km viewed at a range of 500 km, on a frequency of 100 Mc, this formula predicts a correlation bandwidth of $160/\phi^2$ Mc so that the value of ϕ would have to be very high for there to be any effect due to correlation bandwidth at this frequency. As (42) shows, however, the correlation bandwidth depends on the fourth power of the frequency, so that correlation bandwidths of the order of typical radio-astronomy receiver bandwidths can reasonably be expected for the low-frequency end of the VHF band.

One experiment in which correlation bandwidths may have affected the results is that reported by Chivers and Davies [49] in which scintillations were observed simultaneously at 1390 Mc and 79 Mc. At times when small scintillations were present at the higher frequency, the scintillations on the lower frequency were smaller than expected from the application of (39a). However, using (38) it is apparent that if 5 per cent amplitude scintillations occur on 1390 Mc, then assuming irregularities of scale size 5 km at 500-km range, the phase deviations in the 1390-Mc wave are approximately 3 radians. This indicates phase deviations of about 53 radians at 79 Mc and consequently, by application of (43), a correlation bandwidth of only 10 kc at 79 Mc. The receiver bandwidth used was 300 kc so that lack of correlation over the band could have caused the reduction of the 79-Mc scintillations.

Effect of Source Size

Since radio stars are not point sources of radiation, we must consider the effect of their finite angular size on the observed scintillations. It is well known that, whereas the visible stars are observed to twinkle, the planets do not, simply because of the differences in the angular sizes of the stars and the planets. Since radio stars vary greatly in angular size it is important to calculate the critical angular size for which a source will not scintillate using our knowledge of the size of the irregularities.

A rigorous treatment of this problem requires application of the theorems of Booker, Ratcliffe and Shinn [36], but the geometry shown in Fig. 19 lends itself to simple, approximate calculations. AB and CD are wave normals for radiation from opposite edges of an extended source. The scattered cones of radiation of semi-angle θ are shown, the source subtends an angle $\Delta\theta$ and the irregularities are at a range z. A condition at which the correlation of the scintillation pattern from extreme parts of the source begins to diminish, is that the phase-path difference between the length of the generator of one of the scattered cones and that included by the other cone measured along the same path (*i.e.*, the distance x) exceeds 1 radian. Referring to Fig. 19, since θ is small, this condition is approximately

$$\frac{2\pi}{\lambda}\left(z + z\frac{\theta^2}{2}\right) - \frac{2\pi}{\lambda}\left(z + z\frac{(\theta - \Delta\theta)^2}{2}\right) < 1$$

and this becomes

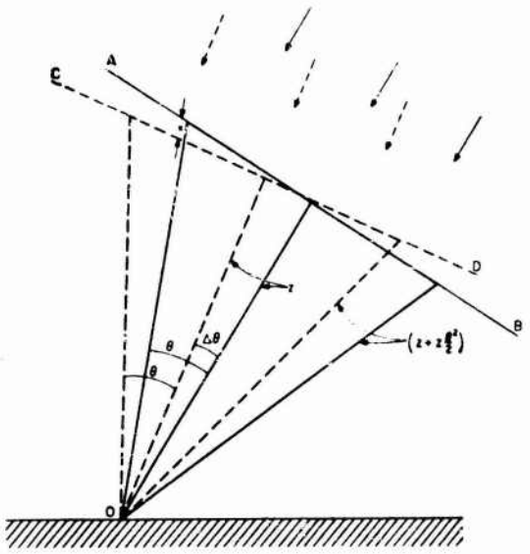


Fig. 19—Scattered cones of radiation from opposite edges of a source of angular size $\Delta\theta$. AB, CD are wave normals for an observer at O, and z is the range of the irregularities. The cone angles are very exaggerated.

$$\frac{2\pi z\theta\Delta\theta}{\lambda} < 1. \tag{44}$$

Weak-Scattering Case: Since for weak scattering the semi-angular width of the cone of scattered radiation is given by $\theta = \lambda/L$, where L is the average irregularity scale size, (44) becomes

$$\Delta\theta < \frac{L}{2\pi z} \tag{45}$$

which is of the order of the angular size of an individual irregularity at the range of the diffracting screen. In other words, correlation begins to diminish if the angular size of the source exceeds that of the individual irregularities. Eq. (45) predicts that for irregularities of scale size 5 km and a range of 500 km the angular size of the source must be less than 6′ arc if there is to be no reduction in scintillations due to the effects of source size. If sources are viewed at low elevation, so that the range of the irregularities might be 1000 km or more, the formula sets a critical source size of less than 3′ arc.

Strong-Scattering Case: To recapitulate, for this case, $\theta = \lambda/L_g$ where L_g is the scale of the irregularities on the ground. Since $L = L_g\phi$ where ϕ is the magnitude of the average phase change in radians ($\phi > 1$) we get $\theta = \lambda\phi/L$. Hence, (44) becomes, for strong scattering

$$\Delta\theta < \frac{L}{2\pi z\phi} \tag{46}$$

which indicates that, for strong scattering, correlation diminishes due to source-size effects when the source subtends an angle which is greater than the angular size of the irregularities divided by the average phase change in radians.

For irregularities of 5-km scale size, viewed at a range of 500 km, with average phase deviations of magnitude

3 radians, (46) predicts that the source must subtend an angle less than 2′ arc to be free from source size effects. When viewing at lower elevations (larger ranges) the source size must be correspondingly smaller.

The strong radio source in Cassiopeia is a disk subtending about 4′ arc, so that a reduced scintillation index might arise during strong scintillation conditions and particularly at low elevations. During the last sunspot maximum period, several groups of workers reported effects which they attributed to the finite angular size of the source [44–48]. We would expect the effect to be dependent on zenith angle because the range of the irregularities varies with this parameter. In the case of strong scattering, we would expect an additional dependence on the observing frequency because the magnitude of the average phase change is frequency dependent.

The above discussion of the effects of receiver bandwidth and source size is applicable to both total power and interferometric receiver systems. However, it should be noted that interferometric systems are subject to additional effects which may become significant under certain circumstances, especially for narrow-band point sources. If the separation of the individual antennas comprising the interferometer exceeds the correlation distance of the diffraction pattern across the ground, interference of the radio waves will become incomplete and the source will tend to disappear from the interferometer recording.

The Role of Ionospheric Focusing in the Scintillation of Radio Stars

We have seen that the theory of diffraction at an irregular screen explains many of the observed features of radio-star scintillation. According to this theory, the instantaneous amplitude observed at the ground depends upon the simultaneous effects of a number of irregularities and, in particular, upon the detailed phase relationships of the waves diffracted by those irregularities. Thus, we expect the details of the amplitude pattern on the ground to be very sensitive to wavelength and we have, in fact, calculated the relatively narrow bandwidths over which strong scintillations at low frequencies should be correlated.

This point was investigated by Wild and Roberts [50] in a series of swept-frequency observations in which they observed scintillations continuously over the frequency band 40–70 Mc. They occasionally found the scintillations to be quite random and unrelated at various frequencies in that band. More commonly, however, the scintillations were very well correlated and, in fact, practically identical at all frequencies between 40 and 70 Mc.

The interpretation of such broad-band fluctuations in terms of diffraction at an irregular screen is by no means obvious. The large bandwidth clearly indicates that only one irregularity can contribute to one fluctuation, that irregularity acting like a lens to focus the radiation on the ground. From the geometry of their observations, which were made at an elevation angle of 15°, Wild and Roberts calculated that the irregularities must have a scale size of at least 3 km to produce broad-band patterns, and they must be smaller than 2.3 km to produce random patterns. Occasionally, both types were observed simultaneously.

Polarization Effects

The Production of Polarization Rotation in the Ionosphere

A radio wave is completely described in terms of the variation, with time and space, of its electric and magnetic fields. These vectors are, of course, simply the sum of their components. The most common way to resolve a time-varying vector into components is to describe it in terms of mutually perpendicular component vectors, each remaining constant in direction but varying with time. This is by no means the only way to resolve a vector. For example, a vector with constant direction (except for sign) and sinusoidally varying amplitude may be resolved into two component vectors which have constant and equal amplitude, but which rotate continuously in opposite directions.

Following this line of reasoning, it is clear that a linearly polarized radio wave is equivalent to two circularly polarized components rotating in opposite senses. The characteristic modes of a radio wave in the ionosphere (the ordinary and extraordinary waves) are, in the general case, elliptically polarized in opposite senses of rotation, and they travel with different phase velocities. In examining the behavior of these waves it is useful to distinguish between two cases: the longitudinal case, where the direction of propagation is parallel to the earth's magnetic field, and the transverse case, where the direction of propagation is perpendicular to the field.

In the longitudinal case the ellipses become circles. Because the two components have different phase velocities, the plane of polarization of the linearly polarized resultant rotates gradually as the wave moves through the ionosphere. This phenomenon is known as Faraday rotation. It may be computed as follows: Consider an element of path length dl in which the rotations of the electric vectors of the ordinary and extraordinary rays are dK_+ and dK_-, respectively. Fig. 20 shows the relationship between dK_+, dK_-, and $d\Omega$, the angle through which the plane of polarization rotates as the wave travels the distance dl. From Fig. 20, $dK_- + d\Omega = dK_+ - d\Omega$, or

$$d\Omega = \frac{dK_+ - dK_-}{2}. \tag{47}$$

Integrating dK_+ and dK_-, each along its own ray path from the transmitter to the observer,

$$\Omega = \frac{1}{2}\left[\int_+ dK_+ - \int_- dK_-\right] = \frac{1}{2}(K_+ - K_-) \tag{48}$$

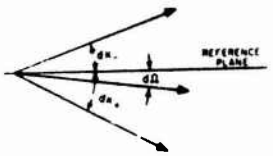


Fig. 20—Rotation of the plane of polarization of the linearly polarized resultant of two circularly polarized waves.

where $K_+ = (2\pi/\lambda)P_+$ and $K_- = (2\pi/\lambda)P_-$ are the angular phase-path lengths of the ordinary and extraordinary rays. If the frequency is sufficiently high that the two ray paths may be assumed to coincide, a further simplification follows. Denoting by λ_+ and λ_- the wavelengths corresponding to the phase velocities of the two waves, then

$$dK_+ = \frac{2\pi}{\lambda_+}\,dl$$

(At such a high frequency, we may clearly equate to unity the cosine of the angle between the wave normal and the ray, a factor which properly appears in this expression.) and

$$dK_- = \frac{2\pi}{\lambda_-}\,dl$$

from which (47) becomes

$$d\Omega = \pi\left(\frac{1}{\lambda_+} - \frac{1}{\lambda_-}\right)dl$$

or,

$$d\Omega = \frac{\pi}{\lambda}(\mu_+ - \mu_-)dl. \tag{49}$$

Here, $\lambda = \mu_+\lambda_+ = \mu_-\lambda_-$ is the free-space wavelength and $(\mu_+ - \mu_-)$ is the difference between the refractive indices of the ordinary and extraordinary waves.

In this quasi-longitudinal case, we have from (8)

$$\mu \approx 1 - \frac{X}{2}(1 \mp Y_L) = 1 - \frac{X}{2} \pm \frac{XY_L}{2},$$

from which $\mu_+ - \mu_- \approx XY_L$ and

$$d\Omega \approx \frac{\pi}{\lambda}(XY_L)dl. \tag{50}$$

In a rationalized system of units,

$$XY_L = \frac{Ne^2}{\epsilon_0 m\omega^2}\cdot\frac{\mu_0 He\cos\theta}{m\omega}$$

and

$$\lambda = \frac{2\pi c}{\omega}, \qquad \omega = 2\pi f$$

so, approximately,

$$\frac{d\Omega}{dl} \approx \frac{Q}{f^2}NH\cos\theta$$

where

$$Q = \frac{e^3\mu_0}{8\pi^2 cm^2\epsilon_0} = 2.97\times10^{-2}\ \text{(mks)}.$$

f = wave frequency, cps.
N = electron density, electrons m^3.
H = magnetic field intensity, amp-turns/meter.
Ω = angle of polarization rotation, radians.
l = ray-path length, meters.

Integrating along the ray path,

$$\Omega = \frac{Q}{f^2}\int_{ray} NH\cos\theta\,dl. \tag{51}$$

This expression gives the correct result only when the two ray paths effectively coincide and when the two components are essentially circularly polarized. When θ is close to the limiting value which separates the quasi-longitudinal and quasi-transverse cases, the circles become ellipses and, when the eccentricity is large, $d\Omega/dl$ depends not only upon θ, but also upon the orientation of the plane of polarization of the wave relative to the plane containing the magnetic field and the direction of propagation.

The sense of rotation is such that, for a plane-polarized wave traveling in the direction of the magnetic field (*e.g.*, a down-coming wave in the northern hemisphere), the electric vector describes a right-handed helix. We can determine this fact by noting first that, looking along the field, electrons appear to revolve in a clockwise direction. At a given point in space the electric vector of the extraordinary wave rotates in the same direction as the electrons. (This seems proper because, of the two characteristic waves, it is the extraordinary which is more affected by the presence of the electrons.) Since the extraordinary wave as a whole progresses without rotating, the end point of its electric vector must describe a left-handed helix in space. (This is analogous to the fact that if a screw is to be pushed, without rotation, through a nut, the screw must have a left-handed thread if it is to cause clockwise rotation of the nut, viewed in the direction of motion of the screw.) Thus, for this case, when traveling along an element of path length dl in the direction of propagation, the changes in the vectors are as shown in Fig. 20, and the plane-polarized resultant is a right-handed helix.

The concept of "rotation angle" seems quite clear when propagation is in a quasi-longitudinal direction and the polarization angle is simply rotated. When propagation becomes transverse, the term loses its significance because the polarization ellipse changes its dimensions along the ray path. In this case, the two elliptical components degenerate into linearly polarized waves, one component always remaining parallel to the magnetic field and the other one perpendicular to it. However, because the components travel with different phase velocities, their resultant gradually changes along

the path from a linearly polarized wave to an elliptically polarized wave, and finally to one which is circularly polarized. Then the sequence is reversed and the wave becomes linearly polarized again. Now, however, the plane of polarization is perpendicular to the original direction, *i.e.*, the linearly polarized resultant has been rotated by 90°. Eq. (49) is still applicable in this case, and we may use it to find the expression for transverse propagation corresponding to (50).

In the transverse case at frequencies well above the plasma frequency, the Appleton-Hartree expression (1) gives, for the refractive index of the ordinary wave,

$$\mu_+ = (1 - X)^{1/2} \approx 1 - \frac{X}{2}$$

and, for the extraordinary wave,

$$\mu_- = \left[1 - \frac{X}{1 - \left(\frac{Y_T^2}{1 - X}\right)}\right]^{1/2} \approx 1 - \frac{X}{2\left[1 - \left(\frac{Y_T^2}{1 - X}\right)\right]}$$

$$\approx 1 - \frac{X}{2}\left[1 + \left(\frac{Y_T^2}{1 - X}\right)\right] \approx 1 - \frac{X}{2} - \frac{XY_T^2}{2}.$$

Thus

$$\mu_+ - \mu_- = \frac{XY_T^2}{2}$$

and (49) becomes, for transverse propagation,

$$d\Omega = \frac{\pi}{2\lambda}(XY_T^2)dl. \tag{52}$$

Y_T is very small at the frequencies which we are considering and, thus, a comparison between (50) and (52) shows that the Faraday rotation is relatively insignificant in the transverse case. It is important to realize that the "rotation" which occurs in the transverse case is actually a complicated series of polarization changes and, if the transverse condition occurs near the boundary of the ionosphere, the wave will commonly emerge with elliptical polarization. If the transverse condition occurs within the ionosphere it is frequently true that the small amount of rotation occurring near the transverse region can be neglected and the longitudinal approximations can be used. However, it is important to realize that the sense of rotation is reversed after passage through the transverse region, and at that point signs must be changed in the integrands of (48). This effect shows explicitly in (51) where $\cos\theta$ changes sign.

The Use of Faraday-Rotation Observations to Study the Ionosphere

We have seen that, for almost all directions of propagation, the angle of Faraday rotation of a linearly polarized radio wave is proportional to the integrated electron content of the ionosphere along the path, weighted by the longitudinal component of the geomagnetic field strength. Ever since Murray and Hargreaves [51] first pointed out that the slow fading of lunar radio echoes is caused by the Faraday effect, the effect has been used by the observers of moon echoes to measure the total electron content of the ionosphere. Although the number of electrons in a column of unit cross section up to a height of 1000 km is believed to be comparable to the number in a similar column from 1000 km on out to the moon, the inverse-cube law of decrease of the geomagnetic field insures that the electrons in the ionosphere (below, say, 1000 km) are the only ones which are effective in producing Faraday rotation. The determination of the total angle Ω of rotation may be accomplished by using two VHF signals spaced in frequency by several Mc, and noting the difference in polarization between them. Such measurements have been made, for example, by Evans [52], and by Taylor [53].

In moon-echo experiments, the frequencies are usually high enough that (51) is highly accurate. Since we are interested in determining the total electron content in a vertical column through the ionosphere, it is convenient to replace the path element dl by $\sec i dh$, where i is the local zenith angle and dh is an element of height. With this substitution, (51) becomes

$$\Omega = \frac{Q}{f^2}\int_0^{1000\text{ km}} NH\cos\theta\sec i dh = \frac{Q}{f^2}\int_0^{1000\text{ km}} NM dh$$

where

$$M \equiv H\cos\theta\sec i.$$

We must remember that this applies to only one half of the round-trip path through the ionosphere. Although further restricting the generality of the expression, it is convenient to take M as constant in the first 1000 km above the earth, and to write

$$\Omega = \frac{QM}{f^2}\int_0^{1000\text{ km}} N dh \tag{53}$$

for one half of the total Faraday rotation observed on a moon echo. This is a good approximation for small zenith angles, and it is the equation which has customarily been used in evaluating the results of moon-echo experiments.

In order to determine the electron content using (53), the value of M must be calculated. Yeh and Gonzalez [54] have made extensive calculations of the variation of the function M for various receiving stations, with particular emphasis upon the variation for their receiving station at the University of Illinois. Yeh and Gonzalez point out there are only a few favored directions for which M is essentially constant below 1000 km. In general, the value of M to be used in (53) must be a weighted average of the actual M along the line of sight, the weighting being done according to electron density. By using an "average" ionosphere profile, Yeh and Gonzalez calculated the weighted average, $\overline{M}$, of M along various directions. They found that $\overline{M}$ is generally equal to M at a height somewhere in the range 340 to 400 km. Fig. 21 gives values of $\overline{M}$ for directions

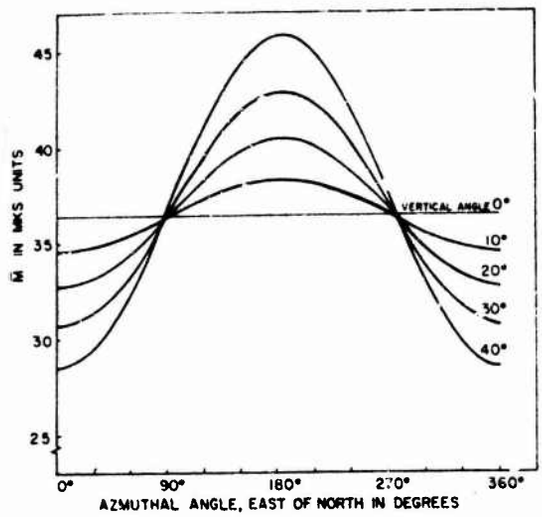


Fig. 21—The variation of $\overline{M}$ at the University of Illinois.

within 40° of the zenith at the University of Illinois.

The advent of satellite-borne radio transmitters has offered a new opportunity for Faraday-rotation studies of the ionosphere [55]. It is important to note that satellite studies, in principle at least, have the advantage that the electron content of the ionosphere can be measured over a sizeable geographical area from a single observing site, within the course of the few minutes duration of a satellite passage. In order to make full use of this advantage, however, it is necessary to remove the restriction to small zenith angles; it is also desirable to use frequencies relatively close to the ionospheric critical frequency in order to improve the sensitivity of the observations.

The failure of (53) at low frequencies and at large zenith angles is strikingly demonstrated by simultaneous observations of the 20- and 40-Mc signals from the Soviet satellites. Eq. (53) predicts that the ratio of the Faraday-rotation rates on the two frequencies shall be 4.0; in fact ratios of up to 5.0 have been observed, and on occasion, for a short period of time the sense of rotation rate has even been opposite on the two frequencies [56].

In addition to the failure of (53), the interpretation of satellite observations is hampered by the difficulty of determining the actual value of Ω. For frequencies as low as 20 or 40 Mc, the polarization vector may be rotated many tens of times in traversing the ionosphere, and so, even if the orientation angle of the transmitting antenna is known, the observed Ω is ambiguous by an additive constant, $n\pi$. One approach, which it has not yet been possible to test, is to install transmitters at nearby frequencies on a single satellite, and then to use the same method of analysis that has been applied successfully to moon-echo observations. Lacking such closely spaced frequencies, several observers have proposed, and used with varying success, a number of methods for removing the ambiguity.

Bowhill [57] has shown that, for a satellite traveling with uniform velocity at constant height over a plane earth, and for a uniform geomagnetic field, the observed Faraday rotation rate at any receiving site would be constant, and proportional to the electron content below the satellite. As one method for avoiding the problem of ambiguities in measurements of Ω, Garriott [58] modified this result to include the effect of a small vertical velocity component and used observations of the Faraday-rotation rate to deduce electron content. Alternatively, if the electron content is otherwise known, it can be combined with the observed Faraday rotation rate to estimate the local electron density at the height of the satellite. Despite Garriott's modifications, many limitations still remain in the derivation of the equations; perhaps the most serious is the requirement for a spherically stratified ionosphere. A more elaborate method, involving several ray tracings through a model ionosphere, has been used by Little and Lawrence [56]. In unfavorable circumstances, the results of even this complex procedure can be affected by unknown horizontal gradients in the ionosphere. The use of closely spaced frequencies on a single satellite would certainly permit a major step forward.

A Brief Summary of the Results of Faraday-Rotation Observations

Faraday-rotation measurements have already been able to contribute something to the difficult problem of determining the electron-density profile of the ionosphere above h_{max}, the level of maximum density. Moon-echo experiments and observations from high satellites can tell us very little about the shape of the electron-density profile but, when combined with true-height profiles up to h_{max} as deduced from ionograms, they do provide a means of estimating the electron content above h_{max}.

From moon-echo experiments, Evans [52] found an average value of about 4 for this ratio of electron contents. Hill and Dyce [59] found evidence for sizeable variations in the ratio and Garriott [60] observing the high passes of Sputnik III, found a ratio of 3 in the daytime and of 5 at night.

The low passes of Sputnik III, ranging in height from 300 to 900 km, enabled Garriott to deduce an average electron-density profile by making the assumption that the shape of the profile has no diurnal or seasonal variation. He found that the density fell to one half its maximum value at a height of about 150 km above h_{max}, and to one fourth its maximum value at about 350 km above h_{max}.

Yeh and Swenson [61], Garriott [60], and Hame and Stuart [62] have all observed a strong diurnal variation in the total electron content of the ionosphere. They agree that the daytime peak is six to eight times greater than the nighttime minimum.

With the analysis techniques which have been available to date, it is difficult to guarantee an absolute accuracy of any one measurement better than ten per cent. However, the sensitivity of the measurements is

enormously better than the absolute accuracy, and it is possible to detect 0.1 per cent changes in total content during a satellite pass [56]. Fig. 22 shows an example of the irregular structure which is commonly displayed by Faraday-rotation measurements. This result is simply another manifestation of large-scale irregularities responsible for the irregular wedge refraction described previously.

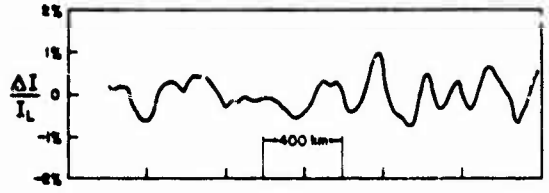


Fig. 22—Irregularities in ionospheric electron content, 0834 MST, September 4, 1958, at Boulder, Colo. $\Delta I/I_L$ is the variation, with horizontal distance, of integrated electron content of the ionosphere relative to that of a smooth, average ionosphere.

Summary

In this survey of ionospheric effects upon earth-space propagation we have mentioned a variety of phenomena. They may be classified as follows:

Amplitude: Ionospheric absorption does occur, but it diminishes rapidly as the radio-wave frequency is increased and hence is unimportant for the frequencies above 100 Mc which would normally be used for space communication. Rather, absorption at the lower frequencies is an important tool for studying the low-lying levels of the ionosphere and the interaction of these levels with incoming radiation and charged particles.

We have mentioned the divergence effect which, in certain cases, may exceed absorption effects but which is nevertheless unimportant for practical space communication.

The irregular structure of the ionosphere is responsible, through the interaction of the diffracted waves which it produces, for amplitude-scintillation effects. We have seen that amplitude scintillations can become extremely severe and thus they may represent a practical limitation for certain types of communication. They are most severe at night and at high latitudes, although a particularly violent form of scintillation is also found near the geomagnetic equator.

The refractive effects of the larger irregularities in the ionosphere might be expected to cause focusing of the radio waves. This phenomenon has been observed to occur at radio frequencies below 70 Mc.

In addition to its direct amplitude effects, the irregular structure of the ionosphere may operate in more subtle ways to produce the appearance of amplitude changes on a particular receiving system. For example, the irregularities may so enlarge the apparent size of a small source, or so destroy the correlation of the radio waves over the aperture of a large antenna, that the source seems to disappear. Also, the ordinary and extraordinary rays, traveling along slightly different paths, may be subjected to uncorrelated phase fluctuation. The polarization of the resultant wave will vary rapidly and, if a linearly polarized antenna is used, the signal will undergo severe amplitude fluctuations, attributable to the "polarization flutter."

Angle of arrival: We have seen that the steady, spherically stratified component of the ionospheric electron-density distribution increases the apparent elevation angle of an external radio source, and that this effect is proportional to the square of the radio wavelength. The magnitude of this systematic refraction is commonly overshadowed by both the regular and irregular components of wedge refraction. We have seen that the long-term average component of wedge refraction at high elevation angles is to some extent predictable from the expected horizontal gradients in the ionosphere. Irregular refraction, which typically varies from one ten-minute period to the next, commonly dominates other types of ionospheric refraction, and is not predictable.

The same irregularities which produce amplitude scintillations also produce angular scintillations. We have described the effects of these short-period fluctuations upon angle-of-arrival measurements. These may be so severe as to cause the source to "disappear" under certain conditions [44].

Frequency: We have mentioned the effects of the ionosphere upon the apparent frequency of a radio signal from a satellite. Since the ionosphere, in this case, produces only a second-order effect, its influence upon a communications circuit is generally unimportant. On the other hand, the dispersive Doppler effect and the related phase-modulation techniques are of considerable interest as a tool for measuring the electron content of the ionosphere and exosphere.

Radar Range: A brief discussion of the effect of the ionosphere upon radar range has shown this effect to be small, typically less than 400 meters at 100 Mc.

Polarization: We have given particular attention to an understanding of ionospheric Faraday rotation. This phenomenon seriously affects any space communication system which uses linear polarization and, perhaps more important, it is the basis for an accurate and sensitive technique to measure the total electron content of the ionosphere.

Bibliography

[1] J. A. Ratcliffe, "The magneto-ionic theory and its applications to the ionosphere," Cambridge University Press, Cambridge, England; 1959.
[2] A. Hewish, "The diffraction of radio waves in passing through a phase-changing ionosphere," *Proc. Roy. Soc. (London) A*, vol. 209, pp. 81–96; October 8, 1951.
[3] C. G. Little, "A diffraction theory of the scintillation of stars on optical and radio wavelengths," *Monthly Notices Roy. Astron. Soc.*, vol. 111, pp. 289–302; 1951.
[4] H. G. Booker, "The use of radio stars to study irregular refraction of radio waves in the ionosphere," Proc. IRE, vol. 46, pp. 298–314; January, 1958.
[5] S. Chapman and C. G. Little, "The nondeviative absorption of high frequency radio waves in auroral latitudes," *J. Atmos. Terr. Phys.*, vol. 10, pp. 20–31; January, 1957.
[6] C. G. Little and H. Leinbach, "Some measurements of high-

latitude ionospheric absorption using extraterrestrial radio waves," PROC. IRE, vol. 46, pp. 334–348; January, 1958.
[7] D. K. Bailey and J. M. Harrington, "A survey of polar cap absorption events (Solar proton events) in the period 1952 through 1960," *J. Phys. Soc. Japan*, vol. 17 (suppl. A-II), pp. 334–336; January, 1962.
[8] R. V. Bhonsle and K. R. Ramanathan, "Studies of cosmic radio noise on 25 Mc/s at Ahmedabad," *J. Sci. Indus. Res. (India)*, vol. 17A, pp. 40–45; December, 1958 (Supplement, 1959).
[9] A. Fredriksen and R. P. Dyce, "Ionospheric absorption investigations at Hawaii and Johnston Island," *J. Geophys. Res.*, vol. 65, pp. 1177–1181; April, 1960.
[10] B. Lusignan, "Cosmic noise absorption measurements at Stanford, California and Pullman, Washington," *J. Geophys. Res.*, vol. 65, pp. 3895–3902; 1960.
[11] W. R. Steiger and J. W. Warwick, "Observations of cosmic radio noise at 18 Mc in Hawaii," *J. Geophys. Res.*, vol. 66, pp. 57–66; 1961.
[12] H. Bremmer, "Terrestrial Radio Waves—Theory of Propagation," Elsevier Publishing Co., Inc., New York, N. Y.; 1949.
[13] H. G. Booker, "Application of the magneto-ionic theory to radio waves incident obliquely upon a horizontally-stratified ionosphere," *J. Geophys. Res.*, vol. 54, pp. 243–274, 1949.
[14] D. K. Bailey, "On a new method for exploring the upper atmosphere," *Terr. Mag. Atm. Elec.*, vol. 53, pp. 41–50; 1948.
[15] S. Weisbrod and L. J. Anderson, "Simple methods for computing tropospheric and ionospheric refraction effects on radio waves," PROC. IRE, vol. 47, pp. 1770–1777; October, 1959.
[16] S. Weisbrod and L. Colin, "Refraction of VHF signals at ionospheric heights," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-8, pp. 107–109; January, 1960.
[17] H. Carru, R. Gendrin, and M. Reyssat, "La refraction ionospherique sur 20, 40 et 108 Mhz et son application a l'effet Doppler des satellites," *Space Research, Proc. 1st Internat'l Space Science Symp.*, Kallman, Ed., North-Holland Publishing Co., Amsterdam, pp. 286–303; 1960.
[18] V. Counter, "Calculations of Ground-Space Propagation Effects," Lockheed Missile Systems Div., Palo Alto, Calif., Rept. No. LMSD-2461; 1958.
[19] F. G. Smith, "Ionospheric refraction of 81.5 Mc/s radio waves from radio stars," *J. Atmos. Terr. Phys.*, vol. 2, pp. 350–355; 1952.
[20] M. M. Komesaroff, "Ionospheric refraction in radio astronomy, I. Theory," *Aust. J. Phys.*, vol. 13, pp. 153–167; 1960.
[21] R. W. Lowen, "Ionospheric refraction, I: Analytic solution for an ionosphere almost spherically stratified," *J. Geophys. Res.*, vol. 67, pp. 2339–2345; 1962.
[22] V. V. Vitkevich, "Investigation of ionospheric non-uniformities by radio-astronomical methods," *Radiotekh. i Elektron.*, vol. 3, pp. 478–486; 1958.
[23] —— and I. L. Kokurin, "Measurement of phase and amplitude fluctuations of radio waves which have traversed the ionosphere," *Radiotekh. i Elektron.*, vol. 3, pp. 1373–1378; 1958.
[24] J. P. Wild, K. V. Sheridan and A. A. Neylan, "An investigation of the speed of the solar disturbances responsible for type III radio bursts," *Aust. J. Phys.*, vol. 12, pp. 369–398; 1959.
[25] R. S. Lawrence and J. L. Jespersen, "Refraction effects of large-scale ionospheric irregularities observed at Boulder, Colorado," *Space Research, Proc. 2nd Internat'l. Space Science Symp.*, North-Holland Publishing Co., Amsterdam, pp. 277–284; 1961.
[26] F. H. Hibberd, "The effect of the ionosphere on the Doppler shift of radio signals from an artificial satellite," *J. Atmos. Terr. Phys.*, vol. 12, pp. 338–340; 1958.
[27] K. Weekes, "On the interpretation of the Doppler effect from senders in an artificial satellite," *J. Atmos. Terr. Phys.*, vol. 12, pp. 335–338; 1958.
[28] W. J. Ross, "The determination of ionospheric electron content from satellite Doppler measurements, I: Method of analysis," *J. Geophys. Res.*, vol. 65, pp. 2601–2606; 1960.
[29] V. R. Eshleman, P. B. Gallagher and R. C. Barthle, "Radar methods of measuring the cislunar electron density," *J. Geophys. Res.*, vol. 65, pp. 3079–3086; 1960.
[30] D. H. Zacharisen, "World maps of F2 critical frequencies and maximum usable frequency factors," NBS, Tech. Note 2; April, 1959 (also supplemented by NBS Tech. Note 2-2, 1960).
[31] R. S. Lawrence, "Continental maps of four ionospheric disturbances," *J. Geophys. Res.*, vol. 58, pp. 219–222; 1953.
[32] J. S. Hey, S. J. Parsons and J. W. Phillips, "Fluctuations in cosmic radiation at radio frequencies," *Nature*, vol. 158, p. 234; 1946.
[33] J. G. Bolton and G. J. Stanley, "Observations on the variable source of cosmic radio frequency radiation in the constellation of Cygnus," *Aust. J. Sci. Res.*, vol. A1, p. 58; 1948.
[34] F. G. Smith, "Origin of the fluctuations in the intensity of radio waves from galactic sources," *Nature*, vol. 165, p. 422; 1950.
[35] C. G. Little and A. C. B. Lovell, "Origin of the fluctuations in the intensity of radio waves from the galactic sources," *Nature*, vol. 165, p. 422; 1950.
[36] H. G. Booker, J. A. Ratcliffe, and D. H. Shinn, "Diffraction from an irregular screen with applications to ionospheric problems," *Phil. Trans. Roy. Soc. (London) A*, vol. 242, pp. 579–609; 1950.
[37] E. N. Bramley, "The diffraction of waves by an irregular refracting medium," *Proc. Roy. Soc. (London) A*, vol. 225, pp. 515–518; 1954.
[38] J. A. Fejer, "The diffraction of waves in passing through an irregular refracting medium," *Proc. Roy. Soc. (London) A*, vol. 220, pp. 455–471; 1953.
[39] S. A. Bowhill, "The scattering of radio waves by an extended randomly refracting medium," *J. Atmos. Terr. Phys.*, vol. 20, pp. 9–18; 1961.
[40] A. Hewish, "The diffraction of galactic radio waves as a method of investigating the irregular structure of the ionosphere," *Proc. Roy. Soc. (London), A*, vol. 214, pp. 494–514; 1952.
[41] H. J. A. Chivers, "The simultaneous observation of radio star scintillations on different radio frequencies," *J. Atmos. Terr. Phys.*, vol. 17, pp. 181–187; 1960.
[42] R. S. Lawrence, J. L. Jespersen, and R. C. Lamb, "Amplitude and angular scintillations of the radio source Cygnus-A observed at Boulder, Colorado," Res. *NBS J. (Radio Propagation)*, vol. 65D, pp. 333–350; 1961.
[43] C. G. Little and A. Maxwell, "Fluctuations in the intensity of radio waves from galactic sources," *Phil. Mag.*, vol. 42, pp. 267–278; 1951.
[44] C. G. Little, G. C. Reid, E. Stiltner, and K. P. Meritt, "An experimental investigation of the scintillations of radio stars observed at frequencies of 223 Mc/s and 456 Mc/s from close to the auroral zone," *J. Geophys. Res.*, vol. 67, 1763–1784, 1962.
[45] J. R. Koster and R. W. Wright, "Scintillation, spread F, and transequatorial scatter," *J. Geophys. Res.*, vol. 65, pp. 2303–2306; 1960.
[46] H. J. A. Chivers, "Observed variations in the amplitude scintillations of the Cassiopeia (23N5A) radio source," *J. Atmos. Terr. Phys.*, vol. 19, pp. 54–64; 1960.
[47] R. Parthasarathy and G. C. Reid, "Signal strength recordings of the satellite 1958 δ2 (Sputnik III) at College, Alaska," PROC. IRE (*Correspondence*), vol. 47, pp. 78–79; January, 1959.
[48] B. H. Briggs, "The correlation of Radio Star Scintillations with Geomagnetic Disturbances," *Geophys. J.*, vol. 5, pp. 306–317; 1961.
[49] H. J. A. Chivers and R. D. Davies, "A comparison of radio star scintillations at 1390 Mc/s and 79 Mc/s at low angles of elevation," *J. Atmos. Terr. Phys.*, vol. 24, pp. 573–584; 1962.
[50] J. P. Wild and J. A. Roberts, "The spectrum of radio-star scintillations and the nature of irregularities in the ionosphere," *J. Atmos. Terr. Phys.*, vol. 8, pp. 55–75; 1956.
[51] W. A. S. Murray and J. K. Hargreaves, "Lunar radio echoes and the Faraday effect in the ionosphere," *Nature*, vol. 173, p. 944; 1954.
[52] J. V. Evans, "The electron content of the ionosphere," *J. Atmos. Terr. Phys.*, vol. 11, pp. 259–271; 1957.
[53] G. N. Taylor, "The total electron content of the ionosphere during the magnetic disturbance of November 12–13, 1960," *Nature*, vol. 189, pp. 740–741; 1961.
[54] K. C. Yeh and V. H. Gonzalez, "Note on the geometry of the earth magnetic field useful to Faraday effect experiments," *J. Geophys. Res.*, vol. 65, pp. 3209–3214; 1960.
[55] F. B. Daniels, "Scientific uses of earth satellites," in "Electromagnetic Propagation Studies with a Satellite Vehicle," J. A. Van Allen, Ed., University of Michigan Press, Ann Arbor, Mich., ch. 30; 1956.
[56] C. G. Little and R. S. Lawrence, "The use of polarization fading of satellite signals to study the electron content and irregularities in the ionosphere," *J. Res. NBS*, vol. 64D, pp. 335–346; 1960.
[57] S. A. Bowhill, "The Faraday rotation rate of a satellite radio signal," *J. Atmos. Terr. Phys.*, vol. 13, pp. 175–176; 1958.
[58] O. K. Garriott, "The determination of ionospheric electron content and distribution from satellite observations, Part 1: Theory of the analysis," *J. Geophys. Res.* vol. 65, pp. 1139–1150; 1960.
[59] R. A. Hill and R. B. Dyce, "Some observations of ionospheric Faraday rotation on 106.1 Mc/s," *J. Geophys. Res.*, vol. 65, pp. 173–176; 1960.
[60] O. K. Garriott, "The determination of ionospheric electron content and distribution from satellite observations, Part 2: Results of the analysis," *J. Geophys. Res.*, vol. 65, pp. 1151–1157; 1960.
[61] K. C. Yeh and G. W. Swenson, "Ionospheric electron content and its variations deduced from satellite observations," *J. Geophys. Res.*, vol. 66, pp. 1061–1067; 1961.
[62] T. G. Hame and W. D. Stuart, "The electron content and distribution in the ionosphere," PROC. IRE (*Correspondence*), vol. 48, pp. 364–365; March, 1960.

Report No. 6

RELATIVISTIC CLOCK ERRORS

CONTENTS

SECTION		PAGE
1.0	Introduction and Statement of the Problem	6-1
2.0	Summary of Findings	6-1
3.0	Perturbations Which Can Be Ignored	6-2
	3.1 Doppler Shift	6-2
	3.2 Equations of Motion of Satellites	6-3
4.0	Relativity Discussion	6-4
	4.1 Background	6-6
	4.2 Alternate Interpretation for the GPS	
5.0	General Discussion of Clocks	6-6
	5.1 Clock Definition	6-6
	5.2 Gravitational Effects	6-7
	5.3 Earth-Bound Clocks	6-8
	5.4 One Clock in Orbit	6-8
	5.5 Relativistic Geopotential	6-9
	5.6 Relativistic Velocity Term	6-11
6.0	Computation of Clock Shifts; Ground Stations	6-14
7.0	Summary of Ground Station Effects	6-15
8.0	Relativistic Clock Shift For Satellites	6-15
9.0	Circular Orbit Considerations	6-17
10.0	Clock Shift For A TRACE-Computed Orbit	6-19
	10.1 Outline of Calculations	6-19
	10.2 Discussion of Computed Results	6-20
11.0	Solar Perturbation	6-22
	11.1 Introduction	6-22
	11.2 Outline of the Calculation	6-22
	11.3 Interpretation of Calculation	6-27
	11.4 Lunar Perturbations	6-27
	11.5 Consideration of Other Orbits	6-27
	11.6 Prediction of the Relativity Shift	6-31

CONTENTS (Cont.)

SECTION		PAGE
12.0	Modelling the Clock Shift	6-31
	12.1 Introduction	6-31
	12.2 Worst-Case Variations	6-33
	12.3 Initial Phase Angles	6-33
	12.4 Stair-Step Approximati. n	6-39
	12.5 Straight-Line-Segment Approximation	6-40
	12.6 Summary	6-41

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Relative Geometry of Sun, Satellite and Ground Station	6-23
2	Two Orientations of Satellite Plane Relative to Earth-Sun Line	6-29
3	Showing How the Satellite-Sun Distance For an Inclined Orbit Can Become Essentially Constant for a Day	6-30
4	Components of the Clock Rate Variation	6-34
5	Possible General Forms of Integrated Time Difference	6-35
6a	Stair-Step Approximation	6-38
6b	Straight-Line-Segment Approximation	6-38

TABLES

TABLE		PAGE
1	Relativistic Geopotential Factor	6-10
2	Relativistic Velocity Term, Squared, Earth-Bound Points	6-11
3	General Relativity Clock Slip Between Ground Stations Time Slip (Picoseconds/Second)	6-12
4	General Relativity Clock Slip Between Ground Stations Time (In Seconds) to Slip 1 Nanosecond	6-13
5	General Relativity Clock Slip Between Ground Stations Time (In Days) to Slip 1 Chip	6-13
6	Orbital Data, 12-Hour Circular Orbit	6-21
7	Physical Constants	6-25
8	Summary of Bit Counts For Eq. (26) Parameters	6-36
9	Bit Count For Segmented Approximation to Eq. (27)	6-40

1.0 INTRODUCTION AND STATEMENT OF THE PROBLEM

In a high-precision multiple-satellite navigation system which involves moving spacecraft and a scatter of fixed earth-bound monitor stations, as well as users employing the system to obtain navigation fixes, it is important to determine whether or not there are any general - or special - relativity effects large enough to affect system performance (ultimate user navigation accuracy), either on an instantaneous basis, or integrated over a finite time span. This question has been explored for the 12-hour GPS navigation satellite orbits, and for the apparent relativistic time shift between clocks located at seven representative locations (SCF station locations). Analysis has revealed that there are indeed certain relativistic effects in the GPS, of a magnitude too large to be ignored. This document summarizes the findings to date.

2.0 SUMMARY OF FINDINGS

The following conclusions constitute the major findings of this analysis; the conclusions will be substantiated in subsequent sections.

a. Although certain relativity effects do in fact occur in the GPS, and are too large to be ignored, the system as a whole may be described as being only very slightly relativistic, that is, relativity effects constitute only minor perturbations of various system characteristics. Several relativistic effects have been sized and are adjudged to be ignorable.

b. The only relativity effect of significance in the GPS is concerned with the apparent rate difference between two identical clocks, located either at different places on the earth's surface, or one clock on the earth, the other clock in orbit. This phenomenon will be carefully explained and analyzed below.

c. For the 12-hour GPS orbits, the relative velocity between any GPS satellite and any ground station is too small to require the relativistic Doppler-shift formula; the usual (ie, non-relativistic) Doppler formula is adequate for all purposes.

d. Satellite orbits may be safely computed using non-relativistic equations of motion; it is not necessary, for example, to consider the relativistic mass change of a satellite, or to distinguish between longitudinal and transverse mass.

e. The sun is the only extra-terrestrial object producing a GPS relativity effect large enough to be considered. The solar relativity perturbation on apparent clock rates is not large, but it also cannot be ignored.

f. The relativity effect on the apparent rate of an orbiting clock, deduced from characteristics of Rf signals radiated by the satellite and received on the ground, can be accurately pre-computed after the clock (ie, its satellite) orbit has been computed. In this sense the relativity effect may be considered to be highly deterministic.

g. One clock in the system - and only one clock - must be defined to maintain system reference time. In general all other clocks, whether they are in satellites, at fixed ground points, or in low-altitude (ie, user) moving positions, will appear to run at different rates, perhaps faster (orbiting clocks) or perhaps slower. In any case the apparent rate difference can be deterministically computed from knowledge of the clock location.

h. The apparent relativistic clock rate difference can be lump-sum included in an overall clock offset, rate and phase model, which will also include the best-available estimates of non-relativistic clock variations.

i. Over any one-hour period the relativity contribution to clock rate and phase can be approximated by straight-line time-dependent variations; higher-order terms (viz, t^2, t^3, etc) are not required. This conclusion includes both terrestrial and solar perturbations.

3.0 PERTURBATIONS WHICH CAN BE IGNORED

3.1 Doppler Shift

When a signal source (transmitter) is moving relative to a receiver, the formula for the Doppler frequency shift, including special relativity considerations, is slightly more complicated that the usual formula. In fact, considering special relativity,

$$f_D = f_t \beta/(1-\beta^2)^{1/2}, \qquad (1)$$

The denominator of Eq (1) may be expanded by the binomial theorem to yield the expression

$$f_D = f_t (\beta + 1/2\,\beta^3 + 3/2\,\beta^5 + \text{---}), \qquad (2)$$

which simplifies to

$$f_D = \beta f_t + 1/2\,\beta^3 f_t \qquad (3)$$

if all terms involving β^5 and higher-order are ignored.

In the case of 12-hour orbits the maximum radial velocity between a satellite and ground stations is less than about 5500 ft/sec, or 1.675 * 10^3 m/sec. For such orbits $\beta < (1.675 * 10^3/3 * 10^8)$, or 0.56 * 10^{-5}, and $\beta^3 < 1.75 * 10^{-16}$. Thus, β^3 is about eleven (11) orders of magnitude smaller than β, which implies that the relativistic correction to an observed Doppler shift is about 11 orders smaller than the shift itself. Now, if it is a system objective to observe and measure a change in radial velocity as small as 0.01 ft/sec, and if the maximum such radial velocity is 5.5 * 10^3 ft/sec, the system must resolve to one part in 5.5 * 10^5. Therefore, a correction 11 orders lower than the Doppler shift itself is safely ignorable.

To summarize, Doppler-shift measurements on 12-orbits need not be corrected for special relativity effects.

3.2 Equations of Motion of Satellites

In accordance with principles of Newtonian (ie, non-relativistic) mechanics, the motion of a satellite is computed by integrating the equation of motion $\vec{f} = m\vec{a}$, if it is assumed the mass is not deliberately being varied (such as by ejecting rocket engine effluent to vary the orbit). According to special relativity it is important to recognize that the mass m is not really a constant, but varies with velocity (as determined in an initial coordinate system centered at earth's center of mass). More precisely,

$$m = m_o/(1-\beta^2)^{1/2}$$

or

$$m = m_o(1+\beta^2/2) \text{ for } \beta << 1,$$

where m_o is the rest mass of the satellite. Now, according to paragraph 3.1, an upper bound for β^2 is about $(0.56 * 10^{-5})^2 = 3.15 * 10^{-11}$. Therefore, the relativistic mass change never exceeds about 10^{-10} - 10^{-11} of the overall satellite mass.

In the equation of motion the acceleration produced by a given force (the earth's gravitational force on a satellite, for example) will vary if the mass varies relativistically. But it can be demonstrated that the gravitational force vector on a satellite, resulting from the earth's mass distribution, is uncertain in magnitude by perhaps 1 part in 3 or 4 million; in other words, the dominant spherical-earth term of the earth's geopotential expression is uncertain to this degree. For any position of the satellite, therefore, the magnitude of the force vector acting on the satellite cannot be stated more accurately than about 3 parts in 10^7, or 3000 parts in 10^{10}.

It is therefore meaningless to include a relativity perturbation of less than 1 part in 10^{10}; this perturbation is lost in the uncertainty of the gravitational forces. Consequently it is entirely adequate to compute the orbits using non-relativistic kinematics.

4.0 RELATIVITY DISCUSSION

4.1 Background

Einstein's Principle of Equivalence states that when an object is accelerated or rotated, a force field is created which can be considered as either inertial or gravitational, depending upon the observer's choice of a reference frame. Stated in other terms, the Principle states that it is impossible to perform a measurement which could distinguish between gravitational and acceleration forces on an object; the two types of forces are equivalent.

Consider two perfect identical clocks placed side by side in a laboratory. By definition these clocks will remain exactly synchronous. Now remove one clock to a different location, where the gravitation potential, or where acceleration forces, or both, are different. Let both clocks radiate electromagnetic signals (photons), the frequencies of which represent the measured, or apparent, rates at which the clocks are running. Finally, let the clock which was moved radiate its signal back to laboratory, and there compare the clock rates, for example by beating the two signals together in a frequency or phase detector circuit.

If the principle of Equivalence is accepted as a fundamental postulate it is possible to show that the two clocks will no longer appear to run synchronously; one will appear to run ahead or behind the other. It is of utmost importance to keep in mind just what experiment is being performed; electromagnetic signals (RF carrier frequencies) are being generated by both clocks, and in the laboratory one is performing a frequency and/or phase comparison of the locally-generated signal and the signal being received from the remote clock, but only after that received signal has propagated from the remote clock back to the laboratory. One is no longer making a side by side comparison! The principle of equivalence then predicts that in general the two signals will not remain synchronous.

Clearly, in this situation, one of two possibilities exists:

a. Either the clock which was moved away from the laboratory has undergone some sort of change, resulting from the accelerations required to actually move it away from the laboratory, or from the changed gravitational potential, or else

b. The clock has not changed, but the electromagnetic signals (photons) radiated from the distant clock undergo a change (increase or decrease of frequency) while propagating back to the laboratory.

Modern relativity theory interprets the phenomenon according to the second alternate, namely, that the energy of a photon is affected by gravity, and that when a photon emanates from a location having a given gravitational potential, later to arrive at a receving point having a different potential, the frequency of the received photon is different from that at the point of emission. Furthermore, the frequency of the received photon is also affected by absolute velocities of the two points, determined in an inertial coordinate system (that is, affected in a way beyond that explained as a Doppler shift).

To a degree of approximation which is adequate for GPS purposes, the following formula may be used to compare two perfects clocks not collocated:

$$f_r/f_t = 1 + (\Phi_t - \Phi_r) + 1/2\,(V_r^2 - V_t^2) + \vec{k}\cdot(\vec{V}_t - \vec{V}_r), \qquad (4)$$

plus higner-order terms of order (Vr^3, Vt^3) and higher.

Where

f_r = frequency (ie, clock rate) at a receiving location

f_t = frequency (ie, clock rate) at a transmitting location

Φ_t, Φ_r = $\phi_t/C^2, \phi_r/C^2$

ϕ_t, ϕ_r = gravitational potential at the transmitting and receiving location, respectively

c = velocity of light

$\vec{V}_t, \vec{V}_r$ = $\vec{V_t}/c, \vec{V_r}/c$, respectively.

$\vec{v}_t, \vec{v}_r$ = vector velocities of the transmitting and receiving points, expressed in an inertial coordinate system (fixed relative to the distant stars)

$\vec{k}$ = a unit vector, directed from the transmitter toward the receiver.

4.2 Alternate Interpretation for the GPS

Although modern relativity theory interprets the frequency shift phenomenon in terms of what happens to the electromagnetic waves, we believe that it is easier to maintain that it is the clock rate itself which changes when the velocity and/or gravitational potential at the clock are varied, compared with some fixed reference clock on the earth's surface. This is asserted for the following reasons:

a. From the point of view of a GPS ground station observer, it is immaterial whether the clock in a satellite actually experiences a rate variation, or whether the clock rate remains intact, but the RF waves propagating from the satellite to the ground undergo a frequency variation on the

way down. All the observer knows is that the observed frequency of the received waves, after correction for Doppler shift, differs from the frequency of a wave generated by an identical clock at his location. The effect is small, percentagewise, but it is large enough to affect GPS operation, and it can be deterministically calculated, once the satellite orbits have been computed.

b. There are other processes, not at all concerned with relativity, which are known to literally affect the clock rate in a satellite, such as proton or electron bombardment, temperature changes, ambient magnetic field variations, and aging. One task for the Control Segment will be to model these effects so as to permit the generation of a clock prediction model for clock offset, rate and phase.

c. Since the relativity effects are small, percentagewise, and since, from a GPS operational point of view it is immaterial how the Control Segment interprets the relativity effects, it seems reasonable to adopt the interpretation that it is the space-borne clocks themselves whose rate is varying, to employ Eq (4) for computing the apparent rate difference, and to ignore the frequency variations of propagating photons. This interpretation permits the inclusion of the relativity effects simply as part of the overall spacecraft clock prediction model, in a direct straightforward manner. Therefore, in the analyses presented in subsequent sections, the relativity phenomenon is interpreted as a shift in the rate of an orbiting clock, or the rate of a clock at some surface-of-the-earth location different than that of the system reference point. Our objective is to ensure that relativity effects are correctly included in the overall GPS.

5.0 GENERAL DISCUSSION OF CLOCKS

5.1 Clock Definition

A "clock" is defined to be an oscillator-counter combination, ie, the counter counts cycles (and perhaps fractions thereof), of some sort of physical oscillator. Identical clocks means literally identical; the oscillators are built exactly alike, using the same physical principles, operate at the same temperature, pressure, magnetic environment, etc, and similarly for the counters. A Perfect clock is defined as a clock running at an absolutely, exactly steady rate, independent of local environmental conditions, excepting variations of the local gravitational potential and/or accelerations.

The rate maintained by a perfect clock depends upon the gravitational potential at the location of the clock, and upon the resultant of any acceleration forces on the clock, as measured in an inertial reference system. Two clocks which are observed to run at exactly the same rate when collocated will not run at the same rate if they are separated,

such that the two experience different gravitational potential, or are affected by different accelerations.

5.2 Gravitational Effects

Consider two identical perfect clocks sitting side by side on a laboratory bench on earth's surface. The clocks are gravitationally attracted to the earth, and this attraction is, in general, partially counteracted by the centrifugal force resulting from the earth's rotation. It is straigthforward to show that, at the equator, the centrifugal force is very close to 1/291 of the gravitational force. Thus, if the earth stopped rotating upon its axis the measured weight of a 291-pound object would increase to 292 pounds. Conversely, at either of the earth's poles the centrifugal force is zero, implying that the measured weight of an object would be insensitive to changes in earth's rotation. (It is implied the measurement is performed by an instrument not affected by the centrifugal force.)

Gravitational potential is frequently, but quite erroneously, equated with the acceleration of gravity at a given point. The acceleration of gravity is a vector quantity, ie, at any point on or above earth's surface the acceleration has both a magnitude and a direction. The potential, on the other hand, is a scalar quantity having magnitude only; the potential is a quantity whose mathematical gradient is the acceleration vector. For a spherical earth the potential is inversely proportional to distance from earth's center, but the acceleration is inversely proportional to the square of this distance. It is the potential which is important in relativity theory, but it is the acceleration which is paramount to orbit calculations.

The gravitational potential at any point in the vicinity of the gravitationally attracting body is defined as the work expended upon a test object of unit mass if that object is brought from a rest condition (zero velocity) an infinite distance from the attracting body, to a rest condition at the point in question. Since, very clearly, if a test object is moved toward an attracting body from a long distance away, work is not done on the object, but in fact the object becomes a source of work (energy), the gravitational potential is a negative quantity, numerically.

From point to point over the earth's surface the gravitational potential in general varies, largely as a function of latitude (because the earth is nearly an ellipsoid of revolution, and not a sphere), and to a lesser extent because of altitude and a somewhat "lumpy" distribution of mass within the earth (ie, considering higher-order deviations from a true ellipsoid of revolution). In addition, because of earth's rotation, a clock located at some arbitrary point on the earth's surface will experience a centrifugal force which modifies the apparent gravity, thereby causing the clock rate to change.

A satellite clock in orbit also experiences both a gravitational potential and an acceleration effect, both of which vary from point to point, and hence from time to time, in the orbit.

Smaller perturbations are produced upon an earth-orbiting clock because of the solar and the lunar gravitational potential.

5.3 Earth-Bound Clocks

If one clock is retained at the equator and the other clock is moved to some higher latitude, the second clock will slow down, for two relativistic reasons:

a. Because the earth is not a sphere, but is more nearly an ellipsoid of revolution, ie, elliptical in a cross-section plane containing the polar axis, surface points at latitudes away from the equator are closer to the center of the earth than equatorial points. Therefore, at higher latitudes the gravitational potential is numerically larger than at the equator.

b. Higher-latitude points are also closer to the polar axis than equatorial points. Since all surface points rotate around the polar axis at the same rate, ie, 2π radians in one sidereal day, the centrifugal (acceleration) force, which is proportional to distance from the polar axis, becomes progressively smaller as the poles are approached.

Both these effects act in the same direction to alter the running rate of an oscillator, as a function of latitude. Among the SCF ground stations, IOS is closest to the equator, and OL-5 is closest to a pole. It is shown in a later section that a clock at OL-5 will lose time relative to an identical clock at IOS, in fact, by 100 nanoseconds over a period of 0.54 day (about 13 hours).

5.4 One Clock in Orbit

If one clock is placed aboard an orbiting satellite, that clock will gain time, ie, run faster, than an earth-bound clock, again for two relativistic reasons:

a. To a first approximation the earth is, gravitationally speaking, spherically symmetric. It can be shown that the gravitational potential due to a homogeneous sphere varies inversely with the first-power of distance r from the center of the sphere. Since an earth-orbiting satellite is further from earth's center than is the surface point, correspondingly the gravitational potential is smaller. For a 12-hour circular orbit, the orbit radius is about 4.17 times the earth's equatorial radius, and therefore, the gravitational potential is 4.17 times smaller.

b. When a satellite is in a (very nearly) (exactly) circular orbit, the centrifugal force (very nearly) (exactly) counteracts the gravitational force, at all points of the orbit. By contrast, on earth's surface centrifugal force never counteracts more than about 0.344 percent of the gravitational force.

The earth is in fact not spherically symmetric; to the next degree of approximation it is an oblate spheroid, ie, an ellipsoid of revolution. Higher-order approximations are often used for high-precision orbit calculations. These effects - often associated with the oblateness, pear shape, and "lumpiness" of the earth only serve to introduce numerical complexity into the computation of orbits and clock rates; the central facts remain that the gravitational potential acting on an earth-orbiting clock is substantially lower than at the surface, and that centrifugal force very nearly or completely counteracts gravity (for nearly-circular orbits).

5.5 Relativistic Geopotential

Eq (4) contains a Φ term and a V^2 term which do indeed yield non-negligible effects. The geopotential term has the form

$$\Phi = \phi/c^2, \tag{5}$$

wherein ϕ is the gravitational geopotential at a point in space (the gradient of ϕ, ie, $\Delta\vec{\phi}$, yields the gravitational acceleration vector of a small mass). It is found that, for points on the earth's surface, Φ has a value around $7 * 1-^{10}$; this value varies by a small amount, ie, by roughly 4 or 5 parts in 10^4, from point to point around the earth. To a degree of approximation which includes only the spherical-earth and oblateness terms for earth's geopotential, Φ can be shown to have the form

$$\Phi = -(g_o r_o/c^2)\left\{\left[1 - J_2 (1 - 3\sin^2 \ell)/2R^2\right]/R\right\}, \tag{6}$$

in which

g_o = acceleration of gravity at earth's equator, after correction for earth's rotation (centrifugal force correction)

r_o = earth's mean equatorial radius

ℓ = latitude of a place on earth's surface

R = distance from earth's center of mass to the point having latitude ℓ, measured in units of r_o; J_2 = geopotential oblateness coefficient = $1.082 * 10^{-3}$.

It is easily verified that, for an earth modelled as an ellipsoid of revolution (polar axis the shorter),

$$R = (1 - 1/f) / \left[\sin^2 \ell + \cos^2 \ell \ (1 - 2/f + 1/f^2)\right]^{1/2}, \qquad (7)$$

f = oblateness factor = 298.30

1/f = (difference between polar radius and equatorial radius) ÷ (equatorial radius)

At the earth's equator the mean value for g , after correction for rotation, is 9.81360 m/sec². Earth's mean equatorial radius r is 6.37839 * 10^6 m, and c = 2.99792 * 10^8 m/sec, all to only six significant digits. It is therefore easily verified that the following values of Φ are obtained for the seven SCF stations, the STC (Sunnyvale) and Holloman AFB (White Sands, NM); stations are listed in Table 1 in order of increasing (less negative) Φ :

TABLE 1
RELATIVISTIC GEOPOTENTIAL FACTOR

STATION	Φ
ISZ*	-6.99380 E - 10
KTS	-6.98578 E - 10
NHS	-6.96603 E - 10
STC	-6.97371 E - 10
VTS	-6.97206 E - 10
HOL**	-6.97119 E - 10
HTS	-6.96558 E - 10
GTS	-6.96281 E - 10
IOS	-6.96111 E - 10

* A fictitious station, Ice Station Zebra, having a latitude close to OL-5.

** Holloman AFB.

5.6 Relativistic Velocity Term

In Eq (4), for a point on the earth's surface, the velocity $\vec{V} = \vec{v}/c$ is the velocity in an inertial reference frame. For a uniformly rotating earth the magnitude of this velocity is simply $v = r\omega_s$ where ω_s is the sidereal rotation rate, 2π radians per 23h 56m 4.09s, and $r = Rr_o$ = distance from earth's center to a surface point having latitude ℓ. Therefore,

$$V^2 = (\omega_s r_o R \cos \ell)^2/c^2; \tag{8}$$

For the same stations listed in Table 1, V^2 is listed in Table 2, following:

TABLE 2
RELATIVISTIC VELOCITY TERM, SQUARED, EARTH-BOUND POINTS
Eqns (7) and (8)

STATION	V^2
ISZ	0.129722 E - 12
KTS	0.680366 E - 12
NHS	1.28[illegible] E - 12
STC	1.51511 E - 12
VTS	1.62935 E - 12
HOL	1.68969 E - 12
HTS	2.07988 E - 12
GTS	2.27280 E - 12
IOS	2.39104 E - 12

TABLE 3

GENERAL RELATIVITY CLOCK SLIP BETWEEN GROUND STATIONS

Reference Station	Field Station: NHS	ISZ	GTS	HTS	KTS	IOS	VTS	HOL	STC
	TIME SLIP (PICOSECONDS/SECOND)								
NHS	.00	-1.10	.93	.75	-.57	1.04	.32	.38	.22
ISZ	1.10	.00	2.03	1.85	.53	2.14	1.42	1.48	1.32
GTS	-.93	-2.03	.00	-.18	-1.50	.11	-.60	-.55	-.71
HTS	-.75	-1.85	.18	.00	-1.32	.29	-.42	-.37	-.53
KTS	.57	-.53	1.50	1.32	.00	1.61	.90	.95	.79
IOS	-1.04	-2.14	-.11	-.29	-1.61	.00	-.71	-.66	-.82
VTS	-.32	-1.42	.60	.42	-.90	.71	.00	.06	-.11
HOL	-.38	-1.48	.55	.37	-.95	.66	-.06	.00	-.16
STC	-.22	-1.32	.71	.53	-.79	.82	.11	.16	.00

Positive Slip Means Field Station Runs Faster Than Reference Station
Negative Slip Means Field Station Runs Faster Than Reference Station

TABLE 4

GENERAL RELATIVITY CLOCK SLIP BETWEEN GROUND STATIONS
TIME (IN SECONDS) TO SLIP 1 NANOSECOND

Reference Station \ Field Station	NHS	ISZ	GTS	HTS	KTS	IOS	VTS	HOL	STC
NHS	--	-910	1077	1337	-1746	963	3077	2620	4601
ISZ	910	--	493	541	1901	468	702	675	760
GTS	-1077	-493	--	-5540	-666	9056	-1657	-1829	-1406
HTS	-1337	-541	5540	---	-757	3437	-2364	-2730	-1884
KTS	1746	-1746	-1901	666	757	--	620	1048	1266
IOS	-963	-468	-9056	-3437	-620	--	-1401	-1522	-1217
VTS	-3077	-702	1657	2364	-1114	1401	--	17608	-9293
HOL	-2620	-675	1829	2730	-1048	1522	-17608	--	-6082
STC	-4601	-760	1406	1884	-1266	1217	9293	6082	--

TABLE 5

GENERAL RELATIVITY CLOCK SLIP BETWEEN GROUND STATIONS
TIME (IN DAYS) TO SLIP 1 CHIP

Reference Station \ Field Station	NHS	ISZ	GTS	HTS	KTS	IOS	VTS	HOL	STC
NHS	--	-1.05	1.25	1.55	-2.02	1.11	3.56	3.03	5.33
ISZ	1.05	--	.57	.63	2.20	.54	.81	.78	.88
GTS	-1.25	-.57	--	-6.41	-.77	10.48	-1.92	-2.12	-1.63
HTS	-1.55	-.63	6.41	--	-.88	3.98	-2.74	-3.16	-2.18
KTS	2.02	-2.20	.77	.88	--	.72	1.29	1.21	1.46
IOS	-1.11	-.54	-10.48	-3.98	-.72	--	-1.62	-1.76	-1.41
VTS	-3.56	-.81	1.92	2.74	-1.29	1.62	--	20.38	-10.76
HOL	-3.03	-.78	2.12	3.16	-1.21	1.76	-20.38	--	-7.04
STC	-5.33	-.88	1.63	2.18	-1.46	1.41	10.76	7.04	--

6.0 COMPUTATION OF CLOCK SHIFTS; GROUND STATIONS

Eq (4) may be used to compute the time-slip (ie, integrated time difference) of two perfect clocks located at different ground points, as follows:

Let identical oscillators run at two points (r,t), and count cycles of oscillation. Define apparent time T (viz, station-measured time) as the total count N times the local oscillator frequency. Then

$$T_r = Nf_r \, , \; T_t = Nf_t$$

if both clocks are counted the same number of counts. Then, if the right-side of Eq (4) is written as $1 + \delta$, and ignoring the Doppler-shift term, it follows that

$$f_r = T_r/N = f_t\,(1+\delta) = T_t\,(1+\delta)/N \; ,$$

or, if

$$\Delta T = T_r - T_t \, ,$$
$$\Delta T = \delta\, T_t$$

and

$$\delta = (\Phi_t - \Phi_r) + 1/2 \left(V_r^2 - V_t^2 \right) \qquad (10)$$

(9)

Stated in simple terms, if a perfect clock at station t runs for an elapsed "time" at that station, as determined by an oscillator count N, another clock at station r will be ahead of, or behind, as δ is positive or negative, the t clock, when station r elapses through the same number N of oscillator cycles.

Tables 1 and 2 may be used to compare pairs of perfect clocks located at the listed ground stations. Either station may arbitrarily be selected as the reference, and the other station is then to be considered the remote station. Table 3 is such a comparison; the tabulation indicates the time slip in pico-seconds per second (1 pico-second = 10^{-12} second)

The data from Table 3 may be used to compute the length of time required for the remote station to advance (positive) or fall behind (negative), relative to the reference station, by a prescribed amount of time. Tables 4 and 5 list the elapsed times for a slip of 1 nanosecond (10^{-9} second) and 100 nanoseconds (1 chip at a 10 MHz rate), respectively.

7.0 SUMMARY OF GROUND STATION EFFECTS

Tables 3 through 5 show very clearly that relativity effects in the GPS are definitely not ignorable. As a representative example from Table 5, for instance, consider VTS to be the reference station. At one extreme, the ISZ clock will fall behind the VTS clock, slipping a full 100 ns chip in 0.81 day (19h 27m), or 9.65 chips per week, or 36.1 chips in a 30-day month. This slip rate is at least as large as, and may be several times larger than the non-relativistic drift rate of high-quality cesium-beam clocks. At the other extreme, if Holloman AFB is the remote clock, Table 5 shows that the HOL clock will run very slightly faster than the reference, in fact, requiring almost three weeks (20.38 days) to slip a full 100 nsec chip; but from Table 4 the slip amounts to 1 nsec in only 17608 seconds (4h, 53m, 28s), which implies that, even in this best case, a 1 nanosecond adjustment should be allowed for several times a day, every day.

It is important to emphasize that relativistic clock slippages have nothing whatever to do with clock imperfections; this effect is a property of nature only. In the GPS, relativistic clock slippage will show up as systematic timing errors, if the effect is ignored. On the other hand, the effect is highly, but not exactly, predictable, hence can be included in the system software. The effect is not exactly predictable, not because relativistic theory is inadequate, but rather because:

a. Numerical values of the various physical constants which enter the calculations contain uncertainties.

b. The model of the geopotential contains small, but finite, residual errors.

c. Station altitude should be, but has not yet been, included in the model.

8.0 RELATIVISTIC CLOCK SHIFT FOR SATELLITES

For a satellite (ie, a clock) in a 12-hour quasi-circular orbit the clock shift relative to a ground station is very substantially larger than the maximum occurring between two ground-based clocks, for two reasons:

a. A 12-hour circular orbit implies that the clock is about 16520 st. mi. from earth's center, which is very nearly 4.1768 equatorial earth radii. Therefore, the gravitational potential, given by Eq (6), is very substantially different (ie, less negative) than for a ground station.

b. For a circular satellite orbit the centrifugal force exactly balances the pull of gravity, whereas for an earth station the centrifugal force never amounts to more than about 0.3 percent (ie, 1/291) of the pull of gravity. In fact, for a circular orbit around a spherical earth, V^2 and Φ terms in Eqns (4) and (10) are exactly equal and oppcsite; but in those equations one must combine Φ with $-V^2/2$.

Here we compute only the approximate magnitude of the satellite-earth clock shift, comparing a 12-hour circular-orbit clock against a single ground station, VTS. Further refinements of this shift rate, and variations during the 12-hour cycle, will be discussed in a subsequent section.

In Eq (6) let R = 4.1768, and ignore the oblateness factor, $J_2/2R^2$. Then Φ = -1.66746 E -10. For a satellite at 16520 st. mi. above earth's center the in-track velocity is computed to be v = 3866.83 m/sec, so that V = v/c = 1.29011 E - 05, or V^2 = 1.66746 E - 10, exactly equal to Φ.

Now, in Eq (10) let t refer to the satellite and r to a receiver at VTS; values of Φ and V^2 for VTS are contained in Tables 1 and 2, respectively. Therefore, in Eq. (10)

$$\delta = (-1.66746\ \text{E-10} + 6.97206\ \text{E-10}) + 0.5\ (1.62935\ \text{E-12} - 1.66746\ \text{E-10})$$

or $$\delta = +4.47902\ \text{E-10} = +\ 447.902\ \text{E-12}.$$

The satellite clock runs faster than the VTS clock, slipping by 447.902 picoseconds per second, or 1 nsec in 2.23 seconds, or 100 nsec = 1 chip in 223 seconds (3h 43m). This slip rate is very much larger than the slip resulting from cesium-beam clock imperfections, and therefore, very much needs to be accounted for as a systematic correction.

9.0 CIRCULAR ORBIT CONSIDERATIONS

If a gravitationally-attracting body is spherically symmetric it is easily demonstrated that its gravitational potential is given by the relation

$$\phi = -GM/r \tag{11}$$

and the force-of-gravity vector $\bar{f}$ on a test object is given by

$$\bar{f} = -\bar{r}\, G\, Mm/r^2 \tag{12}$$

C = universal constant of graviation

M = mass of the attracting body

m = mass of test object

r = distance from the center of the attracting body

$\bar{r}$ = a unit vector, having the direction of increasing r, ie, directed away from the center of the attracting body.

If the test object is allowed to fall freely from infinity to a finite distance r the work W done on the object is calculated as follows:

$$dW = \bar{f} \cdot d\bar{r} = (-\bar{r}\, GM/r^2) \cdot (-\bar{r}\, mdr) = GMmdr/r^2,$$

and the total work W is the integral of the above, from infinite distance to r, ie

$$W = \int_{\infty}^{r} (GMm/r^2)\, dr : -GMm/_r \equiv \phi\ (r). \tag{13}$$

The negative sign simply indicates that the test object has lost potential energy; in the process of so doing the object gains kinetic energy of motion, $1/2\ mv_s^2$, computed as follows (v = free-fall velocity at distance r):

$$\bar{f} = ma = md^2\bar{r}/dt^2 = \bar{r}\, GMm/r^2.$$

Therefore,

$$d^2r/dt^2 = GM/r^2.$$

Multiply through by 2 dr/dt to obtain

$$2\,dr/dt\;\; d^2r/dt^2 = \frac{d}{dt}(dr/dt)^2 = -(2\,GM/r^2)\,dr/dt$$

a further multiplication by dt yields

$$d\,(dr/dt)^2 = -2\,GM\,dr/r^2 = +2\,GM\,d\left(\frac{1}{r}\right),$$

so that both sides are total differentials. Integration between the limits of infinity and r yields the result

$$\left(dr/dt\right)^2 = 2\,GM/r\,.$$

But the left side of this equation is simply the square of the free-fall velocity, v , and the right side is simply -2ϕ. Therefore, $v = -2\phi$ at any distance.

Now, if the test object is placed in a circular orbit of radius r its orbital velocity v can be computed upon equating centrifugal force (acting in the positive r direction) to the negative of gravitation force (acting in the negative r direction), ie,

$$mv_o^2/r = +(G\,Mm/r^2), \text{ or} \qquad (14)$$

$$v_o^2 = +GM/r = -\phi$$

For nearly circular orbits v_o^2 and $-\phi$ generally will not be exactly equal to all points.

This shows that

$$v_o^2 = v_f^2/2$$

for any distance r; in other words, the velocity-squared of an object in a circular orbit of radius r is exactly half the velocity-squared

that the object would have, at the same distance r, if it had fallen in from infinity.

The public press often speaks of a "zero-gravity" condition existing inside an earth-orbiting satellite (such as SKYLAB). In actuality, the zero-gravity condition - that is, exact balance of gravity and centrifugal forces - occurs only at the mass center of the satellite, and at points within the satellite having exactly the same distance from earth's center. At all other points gravity and centrifugal force do not exactly balance, and therefore, a force gradient does exist inside the spacecraft. On the other hand, the force gradient may be so small as to be considered zero for some practical purposes, but certainly not for the purpose of executing a precise calculation of the running rate of a perfect clock!

10.0 CLOCK SHIFT FOR A TRACE-COMPUTED ORBIT

10.1 Outline of Calculations

Standard reference works, such as Moeller's "Theory of Relativity", Oxford Press, 1952, may be consulted for the derivation of formulas giving the frequency of an oscillator in a gravitational/acceleration environment. In all such formulas one finds the gravitational potential and clock velocity-squared both divided by the square of the velocity of light. When these ratios are small as they are for earth-orbiting satellites approximate formulas may be used to compute clock shifts.

For an earth-orbiting satellite one may compute ϕ along the orbit, as a function of the satellite coordinates, relative to a coordinate system fixed in the rotating earth. Similarly, the spacecraft velocity v may be computed.

The geopotential of the earth is a complicated function of position relative to the earth. If the position of a spacecraft is given in terms of earth-centered latitude, longitude and distance from earth's center of mass, the potential is expressible as the summation of a series of zonal, sectoral and tesseral harmonics. Tens or even hundreds of such terms are required for high-precision orbit calculations; however, since the computed relativistic clock shift for an orbiting clock is very small, percentagewise, the geopotential expression may be greatly simplified for computing the relativity shift, as indicated by Eq (6).

TRACE 66 does not employ a geopotential expression directly for orbital computations, and ϕ is not computed for the orbit integration steps. One must take the data which TRACE does compute and print, namely satellite coordinates and velocity, and using this data perform a subsidiary evaluation of Eq (6), for each integration step.

The following technique was used to compute the clock shift for a representative 12-hour orbit:

a. A TRACE 66 run was made for the purpose of computing a trojectory over a 12-hour period, very nearly circular orbit (refer to TRACE 66 run sec 69M).

b. Velocity values at 1-hour integration intervals were printed out; certain of these numbers were used to compute V in Eq (4).

c. The geopotential Φ_t in Eq (6) was hand-computed, using radius and latitude values from the TRACE run, and using a two-term model of the earth's geopotential, ie, the spherical-earth and oblateness terms only.

d. The clock shift rate δ in Eq (4) was computed for only four (4) points of the orbit, ie, ascending node (satellite crossing the equator from south to north), maximum north latitude, descending node, and maximum south latitude. The comparison station was VTS; values of Φ_r and V for this station are provided in Tables 1 and 2.

Table 6 summarizes the orbital calculations. This table reveals three facts of significance to this memo, namely:

a. At the four points computed, v^2 is very nearly equal to $-\Phi$, in accordance with very nearly circular orbit considerations (see Eq (14), section 9).

b. The average clock shift rate between the satellite and VTS is by no means negligible; but, in fact, every second the satellite clock gains 448 picoseconds relative to VTS. This slip rate amounts to 100 nsec, or one full GPS PN chip, in 223 seconds, less than 4 minutes.

c. The clock slip rate is extremely constant, in fact varying (cyclically, if not exactly periodically) above and below the average, by one 1 part in 13532 of the mean itself.

10.2 Discussion of Computed Results

The computed satellite - VTS relativistic clock shift is very definitely large enough to require inclusion in the system software. The sample-of-one calculation indicates that the variation of the shift rate is negligibly small over a 12-hour orbit, thus, suggesting that a sufficient calculation would be concerned only with the average shift rate.

Unfortunately, if variations in the clock shift rate resulting from the sun's gravitational potential are computed it is found that the variations are not small enough to be ignored. Only a few simple computations have been performed on the solar effect; these computations make it clear that the relativity model treated in

TABLE 6

ORBITAL DATA, 12-HOUR CIRCULAR ORBIT

Ref: TRACE 66 RUN SEC 69M

POSITION	RADIUS, FEET	VELOCITY, FT/SEC	LATITUDE DEGREES	GEOPOTENTIAL** $(FT/SEC)^2$	VELOCITY-SQUARED $(FT/SEC)^2$
ASCENDING NODE ORBIT TIME = 0 HOURS	9.729535159 E07	1.269913371 E04	0	-1.612468542 E08	1.6122679970 E08
MAX. NORTH ORBIT TIME = 3 HOURS	8.730999729 E07	1.269793842 E04	63.0394	-1.612317602 E08	1.612376401 E08
DESCENDING NODE ORBIT TIME = 6 HOURS	8.732035167 E07	1.269549061 E04	-0.00000318	-1.612006918 E08	1.611754818 E08
MAX. SOUTH ORBIT TIME = 9 HOURS	8.731006945 E07	1.269793206 E04	-63.0394	-1.612316269 E08	1.612374786 E08
	$(-1/2v^2) * 10^{-10}$	CLOCK SHIFT RATE vs VTS; PICO-SEC. PER SECOND	LAPSED TIME TO SHIFT 100 N SEC, SECONDS		
ASCENDING NODE	-2.50007431	448.013	223.208		
MAX. NORTH	-2.49997434	448.023	223.171		
DESCENDING NODE	-2.49933193	448.087	223.171		
MAX. SOUTH	-2.499972125	48.023	223.203		
AVERAGE		48.037	223.196		

$$**\Phi = -\frac{GM}{r}*\left[1 - \frac{J2}{2R^2}(1-3\sin^2 \ell)\right], \quad R = r/r_e, \quad re = \text{equatorial radius}, \ell = \text{latitude}$$

$$c = 9.835711728 * 10^8 \text{ FT/SEC} \qquad J_2 = 1.083 * 10^{-3}$$

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

Section 10.1, considering the earth's geopotential variations, is by far the dominant effect - therefore, the first order of business would be a proper inclusion of this model in the overall system processing software.

In the operational system TRACE 66 or its equivalent be used to compute and predict orbits. Ordinarily TRACE does not compute geopotential, nor heliopotential, but TRACE does include the geopotential formulas. It would therefore be necessary, but simple, to modify TRACE such as to compute geopotential, and the corresponding clock shift for selected integration steps of each satellite orbit. Heliopotential should also be included in order to compute the clock shift variations over a 24-hour period.

This procedure would utilize predicted orbit position and velocity data to compute ϕ and v in Eq (4), and with the corresponding data for a selected ground reference point, say VTS, the satellite clock shift rate could then be computed. The computation should be repeated for a number of predicted orbit points, perhaps for every half-hour or one-hour of orbit position.

11.0 SOLAR PERTURBATION

11.1 Introduction

Variations in the clock shift rate due to the sun are small, but are not negligible in the GPS system. In fact, it will be shown that clock shift rate variations due to the sun amount to as much as about 1 part in 200.

In other words, the satellite clock does not gain a fixed 448 picoseconds, but can gain as much as 450.2, or as little as 445.8 picoseconds per second, depending upon instantaneous relative distances of the satellite and the earth ground point from the sun. Although such a small variation may appear at first glance to be negligible, the effect is cumulative, such that over a 6-hour period the satellite clock will be ahead or behind, by as much as about 18 nanoseconds, as compared with a computation ignoring the solar effect.

11.2 Outline of the Calculation

In Eq (4) a formula is given for the slip rate between two clocks experiencing relativistic effects.

The coordinate system is an inertial system, centered at earth's center of mass, but fixed relative to the distant stars In this system an earth station rotates around earth's polar axis at a fixed angular rate ω_e; a satellite in a circular orbit rotates at a very-nearly constant angular rate ω_t; both the earth station and the satellite experience a varying distance from the sun.

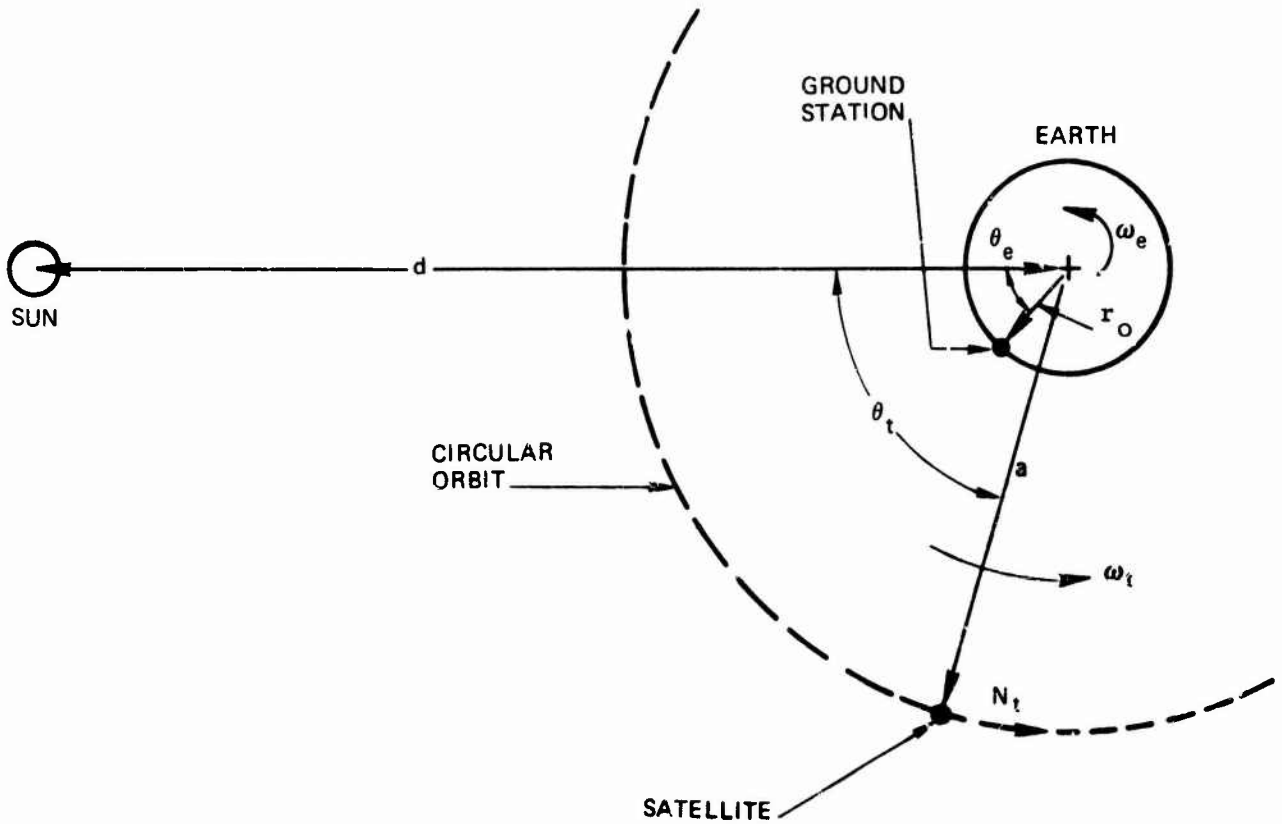


Figure 1 Relati e Geometry of Sun, Satellite and Ground Station

For the purpose of calculating the solar perturbation it is adequate to consider the earth and the sun as gravitational point masses; this approximation permits us to use only the spherical-mass-distribution term to represent geopotential and heliopotential.

The calculation is performed here under the following assumptions:

a. Satellite orbit is circular (fixed distance a from earth's center of mass)

b. The "orbit" of the ground station, rotating around earth's polar axis, the satellite orbit, and the sun all lie in the same plane (refer to Figure 1).

It will become clear that this is a worst-case calculation, but it will also be discussed that, in the GPS system this worst-case condition can be closely approached.

Upon referring to Figure 1 it can be verified that, for the satellite,

$$v_t = a\omega_t \tag{15}$$

$$\Phi_{te} = -GMe/a = \text{geopotential at the satellite,} \tag{16}$$

$$\Phi_{ts} = -GMs/(d - a\cos\Theta_t) = \text{heliopotential at the satellite,} \tag{17}$$

and for the ground station,

$$v_e = r_o\omega_e \tag{18}$$

$$\Phi_{re} = -GMe/r_o = \text{geopotential at the ground station,} \tag{19}$$

$$\Phi_{rs} = -GMs/(d - r_o\cos\Theta e) = \text{heliopotential at the ground station} \tag{20}$$

In the above expressions Me and Ms are the earth's and the sun's masses, respectively, and G is the universal gravitational constant.

In Eq (1),

$$\Phi_t = (\Phi_{te} + \Phi_{ts})/c^2,\ V_t = v_t^2/c^2 \text{ and similarly for } \Phi_r \text{ and Vr.}$$

In section 9, Eq (14), it was shown that, for a circular orbit the in-track velocity, squared, is just equal to the negative of the geopotential at the satellite; in other words combination of Eqns (15) and (16) above lead to the result

$$G = a^3 \omega_t^2 / Me; \tag{21}$$

this result may be used to eliminate G explicitly from Eqns (16), (17), (19), and (20). One may then substitute Eqns (15) - (21) into Eq (4), note that both a and r are much less than d (numerical values are tabulated in Table 7), and obtain the approximate result

$$c^2 \delta = - a \omega_t^2 \left[\begin{array}{l} 3/2 - a/r_o - 1/2 \, (ro/a)^2 \, (\omega e/\omega t)^2 \\ + (a/d) \, (Ms/Me) \left\{ \frac{a}{d} \cos \Theta_t - \frac{ro}{d} \cos \Theta_r \right\} \end{array} \right] \tag{22}$$

TABLE 7

PHYSICAL CONSTANTS

QUANTITY	NUMERICAL VALUE
a, 12-hour orbit radius	26610 Km
r_o, earth's equatorial radius	6378 Km
ω_e, earth's rotation rate	2π radians per sidereal day
ω_t, satellite rotation rate	$2 * \omega_e$ (12-hour orbit)
d, mean earth-sun distance	$1.50** 10^8$ Km
M_e, earth's mass	$5.98 * 10^{24}$ Kg
Ms, sun's mass	$1.99 * 10^{30}$ Kg
c, velocity of light	$3 * 10^8$ m/sec

DERIVED RATIOS	VALUE
a/r_o, r_o/a	4.172, 0.2397
a/d	$1.774 * 10^{-4}$
r_o/d	$4.252 * 10^{-5}$
M_s/M_e	$3.33 * 10^5$

Substitution of numerical values from Table 7 leads to the expression

$$\delta = + 448.477 * 10^{-12} \left[1 - 0.0039303 \cos \theta_t + 0.000942 \cos \theta_r\right]. \quad (23)$$

It is now noted that, because both the satellite and the ground station are moving,

$$\theta_t = \theta_{ot} + T\omega_t, \theta_e = \theta_{oe} + T\omega_e \quad (24)$$

in which T = time measured from an arbitrary zero reference, and θot, θoe are the position angles of the satellite and the ground station at T = 0. Consequently δ, computed from Eq (23), has an average value of 448 * 10a12, and possesses two small cyclic components, one have a 12-hour period, the other with a 24-hour period.

The quantity δ in Eq (23) is a dimensionless quantity which represents the relative difference of rates between two clocks, ie, one clock (in the satellite) is running about 448 parts in 10^{12} faster than the other clock (on the ground). In order to compute the net time slip between the two, one must integrate δ from an initial time T = To to some later time T_f. This time difference D_{of} is easily found to be represented by

$$D_{of} = 0.448477 (T_f - T_o) - 11.69 \left[\sin (\theta_{ot} + T_f \omega_t) - \sin (\theta_{ot} + T_o \omega_t)\right] \quad (25)$$

$$+ 5.75 \sin (\theta_{oe} + T_f \omega_e) - \sin (\theta_{oe} + T_o \omega_e) \text{ nanoseconds.}$$

with T_f and T_o expressed in seconds.

11.3 Interpretation of Calculation

The first term on the right side of Eq (25) represents a linear-increasing time difference resulting from the fact that the satellite clock is further from earth's center than the ground-based clock. This term accumulates to a time slip of 1 nanosecond every 223 seconds (3 minutes, 43 seconds) of lapsed ground station time.

The two sinusoidal terms in Eq (25) represent small variations above and below the linearly-increasing time slip. Depending upon the initial angles θot, and θoe these two sinusoids will combine inphase positive, or inphase negative, once in each 24-hour period. When the terms are inphase the clocks will differ by 11.69 + 5.75 = 17.44 nanoseconds, above or below the time difference computed from the (much larger) linearly increasing term.

For example, if To = θot = 0 and θoe = 90°, then after 9 hours (33,400 seconds) the linear term will have increased by 0.448,447 * 33400 = 14960 nanoseconds; at this time the sinusoidal terms will be inphase positive, indicating that the satellite clock is an additional 17.4 nanoseconds ahead of the ground.

Tnus, the solar effect, although considerably smaller than the linear term, nevertheless must be considered when time differences of a nanosecond or more are significant.

11.4 Lunar Perturbations

The moon also produces small, but negligible, clock shift variations. In order to demonstrate that these perturbations are in fact negligible, Eq (22) may be used again, but with the moon's mass Mm replacing the solar mass Ms, and with the average earth-moon distance dem replacing the average earth-sun distance d. Standard references list Mm = 7.35 * 10^{22} Kg, dem = 3.84 * 10^{5} Km. When these values are used in Eq (22), and the result converted into forms equivalent to Eqns (23) and (25), it is found that the sinusoidal variation terms are smaller by a factor of 178 than the terms due to solar perturbations. Therefore, in Eq (25) the linear term is all that is required; the other two terms never contribute a deviation of more than 17.44/178 = 0.098 nanosecond. For this reason lunar perturbations are considered negligible.

11.5 Consideration of Other Orbits

The analysis given above is for the special case wherein the ground station motion, the satellite motion, and the earth-sun direction line all lie in the same plane. More generally, the satellite and ground station move in different planes, neither of which contains the earth-sun direction line.

The solar relativistic effect on the GPS system is caused by the solar gravitational potential acting at the satellite and at the ground station. This effect varies cyclically, as the sun-satellite and sun-ground station distances constantly vary.

If the ground station is located at a high latitude (KTS, OL-5) then the sun-ground station distance would be more nearly constant than for a ground station at or near the equator (ICS). In this case the term in cos θr, in Eqns (12) and (23) would possess a smaller coefficient; for a station located at a pole the coefficient would vanish.

When the satellite orbit is inclined relative to earth's equator (such as the so-called SIGMA orbits, with inclination of 63°) the direction of a normal to the orbital plane relative to the earth-sun direction line, is a vital parameter. If, at one extreme, these two directions happen to coincide, then over a 24-hour period the satellite revolves around the earth at an almost constant distance from the sun. In this case the cos θt term in Eqns (22) and (23) vanishes. (Refer to Figure 2a.)

At the other extreme, if the satellite orbital plane contains the earth-sun directions line then the satellite-sun distance experiences the maximum possible variation through one satellite period. This is the condition reflected in Eqns (22) and (23), and illustrated in Figure 2b.

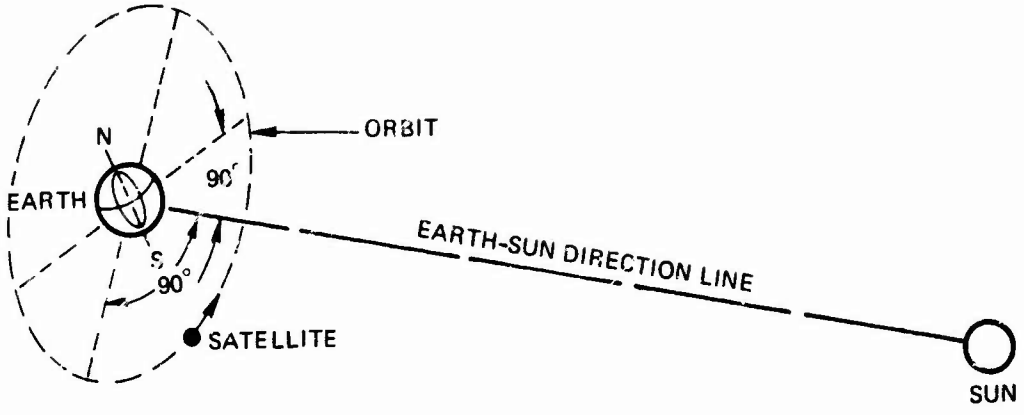
The earth's equatorial plane is tilted 23.5° relative to the earth's orbital plane around the sun (the ecliptic). If a satellite orbit is inclined 66.5° it would be possible for the orbital plane to lie exactly perpendicular to the earth-sun line. (See Figure 3). During the 24-hour period wherein this condition occurred the cos θt term in Eqns (22) and (23) would virtually vanish.

However, several weeks later, when the earth has revolved about 90° around the sun, and the satellite orbit has precessed around the earth, it would be possible for the orbital plane of the satellite to then contain, or nearly so, the earth-sun line. On this date the relativistic satellite clock rate variation, would be at a maximum.

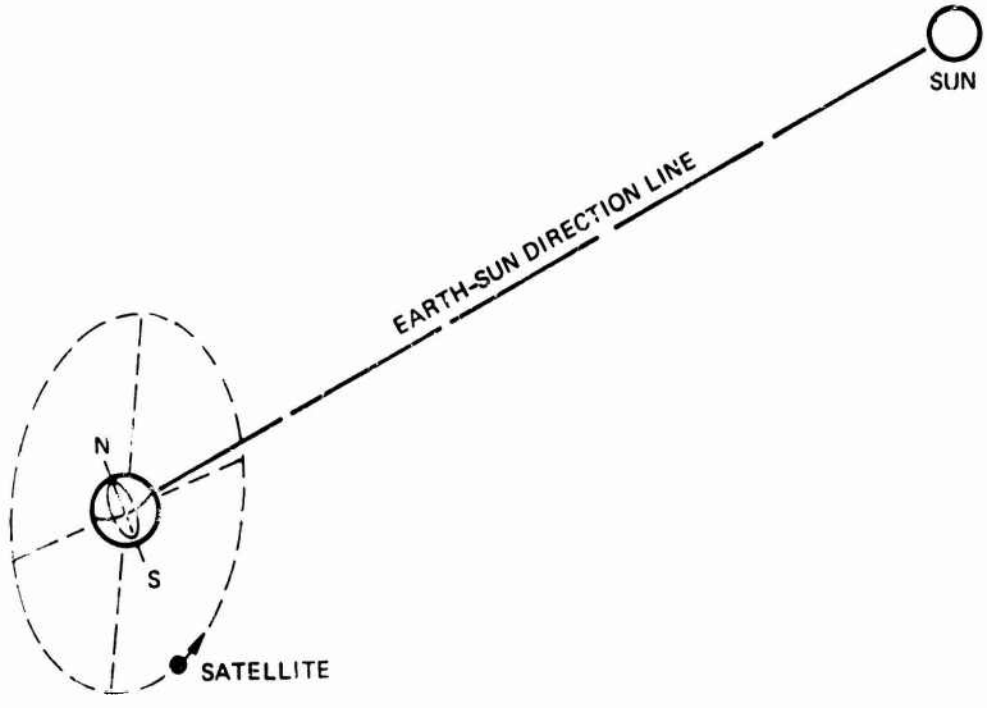
Because the SIGMA orbits are inclined 63°, very close to 66.5° the day-to-day variations explained above would be closely approached by satellites in such orbits.

In the GPS Phase I configuration satellites will revolve in two planes, separated by 60° (or 120°). In general the satellites in a given plane will experience much the same relativistic time shift variations, except that two satellites at differing phases will have the maxima and minima of their rate shifts correspondingly separated in time (that is, in Eqns (23) and (25), the initial phase θot would be satellite-peculiar).

Satellites in two different planes would not in general experience the same time shifts, because the earth-sun line would make different angles to the perpendiculars to the orbital planes. Thus, the relativistic time shift variation will be unique, although predictable, for each individual satellite.



(a) SATELLITE ORBIT PERPENDICULAR TO EARTH-SUN LINE; SUN-SATELLITE DISTANCE ESSENTIALLY FIXED



(b) SATELLITE ORBITAL PLANE CONTAINS EARTH-SUN LINE

Figure 2 Two Orientations of Satellite Plane Relative to Earth-Sun Line

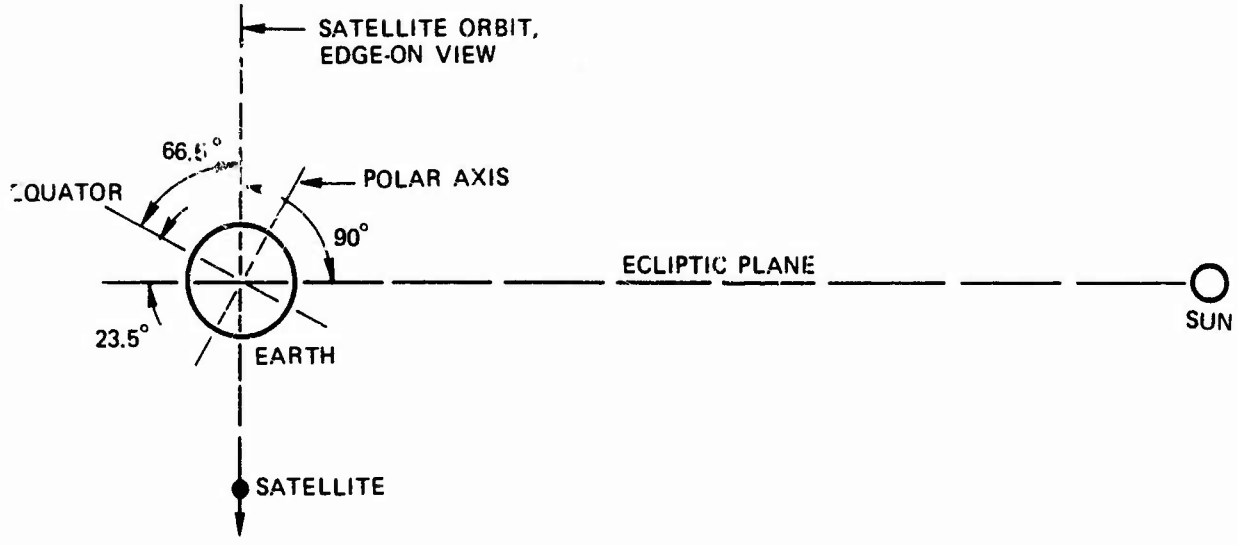


Figure 3 Showing How the Satellite-Sun Distance for an Inclined Orbit Can Become Essentially Constant for a Day

11.6 Prediction of the Relativity Shift

In the operating GPS, system computers will maintain predictions of the orbits of all satellites, and the orbital predictions will contain solar perturbation effects, as well as satellite in-track velocity. The relativistic time shift discussed in this memorandum can be handled by having the computer also compute the geopotential and heliopotential acting on both the satellite and a reference ground point, then making use of Eq (4).

This computation does not in any way change the procedure for predicting orbits. It does require relativly simple additional computation, which involves the gravitational potentials themselves, and not their gradients. On the other hand, whereas dozens of geopotential-gradient terms may be used for orbital predictions, only the first two terms (spherical earth and oblateness terms) are required for the relativity computation, as per Eq (6), and the sun may be treated as a homogenous sphere.

Unfortunately, the solar effect is large enough that it cannot be ignored; on the other hand it is fortunate that the effect is predictable, and that the solar effect is the only perturbation which must be considered (ie, lunar and planetary perturbations are ignorable).

12.0 MODELLING THE CLOCK SHIFT

12.1 Introduction

In Eq (25) it is shown that the clock rate difference between a clock in a 12-hour earth orbit and an identical clock on the ground can be expressed in the form (refer to Figure 4)

$$\delta = A\left[1 - B\cos(\Theta_{ot} + t_{\omega t}) + C\cos(\Theta_{oe} + t_{\omega e})\right] \qquad (26)$$

in which

A = a constant, approximately $448 * 10^{-12}$

B = a constant (over a 24-hour period)

C = a constant (over a 24-hour period);

It was also argued that B may be as large as about 0.0039, or that it may be very close to zero, depending upon how the satellite orbital plane is oriented relative to the earth-sun line.

Similarly, C may be as large as 0.000942 for an earth station at a low latitude, or very nearly zero for a earth station near one of earth's poles (KTS, OL-5).

In Eq (26), δ is the rate difference; the integral of δ with respect to time represents the actual indicated time difference of the two clocks.

The leading term of Eq (26) is the constant A, which is very nearly $448 * 10^{-12}$; this term represents the average rate difference. In an elapsed time T, with a given value A, the accumulated (average) time difference D is simply

$$D = AT$$

Suppose constant A is in error by an amount ϵ. Then the accumulated (average) time difference will be $D + \Delta$, ie

$$D + \Delta = (A + \epsilon)T = AT + \epsilon T = D + \epsilon T, \text{ or}$$

$$\Delta = \epsilon T, \ \epsilon = \Delta/T.$$

We may consider Δ as the (average) time difference error, due to incorrect statement of the average rate difference.

Suppose we wish to keep this error below 1 nanosecond. Then, for 1 hour or for 24 hours ϵ is computed to be

ϵ (1-hour) = 10^{-9} second/3600 seconds = $0.278 * 10^{-12}$

ϵ (24-hours) = 10^{-9} second/86400 seconds = $1.158 * 10^{-14}$

Since A is about $448 * 10^{-12}$, ϵ (1-hour) is smaller than A by 1 part in 448/0.278 = 1620 parts; correspondingly, ϵ (24-hours) is smaller than A by 1 part in 448/0.01158 = 38700 parts.

If constant A is represented by a binary number, the number must contain sufficient bits to avoid round-off errors. In order to guarantee that A will not be in error by more than 1 part in 1620 (or 38700), the number of bits for the two cases must be, respectively, 11 bits ($2^{11} = 2048$) and 16 bits ($2^{16} = 65536$).

Therefore, if Eq (26) is used to predict the relativistic clock shift for 1 hour or 24 hours, constant A must be represented by at least 11 bits, or 16 bits, respectively.

12.2 Worst-Case Variations

The worst case of the cosine terms in Eq (26) occur if the initial time is zero, and if $\theta ot = \pm 90^{\circ}$. (Refer to Figure 5.) Then

$$D = \int \delta \, dt = AT - AB \int_o^T \sin(\pm t\omega_t)\, dt + AC \int_o^T \sin(\pm t\,\omega_e)\, dt$$

$$= AT + \frac{AB}{\omega t}(1 \mp \cos(T\omega_t)) + \frac{AC}{\omega_e}(1 \mp \sin(T_{\omega e})); \qquad (27)$$

In Eq (27) the worst case of the second term occurs if $T\omega t = \pm\pi$. Since B evidently never exceeds about 1/254, and since ωt, the satellite angular rate, is $2\pi/43200$ radians per second, the magnitude of the second term never exceeds

$2 * 448 * 10^{-12} * (1/254) * 43200/2\pi = 24.25$ nanoseconds.

The uncertainty in this quantity, due to finite bit count, will be less than 1 nanosecond if $2AB/\omega t$ is expressed by 5 bits ($2^5 = 32$). Since the uncertainty in A cannot exceed 1 part in 1620, and since the period of the satellite will be known to 1 second (or much better), implying that ωt is known to at least 1 part in 43200 parts, all the uncertainty in $2AB/\omega t$ can be attributed to B. Therefore, B must be represented by at least 5 bits.

Similarly, the worst case of the third term in Eq (27) amounts to $2AC/\omega e$. Here C is evidently not larger than 1/941, and ωe, the earth's angular rate, is $2\pi/86400$ radians per second. Therefore the worst case of this third term never exceeds

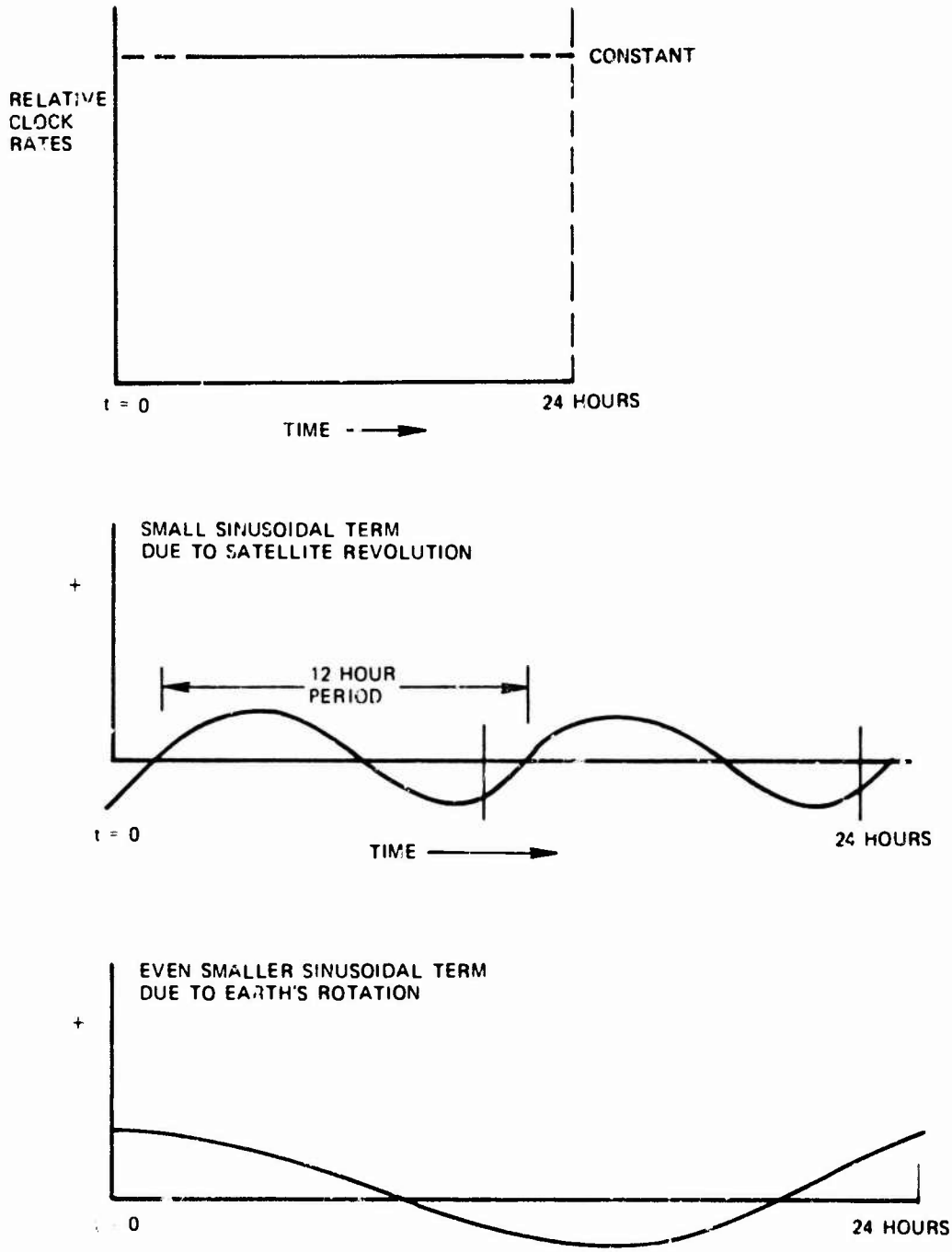
$2 * 448 * 10^{-12} * (1/941) * 86400/2\pi = 12.75$ nanoseconds.

As before, the uncertainty in this quantity, due to finite bit count, can be attributed to C, which must be given to 1 part in 12.75 parts, requiring at least 4 bits ($2^4 = 16$).

In order to avoid compounding of roundoff error both B and C should be represented as at least 6-bit words. In fact, since the two phase angles in Eq (1) require 6 and 7 bits, respectively, all quantities except A in Eq (26) should be represented by 8 bits.

12.3 Initial Phase Angles

The phase angles θot and θoe in Eq (26) have an impact on (1) the maximum deviation of the integrated clock time difference and (2) the time at which that maximum occurs (see figure 2). The integral of the second term in Eq (26) is sufficient to size the bit count required for those two phase angles.



NOTE: INITIAL PHASES OF BOTH 12-HOUR AND 24-HOUR TERMS VARY FROM STATION-TO-STATION AND FROM DAY-TO-DAY AND ARE SATELLITE-PECULIAR PHASES OF THE SINUSOIDAL TERMS, AND AMPLITUDES OF ALL THREE TERMS ARE PREDICTABLE, GIVEN THE ORBITS AND THE GROUND STATION LOCATION

Figure 4 Components of the Clock Rate Variation

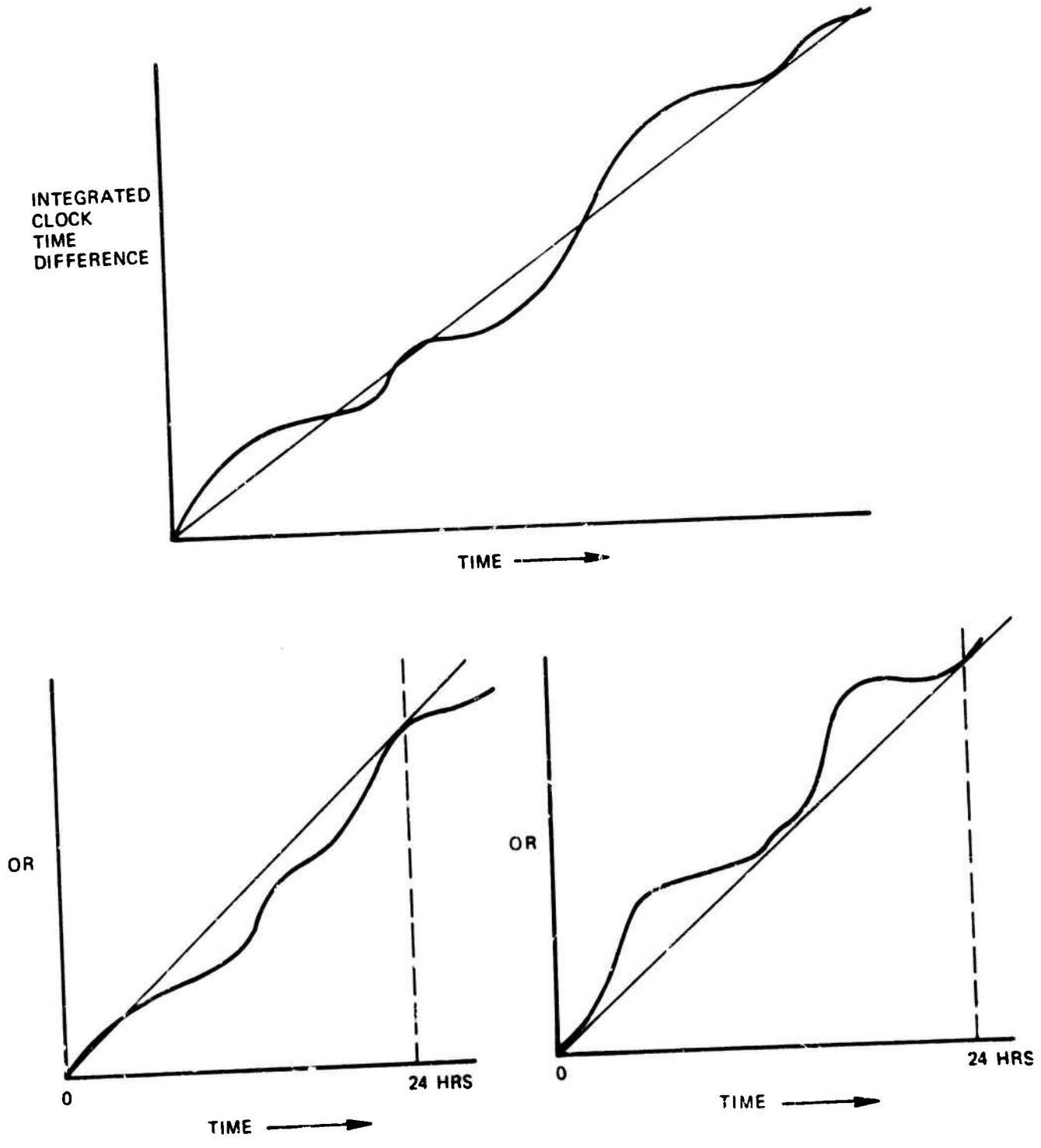


Figure 5 Possible General Forms of Integrated Time Difference

TABLE 8

SUMMARY OF BIT COUNTS
FOR EQ (26) PARAMETERS

QUANTITY IN EQ (1)	COMPUTED MINIMUM BIT COUNT	RECOMMENDED BIT COUNT
A (1-hour, 24-hours)	11 bits, 16 bits	11 bits, 16 bits
B	5 bits	8 bits
C	4 bits	8 bits
θ_{ot}	7 bits	8 bits
θ_{oe}	6 bits	8 bits
	TOTALS (1-hour)	43 bits
	(24-hours)	48 bits

The integral between 0 and T of the middle term of Eq (26) is

$$d = \frac{AB}{\omega t}\left[\sin\left(\theta_{ot} + T_{\omega t}\right) - \sin\theta_{ot}\right]. \qquad (28)$$

From the discussion given above, AB/ωt never exceeds 12.13 nanoseconds.

The time T at which d is a maximum (or minimum) is obtained by differentiating Eq (28) partially with respect to T, and setting the derivative to zero. This yields the result

$$\tan(T_{\omega t}) = \cot\theta_o;$$

upon using this result in Eq (28) it follows that

$$d_{max} = AB/_{\omega t}\left(1 - \sin\theta_{ot}\right).$$

Now, the derivative of d_{max} with respect to θot gives the rate at which the maximum integrated time error, due to the middle term in Eq (26), varies with respect to the phase angle θot; this derivative is simply

$$\partial d_{max}/\partial\theta_{ot} = -AB/_{\omega t}\cos\theta_o \text{ nanoseconds per radian,}$$

and this expression has the maximum magnitude if θ = ±90°. Thus, the error in dmax due to an error ε in θ is given by

$$\delta(d_{max}) \leq \frac{AB}{\omega t}\epsilon \leq 12.13 \text{ nanoseconds.}$$

If (dmax) is not to exceed 1 nanosecond, then ε may not exceed 1/12.13 radians. Since θot may have any value between 0 and 2π, the permissible uncertainty in θot is one part in 2 * 12.13 = 76.7 parts. Therefore, at least 7 bits (2^7 = 128) are required to express θot.

It is discussed above that the third term in Eq (26) does not contribute more than about half as much integrated clock time difference as does the second term; therefore, 7 bits is also adequate for the representation of phase angle θoe. (In fact, literally 6 bits would suffice). These results are summarized in Table 8.

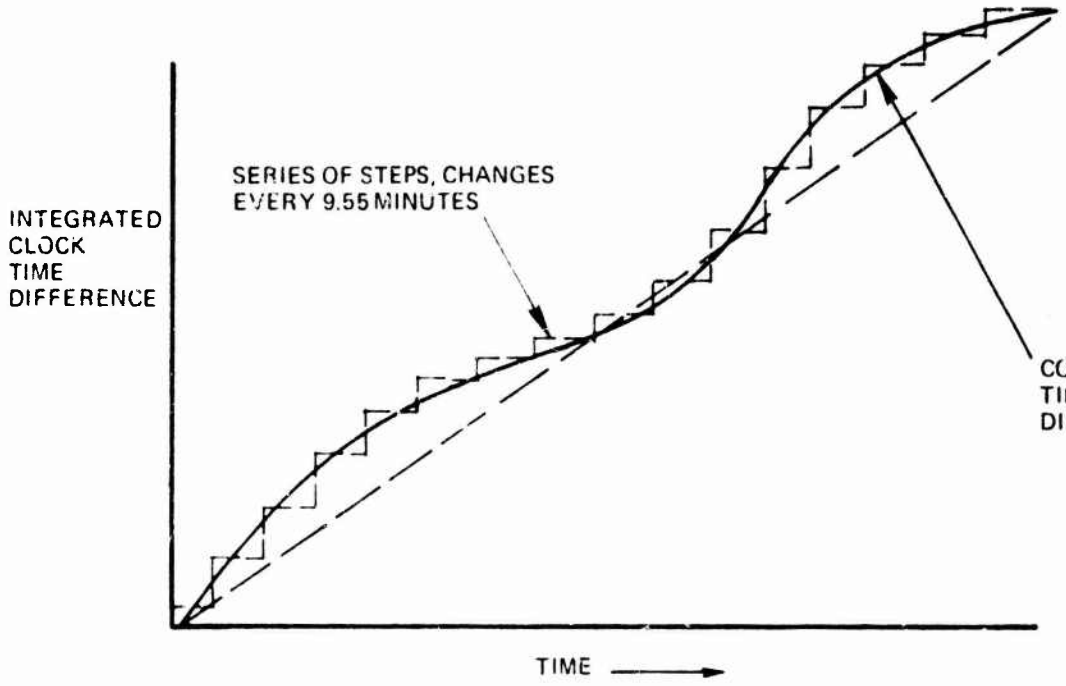


Figure 6a Stair-Step Approximation

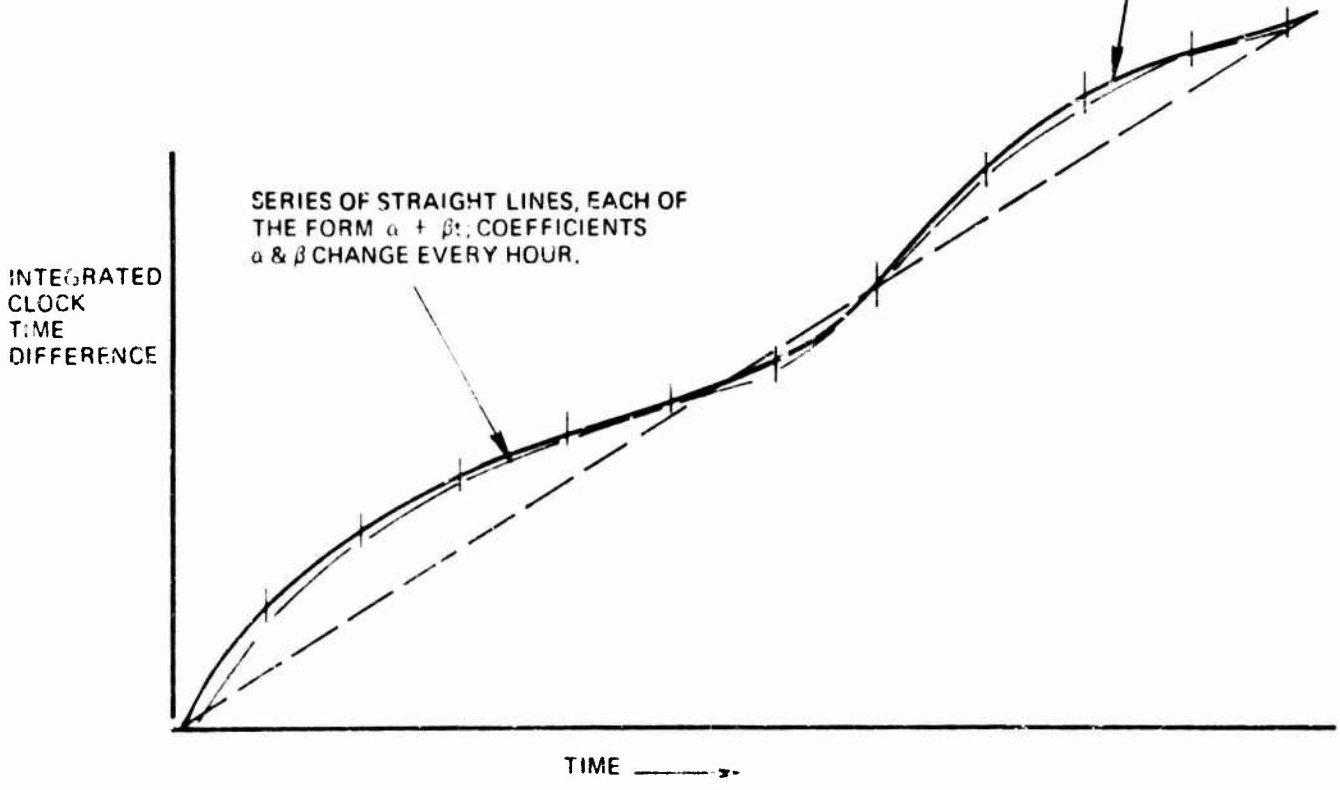


Figure 6b Straight-Line-Segment Approximation

12.4 Stair-Step Approximation

As an alternate to computing the clock time rate difference from Eq (26), or the integrated time difference D from Eq (27), initial time To to final time Tf, one may use either a stair-step approximation, or a series of straight-line approximations, as shown in Figures 6a and 6b, respectively.

If constant A has the value $448 * 10^{-12}$, then because of just this term alone D will increase by 1 nanosecond in about 2.235 seconds. Representation of Eq (27) by a series of stair steps would require a new step at this interval, implying 1614 steps per hour, and 38630 steps per day, if the step-to-step increment is 1 nanosecond. Since the accumulated time difference for a full 24-hour period is about 38630 nanoseconds, each step would have to be represented by a 16-bit number, which is obviously completely out of the question.

A much more practical approach would be to provide the user with a value of A, good for 24 hours, and requiring a 16-bit number; in addition, the sinusoidal terms in Eq (27) would be represented either in the form of stair steps, or as a series of straight-line approximations, each of the form (a + bt), each thus requiring two numbers. The user would compute the linear time term in Eq (27).

If the stair-step approach is used the time interval per step must be small enough that the sinusoidal function changes at most by 1 nanosecond from step to step. The maximum rate of change of D in Eq (27), considering only the second term, occurs when $Tf\omega t + \theta ot = 0$ or π, and amounts to AB; since A is about $0.448 * 10^{-9}$ and B is not more than 0.0039, this rate of change cannot exceed $1.745 * 10^{-12}$ second per second, which accumulates to 1 nanosecond in 573 seconds, or 9.55 minutes, or 0.159 hour. Therefore, one must use 24/.159 = 151 steps for a 24-hour representation. And since the peak value of the middle term in Eq (27) can be as large as about 24 nanoseconds, one must use 5 bits ($2^5 = 32$) to represent each step, if roundoff error is not to exceed 1 nanosecond.

Thus, the stair-step approach requires a total of 16 bits for A and 5 * 151 = 755 bits for the sum of the two terms in Eq (27), for a total (minimum) of 771 bits.

12.5 Straight-Line-Segment Approximation

A series of straight lines could also represent the sinusoidal terms in Eq (27). An analysis of this approach has been made; details are not included here, but the conclusions are that

a. A straight line approximation can represent the sinusoidal terms for the period up to 1 hour, with residual error of less than 1 nanosecond (actually, about 0.8 nanosecond).

b. The two coefficients of each straight line segment require 5 bits each, for a total of 10 bits per segment.

For a 24-hour representation there must be 24 straight line segments, requiring 24 * 10 = 240 bits; to this one must add 16 bits for A, for a total of 246 bits.

These considerations are summarized in Table 9, which follows:

TABLE 9
BIT COUNT FOR SEGMENTED APPROXIMATION TO EQ (27)

METHOD OF APPROXIMATION	COMPUTED BIT COUNT (MINIMUM)	SEGMENT INTERVAL (MAXIMUM)
Stair Step	771 Bits	9.55 Minutes
Straight Line Segment	256 Bits	1 Hour

A final alternate technique is to use straight-line segments, valid for one hour, but using only two coefficients for the full hour. In this technique one coefficient would represent the rate at the beginning of the hour, and would be derived from Eq (26); as argued above, an 11-bit number is required for this coefficient. The second coefficient represents rate slope for the next hour; it was shown earlier that 5 bits is sufficient for this slope coefficient.

Thus, for one hour this technique requires 16 bits, and for 24 hours the bit count is 16 * 24 = 384 bits.

Undoubtedly this is the preferred technique, because it permits the (deterministic) relativistic correction to be easily included in the overall clock model, and the bit count for the components of this model (offset, frequency and phase) can easily absorb the relativity correction.

12.6 Summary

By far the lowest bit count is obtained through direct computation of Eq (27); this approach requires the user to compute the sine function of four angles.

For this approach the bit count is only 48 bits, and the calculation is valid for 24 hours in advance. (NOTE: the coefficients in Eq (27) will change from day to day; it would be considerably more complicated to provide coefficients valid for more than 24 hours. However, if the satellite orbits are computed for say 5 days in advance, the coefficients in Eq (27) could be separately computed for each of those days. But the user should be given only a single set of 24-hour coefficients on any one day.)

The next more complicated approach is to use 24 straight-line segments, plus computation of the first term in Eq (27); this approach avoids the need to compute sine functions, but requires 256 bits (for each 24-hour period).

The most complicated approach, from the point of view of bit count is the stair-step approximation, which requires a minimum of 771 bits, and involves steps which change at the interval of 9.55 minutes (if this approach is adopted the interval could be reduced to say 5 minutes, requiring more bits (in fact, 1456 bits), or it could be raised to 10 minutes, requiring fewer bits (736), but would permit a roundoff error very slightly exceeding 1 nanosecond under extreme worst-case conditions).

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

Report No. 7

SECURITY OF NAVIGATION SIGNAL

CONTENTS

SECTION		PAGE
1.0	Introduction	7-1
2.0	Summary of Approaches	7-1
	2.1 Encryption Technique	7-1
	2.2 Remote Keying Technique	7-2
	2.3 Alternate Rekey Distribution Methods	7-4
3.0	Effect of Non-Stationary Orbit on Rekeying Concept	7-4
4.0	Key Issues	7-6
5.0	Security Plans for Phase I	7-7
	References	7-8

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Protected Code Generator	7-3

TABLES

TABLE		PAGE
1	Alternate Rekey Distribution Method	7-5

SECURITY OF NAVIGATION SIGNAL

1.0 INTRODUCTION

An ultimate goal of the GPS is to (a) provide a navigation capability to the user in the presence of enemy jamming and spoofing and (b) to discourage use of the high accuracy P-signal by an unauthorized user. The approach in meeting this goal is to use some form of signal encryption which would prevent unauthorized users from either receiving and using the signal himself, or from transmitting a replica of the signal which could jam or spoof the user.

In 1972, Signal Definition Studies were performed by Philco-Ford and Magnavox Research Laboratories for the USAF/SAMSO (References 1 and 2). In these studies the security problem was considered in detail. The National Security Agency also participated in these studies. As a result of these studies a system was recommended which (a) employed Crypto keying of the P-signal and (b) provided a means of selectively rekeying the P-signal code in the event of compromise.

In 1973 the GPS was organized and the security requirements were modified. The P-signal would be encrypted in the satellite (by ground command) during the Phase I program. Tests would be made with a limited number of receivers provided with key generators. No provisions would be made for rekeying, however. After these tests were completed and evaluated, security of the operational system (Phase III) would be defined and implemented.

The remainder of this section will discuss some aspects of the security problem.

2.0 SUMMARY OF APPROACHES

2.1 Encryption Technique

In the satellite both the clock and PRN code shall be fed to a transmission security device (TRANSEC), which shall provide secure output to the P modulator.

(See Figure 1.) This TRANSEC device shall utilize the NASA KEESEE hybrid as the basic security source. Approximately 150 bits are to be provided from the command receiver to fill the TRANSEC. Provision shall be made to bypass the TRANSEC on external command (Ref. 3).

The user equipment will generate the P-code in the same manner. If the user has the correct key in the TRANSEC and the PRN sequence generators are timed correctly, then the user code will match the received code from the satellite and signal reception will take place.

2.2 Remote Rekeying Technique

Due to the large user population and the resultant relatively large probability of security compromise, it is desired that the system provides for (a) automatic a priori keying of the system, (b) automatic selective rekeying by user group or category, and (c) a means for a user to enter the system after rekeying takes place. The general technique for rekeying is described as follows:

The user population is rekeyed by (1) transmitting the new key as part of the navigation data as a "special data message." (Ref 4). This key word is stored by the user and inserted into his TRANSEC at a predetermined time when the key is changed (ie, midnight every other day), or alternatively, upon receipt of a command to change key via the navigation signal channel.

The user may initially receive his key manually from an area user terminal or other source. Once in operation, his key is automatically updated via the navigation message. With this approach, the rekeying message need be transmitted only over the P-signal, which is the secure link. However, a danger exists in that the user, for some reason, misses receiving the new key word, he would automatically lose the P-signal at the time of key change. To provide for this contingency and also to provide a means for a new user to enter the system without knowledge of the current P-signal key, it has been proposed that the rekey word be transmitted in the clear over the C/A signal. Since the C/A signal is unprotected, the rekey word itself must be encrypted. Thus each user must carry

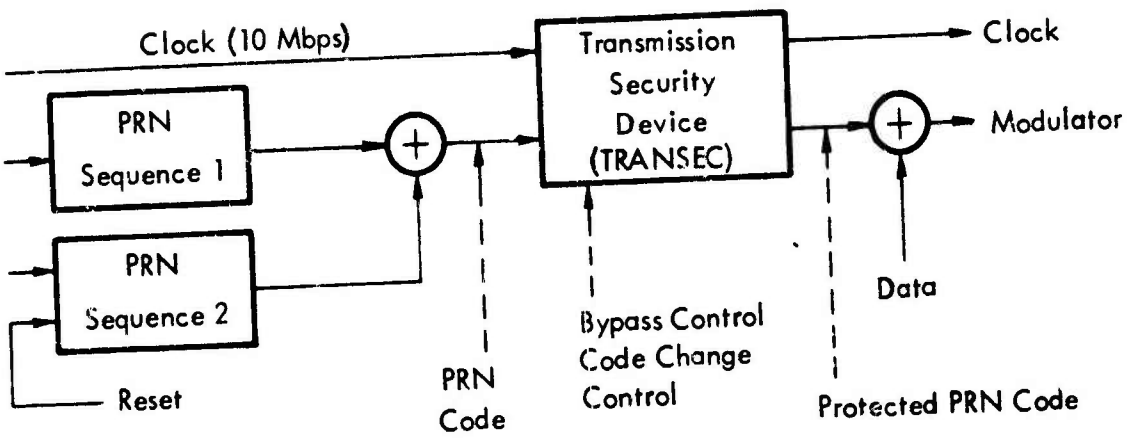


Figure 1 Protected Code Generator

a data encryption key to decode the rekey word. This encryption key vector would be assigned to the users and manually inserted into their equipment.

2.3 Alternate Rekey Distribution Methods

Table 1 lists three methods of distributing the P-signal keys to the user. The method described in the bottom line of the table meets all of the desired requirements of rekeying but has the greatest impact on the total system, particularly if selective rekeying to unique users is employed, thus placing a heavy demand on the C/A data channel.

3.0 EFFECT OF NON-STATIONARY ORBIT ON REKEYING CONCEPT

The studies performed in 1972 were based on a synchronous satellite orbit configuration. The current GPS orbits are non-stationary orbits with a period of 12 hours. The impact of the 12 hour orbit is that the rekeying sequence must be initiated well in advance of the time the satellite key is changed. Normally, each satellite is loaded with the updated navigation message once a day. Thus at least 24 hours must elapse prior to rekey.

The time of key change must be executed in each satellite by a stored program since (a) all satellites must change key simultaneously and (b) some of the satellites will be out of view of the ground control segment at the time of rekey.

If rekeying by groups is employed, it may be necessary to time phase the rekeying words in the navigation message with the geographical location of the satellite suborbital point relative to the location of the groups. (The groups would be arranged by location as well as by operational missions.) If this isn't done, there may not be sufficient time to transmit all of the group keys before the satellite moves beyond the view of the user. Limitations on the size of the

TABLE 1

Alternate Rekey Distribution Methods

METHOD	MEDIUM	RISKS	RELATIVE COST	IMPACT ON GPS
REKEY ALMANAC TO AIRFORCE AND NAVAL INSTALLATIONS	COURIER, DISTRIBUTION POINT TO USER	PHYSICAL INTERCEPTION OR COMPROMISE	HIGH	NONE
	RADIO COMMUNICATIONS, DISTRIBUTION POINT TO DCA TO USER	ELECTROMAGNETIC COVERT INTERCEPTION TRANSMISSION ERRORS POSSIBLE BUT CAN BE CORRECTED	MODERATE	NONE
ADVANCE REKEY ALMANAC RELAY OVER GPS	RADIO REPLAY OVER GPS P-CHANNEL DISTRIBUTION POINT TO SGLS TO USER	ELECTROMAGNETIC COVERT INTERCEPTION TRANSMISSION ERRORS CANNOT BE CORRECTED	LOW	GPS P-DATA CHANNEL BURDEN INCURRED PERMITTING LESS FREQUENT NAVIGATION DATA RELAY
A POSTIORI REKEY RELAY OVER GPS TO PERMIT LATE USER ENTRY	RADIO REPLAY OVER GPS C/A-CHANNEL, DISTRIBUTION POINT TO SGLS TO USER	ELECTROMAGNETIC COVERT INTERCEPTION TRANSMISSION ERRORS CANNOT BE CORRECTED	LOW	GPS C/A-DATA CHANNEL BURDEN INCURRED PERMITTING LESS FREQUENT P-SIGNAL ACQUISITION DATA

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

navigation data storage equipment on board the satellite places a limit on the number of types and frequency of rekey messages which can be transmitted by a satellite between upload periods. By proper allocation of different group or unique crypto vectors to be stored and transmitted by each satellite, the effective rate of transmission of the rekey messages to the total user population is enhanced by the total number of satellites in the space segment (ie, 24 or 27 satellites). Details of rekeying sequences employing such strategies have not been worked out. Obviously they are highly dependent on the deployment of the secure user population. A deployment model would have to be postulated before conducting any analysis.

4.0 KEY ISSUES

The key issues which we believe must be resolved before implementing an operationally secure system are:

a. Can a reliable rekeying sequence be established which is compatible with a reasonable satellite uploading and distribution schedule?

b. Does the group/unique dual crypto vector concept provide adequate security for transmission of the P-signal code key over the C/A signal data link?

c. Does a selective rekeying technique require too much satellite data storage and take too long to rekey all unique users?

d. Is the rekeying concept simply too complex for an operational system? (ie, keeping track of all user equipment codes, locating compromised users, etc.)

e. Is the added cost to the user equipment worth the added protection? (Note that the added logic circuit cost is probably a minor part of the cost increment. More significant is the special handling of Red/Black interfaces, logistics, etc.)

The answer to these questions may lead to the definition of a less complex system which would have some degree of security, but not the features of remote rekeying discussed above.

5.0 SECURITY PLANS FOR PHASE I

As a result of the above considerations, it was decided to limit the security objectives for Phase I. The objective of the Phase I program is to demonstrate the feasibility of transmitting and receiving a crypto-keyed navigation P-signal. No attempt will be made to secure the code in the control segment.

Specific functions which will be demonstrated in Phase I are:

a. Uploading into satellite memory via secure uplink
b. Generation of secure code in satellite
c. Acquisition and reception of secure code by user receiver
d. Mutual interference effects (if any) from other three satellites operating in secure mode

REFERENCES

1. System 621B Signal Definition Study Final Report, Philco-Ford Corporation, SAMSO-TR72-247, October, 1972 (Confidential), Section 3.3 and Appendix E.

2. System 621B Signal Definition Study, Magnavox Research Laboratories, SAMSO TR-72-248, Volume 2 (Secret), Section II. October, 1972.

3. Space Vehicles Requirements Document DNSDP Space Vehicle System, USAF/SAMSO DNSDP-SVR-101, 73 Dec 28 (Section 3.7.7.2.2.1, p. 32).

4. System Specifications for the GPS USAF/SAMSO SS-GPS-101A Appendix II 29 January 1974, Par. 20.5.2.6.1.

REPORT NO. 8

ERROR BUDGET ANALYSIS

ERROR BUDGET ANALYSIS

TABLE OF CONTENTS

SECTION		PAGE
1.0	INTRODUCTION AND SUMMARY	8-1
	1.1 Analysis Approach and Report Organization	8-11
2.0	EPHEMERIS ERROR	8-13
	2.1 Accuracy of Data Input to Ephemeris Processor	8-14
	2.1.1 Monitor Station Data Accuracy	8-14
	2.1.2 Ionospheric Correction Accuracy	8-17
	2.1.3 Range Sample Smoothing	8-19
	2.1.4 Tropospheric Correction Accuracy	8-20
	2.1.5 Net Accuracy of Range Data	8-20
	2.2 Net Ephemeris Error	8-23
	2.3 Effects of Baseline Variations	8-26
	2.3.1 Distributed Processing	8-26
	2.3.2 Satellite Clock Frequency Offset Uncertainty	8-26
	2.3.3 An Alternative Tracking Station	8-26
	2.3.4 BOS Excluded from the Tracking Network	8-28
	2.3.5 Reference Longitude at VTS	8-28
	2.3.6 Station Location Uncertainties	8-28
	2.3.7 Range Rate Noise - .05 feet/second	8-28
	2.3.8 Distributed Processing (1/30/74 Baseline)	8-28
	2.3.9 Simultaneous Multi-Satellite Processing	8-29
	2.3.10 Geopotential Uncertainty Effects	8-29
	2.4 Ephemeris Representation Errors	8-30
	2.5 Relative Accuracy	8-31
3.0	IONOSPHERIC ERRORS	8-32
	3.1 Two Frequency Correction Accuracy	8-34
	3.1.1 User's Ionospheric Correction Algorithm	8-35
	3.1.2 Accuracy of the User's Two Frequency Correction	8-37

TABLE OF CONTENTS (continued)

SECTION		PAGE
	3.2 Fixed Bias Correction Accuracy	8-38
	3.2.1 Monte Carlo Simulation	8-38
	3.2.2 Total Electron Content Measurements	8-41
	3.2.3 Summary of Fixed Bias Analysis	8-41
	3.3 Relative Accuracy	8-42
4.0	TROPOSPHERIC ERRORS	8-43
	4.1 Tropospheric Error Analysis	8-49
	4.2 Humidity Errors	8-52
	4.3 Sea Level Pressure Errors	8-52
	4.4 Sea Level Temperature Errors	8-55
	4.5 Temperature Lapse Rate Errors	8-55
	4.6 Summary of Errors	8-58
5.0	MULTIPATH ERROR	8-63
	5.1 Characteristics of the Multipath Signal	8-63
	5.2 Delay Lock Loop Effects	8-65
	5.2.1 Summary of Results	8-65
	5.2.2 Range Error	8-67
	5.2.3 Range Rate Error	8-76
6.0	RECEIVER NOISE AND QUANTIZATION	8-83
	6.1 Pseudorange and Range Rate Tracking Errors	8-84
	6.1.1 Range Errors	8-84
	6.1.2 Range Rate Errors	8-87
	6.1.3 Performance Summary	8-89
	6.2 Reference Link C/No	8-95
	6.3 User Accuracy	8-100
7.0	GPS CLOCK ERROR	8-104
	7.1 Clock Noise Characteristics	8-106
	7.2 Clock Prediction Algorithms	8-108
	7.2.1 Least Squares/Exponential	8-108
	7.2.2 Linear	8-110
	7.2.3 Carlson Triangular	8-110

LIST OF FIGURES

FIGURE		PAGE
1-1	Absolute UERE vs Time	8-4
1-2	Relative UERE vs Time	8-5
1-3	Relative UERE vs Time	8-6
1-4	Net UERE vs C/No Aircraft/Ships/Manpacks	8-8
1-5	Ephemeris Contribution to User Equivalent Range Errors (with 10 foot Spherical Station Errors)	8-9
1-6	Ephemeris Contribution to User Equivalent Range Errors (without Station Location Errors)	8-10
1-7	UERE Analysis Approach	8-12
2-1	Monitor Station Elevation Angle Probability Distribution	8-15
2-2	Monitor Station Receiver Error Data	8-16
2-3	Monitor Station Multipath Error vs Elevation Angle and Antenna Height	8-18
2-4	Range Error Flow and Values	8-22
2-5	User Equivalent Range Error, Baselines and Variations	8-27
3-1	Ionospheric Correction Error	8-33
3-2	Ionospheric Error Statistics	8-40
4-1	1 Sigma Error due to Deviations from Assumed Standard	8-45
4-2	Phase I Satellite Elevation Angle Distribution	8-46
4-3	User Error Distribution Curves	8-47
4-4	Tropo Range Correction	8-51
5-1	Maximum Surface Roughness for Specular Reflection vs Elevation Angle	8-64
5-2	Range Error During and ILS Approach	8-66
5-3	Generalized Discriminator Curve	8-67
5-4	Statistical Range Error vs Multipath Distance	8-72
5-5	Deterministic Range Error vs Multipath Distance	8-74
5-6	Probability Density of Statistical Range Error	8-75

LIST OF FIGURES (continued)

FIGURE		PAGE
5-7	P Signal Multipath Errors in Delay Lock Loop	8-77
6-1	P Signal Receiver Range Errors	8-91
6-2	C/A Signal Receiver Range Errors	8-92
6-3	P or C/A Signal Receiver Range Rate Errors	8-93
6-4	Elevation Angle Effects	8-98
6-5	C/No vs Elevation Angles (Isotropic Antenna)	8-99
6-6	User/Satellite Elevation Angle Probability Distribution	8-101
7-1	Allan Variance vs Averaging Time	8-107

LIST OF TABLES

TABLE		PAGE
1-1	GPS Error Budget Specification	8-1
1-2	GPS Error Budget	8-2
2-1	Ionospheric Error Correction Algorithm	8-19
2-2	Tropospheric Error Correction Algorithm	8-21
2-3	User 1 Sigma Positioning Errors with Distributed Processing Current Baseline	8-24
2-4	Baseline Covariance Analysis Conditions Distributed Processing Concept	8-25
2-5	Representation Error vs Size	8-31
3-1	Net Error Due to Ionospheric Correction	8-32
3-2	1 Sigma Two Frequency Ionospheric Correction Error	8-37
3-3	Relative Ranging Error of Fixed Correction User	8-42
4-1	1 Sigma Troposphere Range Error	8-44
4-2	Humidity Errors	8-53
4-3	Sea Level Pressure	8-54
4-4	Sea Level Temperature	8-56
4-5	Temperature Lapse Rate	8-57
4-6	1 Sigma Troposphere Errors, 90 and 10 Deg. Angles	8-59
6-1	1 Sigma Ranging Error due to Noise and Quantization	8-83
6-2	1 Sigma Range Rate Error due to Noise and Quantization	8-83
6-3	User Receiver Parameters	8-90
6-4	Range and Range Rate Acceleration Bias Errors	8-90
6-5	Reference C/No for Isotropic Antenna at 5^{o} Elev Angle	8-97
6-6	1 Sigma Receiver Noise and Quantization Errors	8-102
7-1	Mean Clock Prediction Error	8-109

I.0 INTRODUCTION AND SUMMARY

This report addresses the GPS error budget and analyzes the various sources of error in the system. The errors are discussed in terms of User Equivalent Range Error (UERE) which is defined as follows:

> Definition of User Equivalent Range Error. User navigation error is the statistical estimate of the error in computed user position when the user is uniformly distributed in time and position. The navigation error is the uncorrelated portion of the observed user range error multiplied by the Geometric Dilution of Precision (GDOP) for the user's position. This uncorrelated portion of the observed range error is called "User Equivalent Range Error".
>
> The geometric dilution of precision (GDOP) is defined as the square root of the trace of the covariance matrix C_v for the case where the covariance matrix for the errors in the pseudo-range measurements is an identity matrix (implying unit, random, uncorrelated user ranging errors to each satellite). For this special case the covariance matrix is as follows:

$$C_v = G^{-1} G^{-T} = G^{-1} (G^{-1})^T$$

Table 1-1 gives the error budget contained in the GPS system specification. Table 1-2 summarizes the results of the analysis and shows the one sigma UERE two hours after the satellites have been updated. The relative range applies to two similarly equipped users making nearly simultaneous measurements.

TABLE 1-1

GPS ERROR BUDGET SPECIFICATION

	PHASE III	PHASE I
Space Vehicle Ephemeris	5	12
Atmospheric Delay	8 to 17*	8 to 17*
Space Vehicle Group Delay	3	8
Receiver Noise and Resolution	5	5
Multipath	4 to 9*	4 to 9*
Total R.S.S.	12 to 12	18 to 25

*These error quantities represent the use of different atmospheric delay correction methods and different user environmental conditions.

TABLE 1-2

GPS ERROR BUDGET

(a) ABSOLUTE RANGE ERROR (ft)

USER	AIRCRAFT			SHIPS/MANPACK		
ACCELERATION	1 g			0		
TYPE RCVR	P/2	P/F	C/F	P/2	P/F	C/F
RCVR NOISE & QUANT.	1.4	1.4	8.2	1.3	1.3	8
IONOSPHERE	1.2	30	30	1.1	30	30
TROPOSPHERE	0.9	0.9	0.9	0.9	0.9	0.9
MULTIPATH	2.0	2.0	20	2.0	2.0	20
CLOCKS/GROUP DELAY	10	10	10	10	10	10
EPHEMERIS	9.2	9.2	9.2	9.2	9.2	9.2
Total RSS (ft)	13.9	33	39	13.8	33	39

(b) RELATIVE ERROR BETWEEN TWO SIMULTANEOUS USERS (ft)

RCVR NOISE & QUANT.	2.0	2.0	11.6	1.8	1.8	11.3
IONOSPHERE	1.7	0	0	1.6	0	0
TROPOSPHERE	0	0	0	0	0	
MULTIPATH	2.8	2.8	28	2.8	2.8	28
CLOCKS/GP DELAY	0	0	0	0	0	0
EPHEMERIS	0	0	0	0	0	0
Total RSS (ft)	3.9	3.5	30	3.8	3.4	30

P/2 = P-Code, 2 Frequency Ionospheric Correction

P/F = P-Code, Fixed Ionospheric Correction

C = C-Code, Fixed Ionospheric Correction

Figure 1-1 shows the UERE associated with the absolute positioning accuracy of the user as a function of the time after the latest update of the satellites' clock data. This determines how well he can position himself against a given fixed co-ordinate system.

Relative navigation accuracy is also of interest as the user may wish to determine his relative position in three modes: 1) another simultaneous user, 2) his or another user's position at some previous time, or 3) a ground station.

In mode 1, each user estimates his position independently, with common error sources, especially satellite clock errors, causing common position estimation errors. The users then communicate their estimates to each other and compute their relative positions. It can be expected that the common errors will tend to cancel in the subtraction, leaving only the effects of the independent random errors made by the two users.

For the two frequency users, their noise, quantization, ionospheric correction, and multipath errors are RSS'd. Only the noise, quantization and multipath errors are RSS'd for the fixed correction users. The errors associated with this mode are shown in Table 1-1 (b). It can be seen that multipath effects dominate the results if the path differences are less than 1.5 chip widths.

In mode 2 the same noise contributions occur, however the satellite clock, ephemeris troposphere and ionospheric errors grow significantly larger with time. The relative position UERE for this mode is shown as a function of time in Figure 1-2.

In mode 3, a ground station replaces the second user. The station makes measurements like the user, but has the advantage of being stationary and able to take many measurements, thereby reducing the effect of noise. Hence, not only do the common error sources cancel, but the doubling of the noise and multipath power which occurs in mode 1 does not occur in mode 2. Furthermore, if the absolute position of the ground station is known, the computed position can be used for calibration purposes to enable the user to obtain a more accurate absolute position determination. Figure 1-3 shows the user's error relative to a fixed ground station as

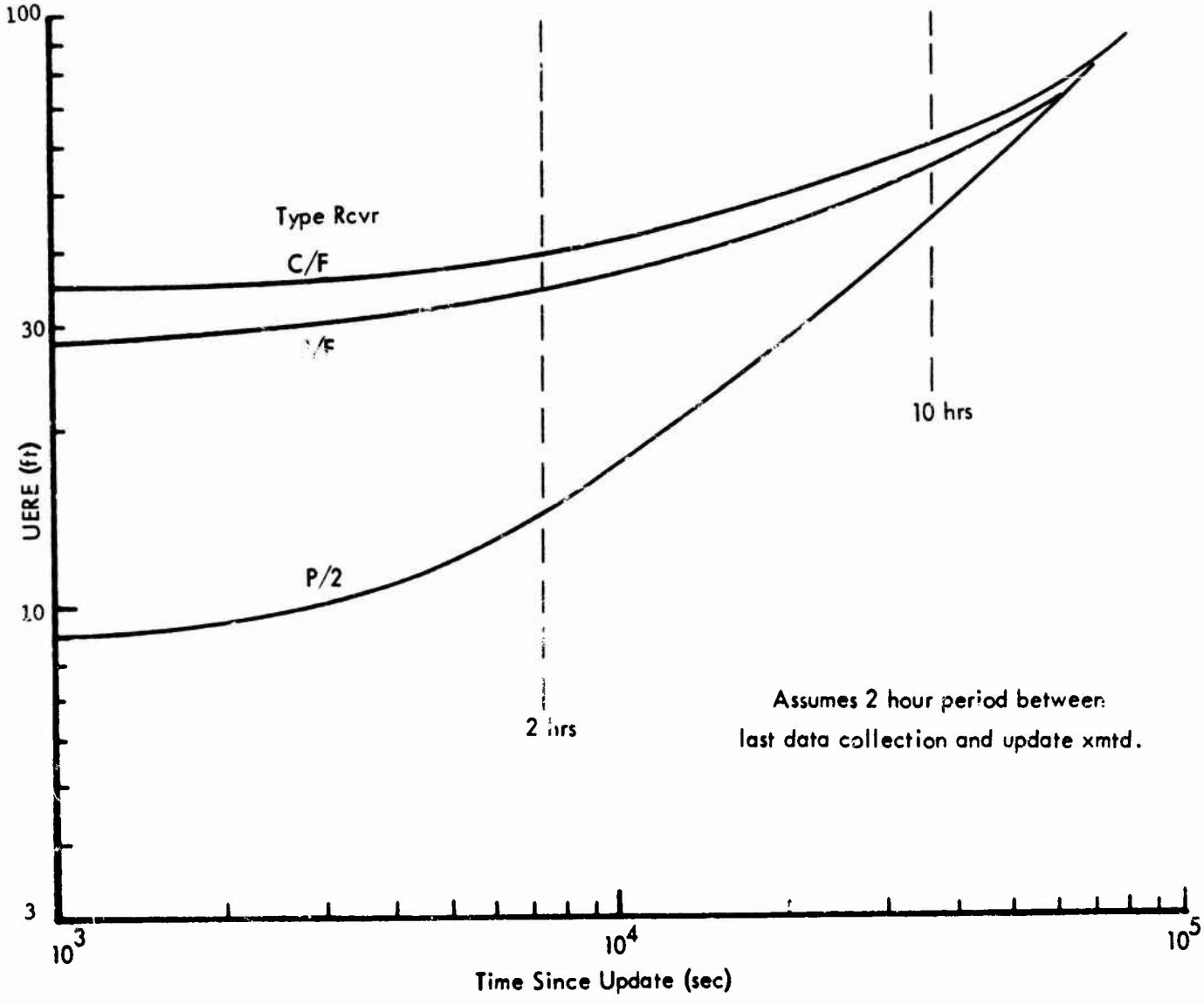


Figure 1-1 Absolute UERE vs Time

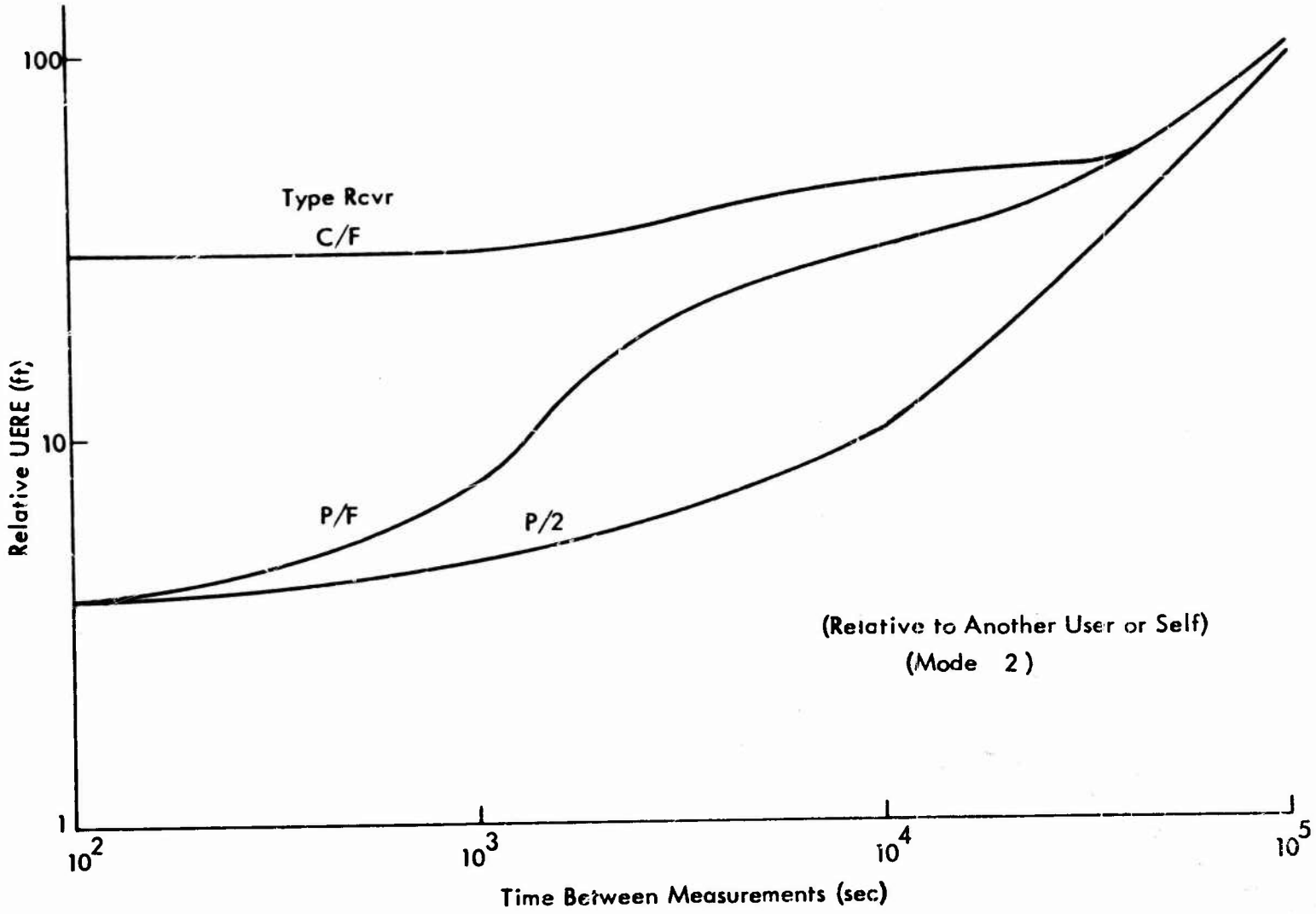


Figure 1-2 Relative UERE vs Time

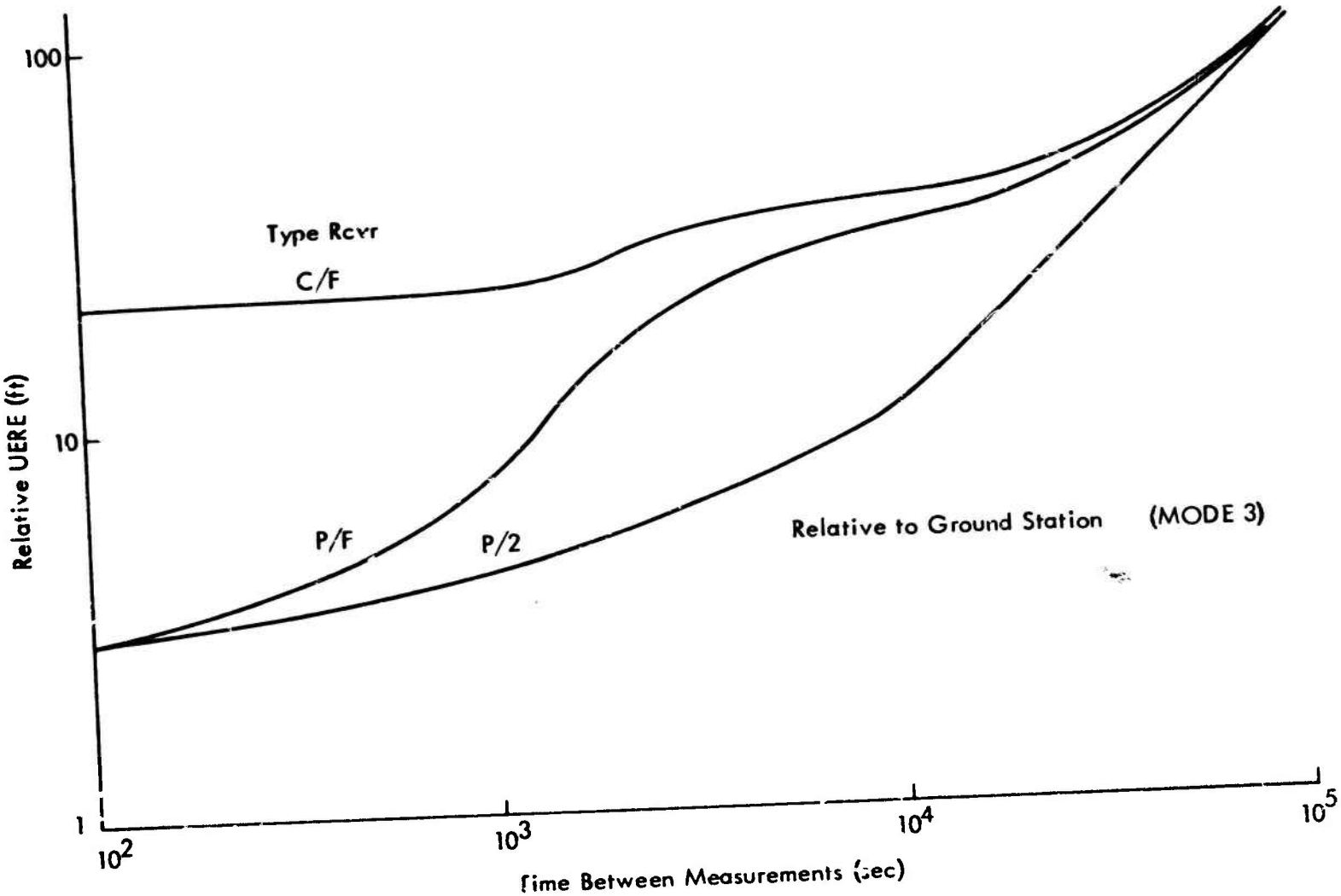


Figure 1-3 Relative UERE vs Time

a function of time between the ground station measurement and the user measurement.

The distance between the users or between the user and the ground station will have some effect on relative accuracy. Tropospheric correction and ephemeris accuracy are geographic separation sensitive error sources. Each can contribute one to two feet to the relative error of the two frequency user. The ionospheric error for the fixed correction user will determine the permissible separation for a given relative error.

In all of the above relative accuracy modes, the users must be operating with the same satellite constellation. If they are not, satellite clock errors will dominate if the time since clock update is greater than 30 minutes for the two frequency users. In this case, the user's absolute accuracy and relative accuracy are similar and Figure 1-1 applies.

The effect of C/N_o variations on UERE is shown on Figure 1-4 for a time 2 hours after the last update. The P code two frequency user's accuracy is essentially determined by the accuracy of the ephemeris and the predictability of the satellite clock. The higher receiver threshold shown is caused by the 3 dB lower signal level of the L_2 signal. The P code, fixed ionospheric correction receiver is limited by the accuracy of this correction down to the threshold of the receiver. The ULRE of the C code receiver is determined by signal level for weak signals and by the accuracy of the ionospheric correction above 36 dB-Hz (approximately -163 dBw out of the antenna).

The Ephemeris Contribution to UERE at four geographic locations is shown in Figure 1-5 for both the Distributed Processing (baseline) and Multi-Satellite Processing techniques. The values shown assume monitor station location uncertainties of 10 feet spherical. This is very conservative and uncertainties considerably smaller than this are more likely after the GPS becomes operational. Figure 1-6 shows the Ephemeris Contribution if this error source were removed.

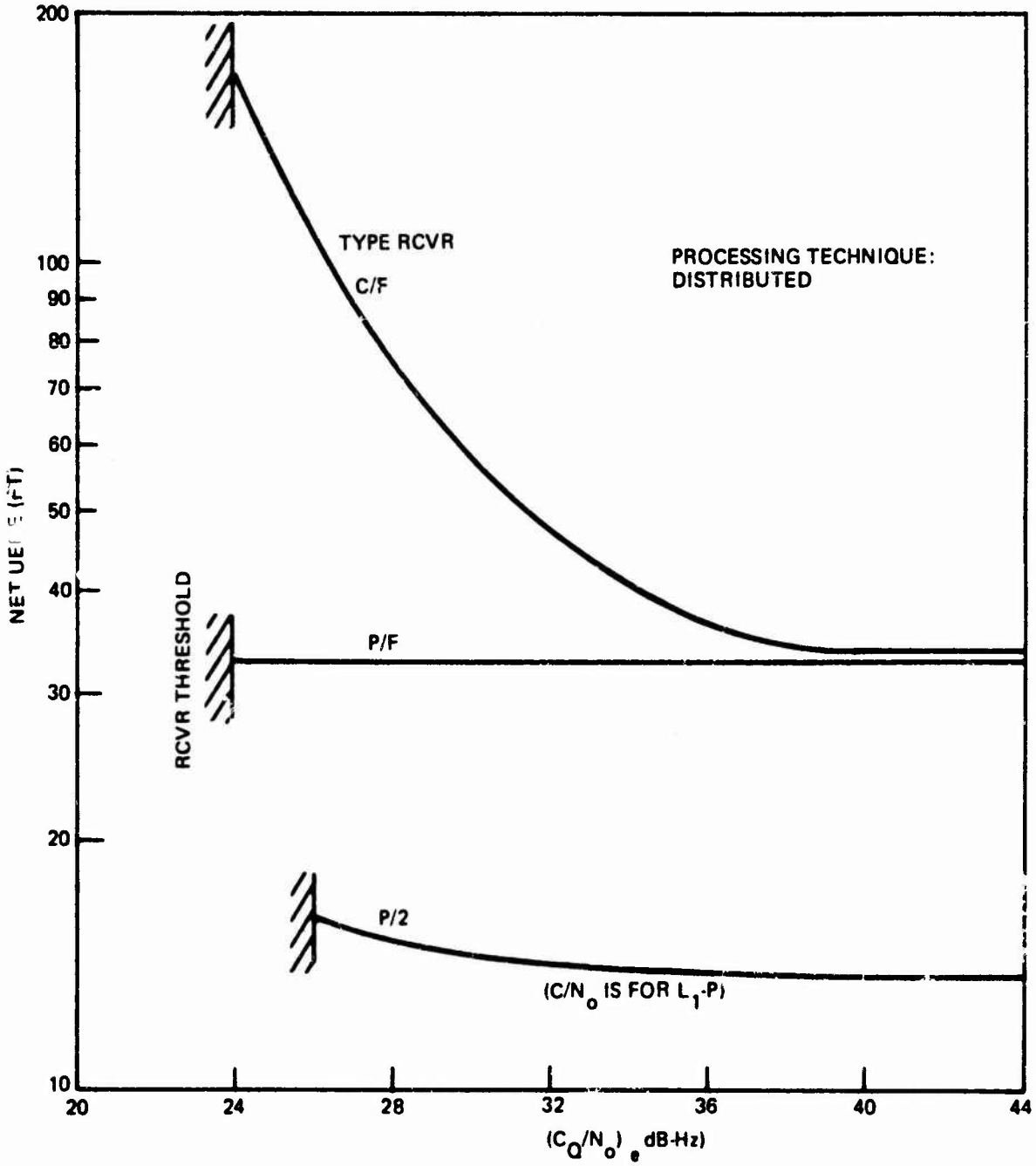


FIGURE 1-4 Net UERE vs C/N$_o$
Aircraft / Ships / Manpacks

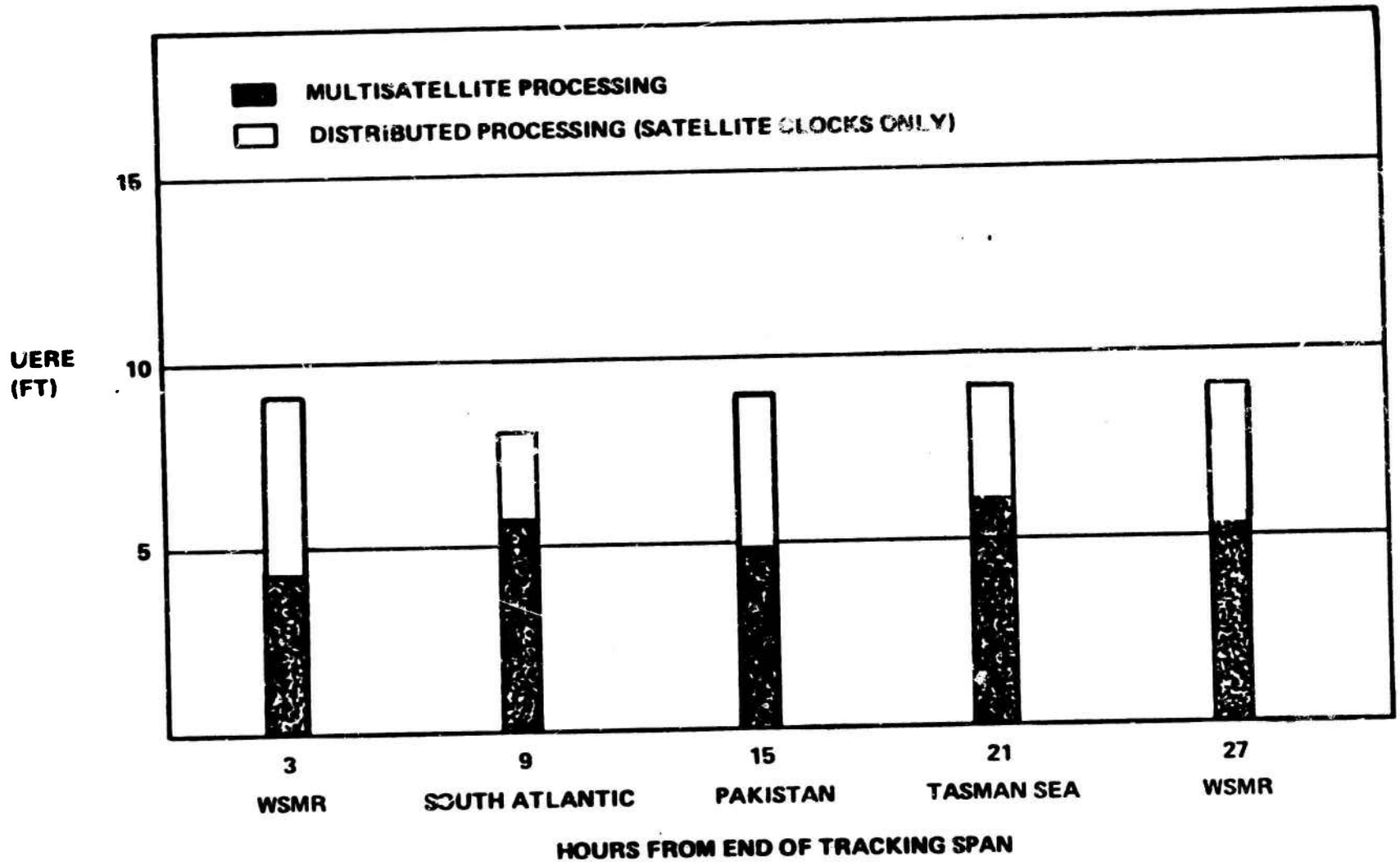


Figure I-5 Ephemeris Contribution To User Equivalent Range Errors (With 10 Foot Spherical Station Errors)

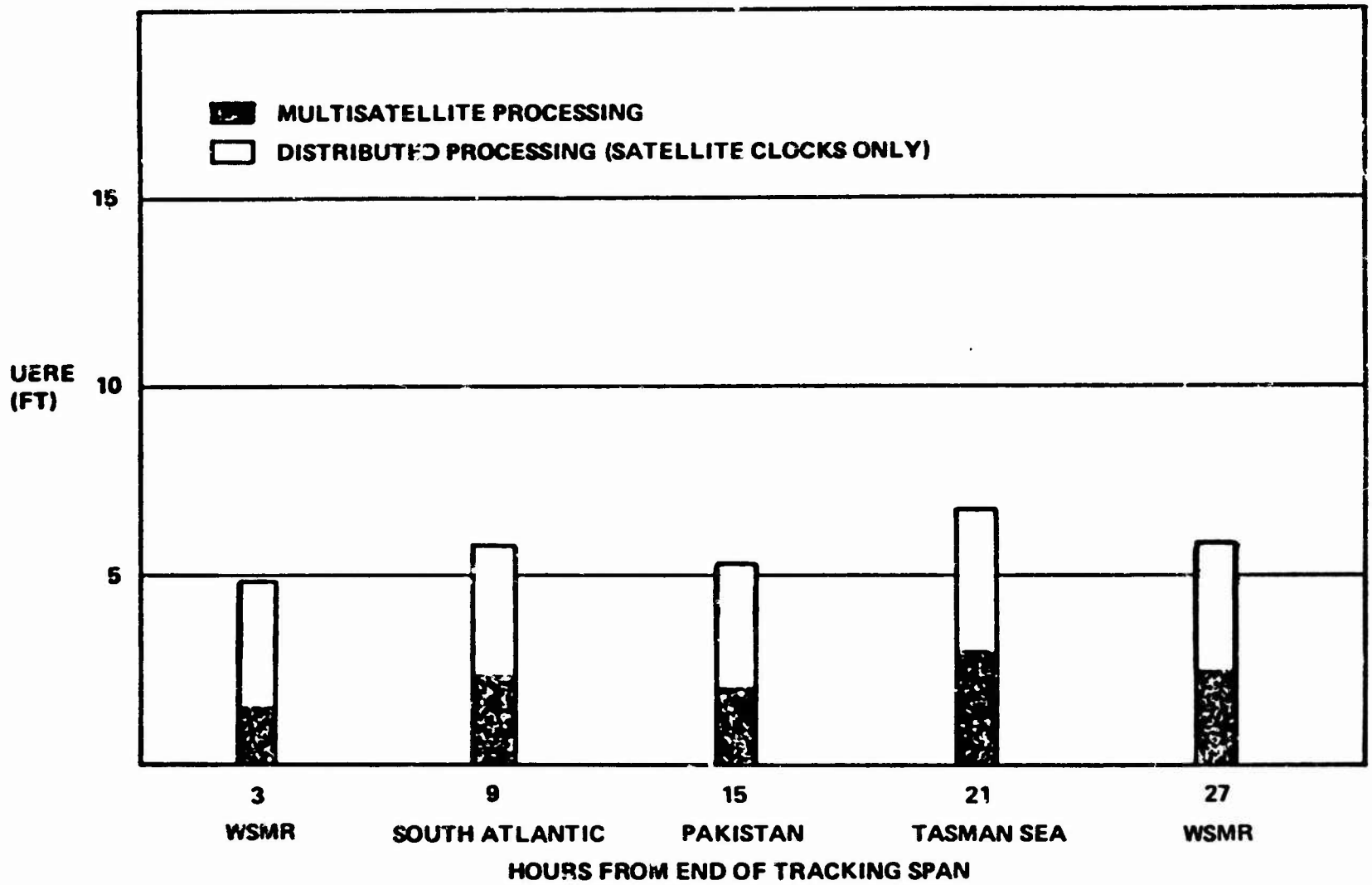


Figure I-6 Ephemeris Contribution to User Equivalent Range Errors (Without Station Location Errors)

1.1 Analysis Approach and Report Organization

TRACE 66 was used to simulate the data sampling and ephemeris generation functions accomplished by the Control Segment. The user position and clock errors calculated by TRACE were then divided by GDOP to produce the Ephemeris Contribution to UERE. The derivation and assumptions concerning this parameter are discussed in Section 2. A detailed discussion of the simulations conducted is contained in Part II, Vol. C, Report C-9, Ephemeris and Clock Simulations.

The ionosphere induced errors in the user's range measurement are addressed in Section 3. The two frequency correction accuracy is determined and an estimate of the one sigma error associated with a fixed bias correction is presented. A detailed discussion of the problems assoicated with ionospheric models is contained in Part II, Volume A, Report 5.

The errors caused by variations in the troposphere are analyzed in Section 4. The individual effects of humidity, sea level pressure, sea level temperature and lapse rate are investigated.

The effect of specular multipath is analyzed in Section 5. The characteristics of the reflected signal are briefly discussed and the ranging error induced in the Delay Lock Loop is determined.

The ranging accuracy of the P and C code receivers are calculated as a function of C/N_o in Section 6. A reference link is established based upon the GPS satellite specification and the RF Analysis contained in Part II, Volume A, Report 2. The user error is then calculated.

The ranging error caused by the unpredictability of the satellite clock is then investigated in Section 7. The characteristics of the clock noise are discussed and various prediction algorithms are described. The accuracy of a typical approach is determined. A detailed discussion of clock noise and prediction algorithms is contained in Part II, Volume A, Report 3.

The errors thus generated are then RSS'd to produce the net UERE. This approach is depicted in Figure 1-7.

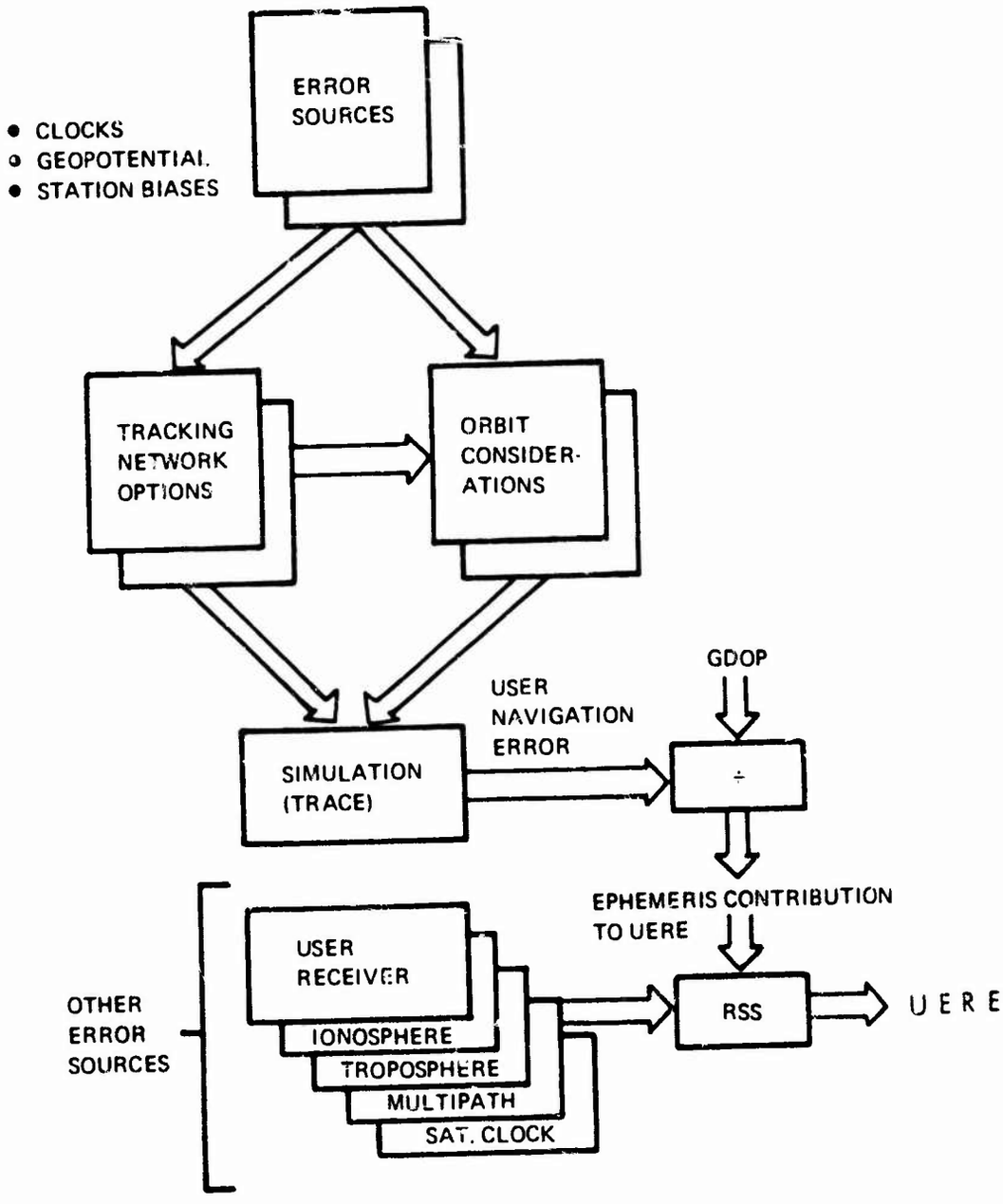


FIGURE 1-7 U E R E Analysis Approach

2.0 EPHEMERIS ERROR

This section analyzes tne UERE contribution of the Control Segment which is passed to the user in the form of slightly erroneous navigation data. Factors which cause this error are: noisy ranging data at the monitor stations, quantization errors, ionospheric, tropospheric and multipath effects at the monitor station, station location uncertainties, clock errors in the space craft and Control Segment stations and geopotential effects. Section 2.1 determines the quality of the data which is input to the MCS data processor to generate the future update message. Section 2.2 summarizes the simulation results of the baseline configuration and presents the net UERE contribution of the Ephemeris. Section 2.3 shows the effect of variations to the baseline and gives UERE contributions under these various configurations or assumptions. Section 2.4 summarizes the analysis of the error caused by the ephemeris representation and Section 2.5 presents the results of simulations which determine the relative error introduced by the ephemeris.

2.1 ACCURACY OF DATA INPUT TO EPHEMERIS PROCESSOR

This section analyzes the quality of the data input into the ephemeris generating data processors. The sampling frequency is one range sample every 15 seconds.

2.1.1 Monitor Station Data Accuracy

The accuracy of the data forwarded from the Monitor Station (MS) to the MCS is dependent upon the signal level at the MS and the antenna and receiver characteristics. Section 6, Receiver Noise and Quantization Errors, develops a reference link C/N_o as a function of elevation angle for an isotropic antenna (Figure 6-5) and determines the range and range rate accuracies as a function of C/N_o (Figures 6-1 and 6-3).

The probability distribution of the satellite elevation angle for four monitor station locations during Phase I is shown on Figure 2-1. It can be assumed that this distribution will change only slightly during the later phases of the program.

The approximate gain versus elevation angle of a broad band, circularly polarized, off-the-shelf, antenna which is the current monitor station baseline approach is shown in Figure 2-2(a). The data has been extrapolated from linear polarization tests so the information shown is quite approximate. Nevertheless, the dip in gain between 10^o and 35^o does represent the typical characteristics of a bifilar wound conical antenna.

The resultant C/N_o of the L_1 and L_2 P-code signals is shown in Figure 2-2(b). This is based upon the reference C/N_o levels shown on Figure 2-2. The range and range rate errors produced are shown on Figure 2-2(c) and (d).

The L_1 and L_2 P-code range and L_1 range rate errors at the average elevation angle of 25^o are 1.7 and 2.6 feet and 0.045 feet/second, respectively.

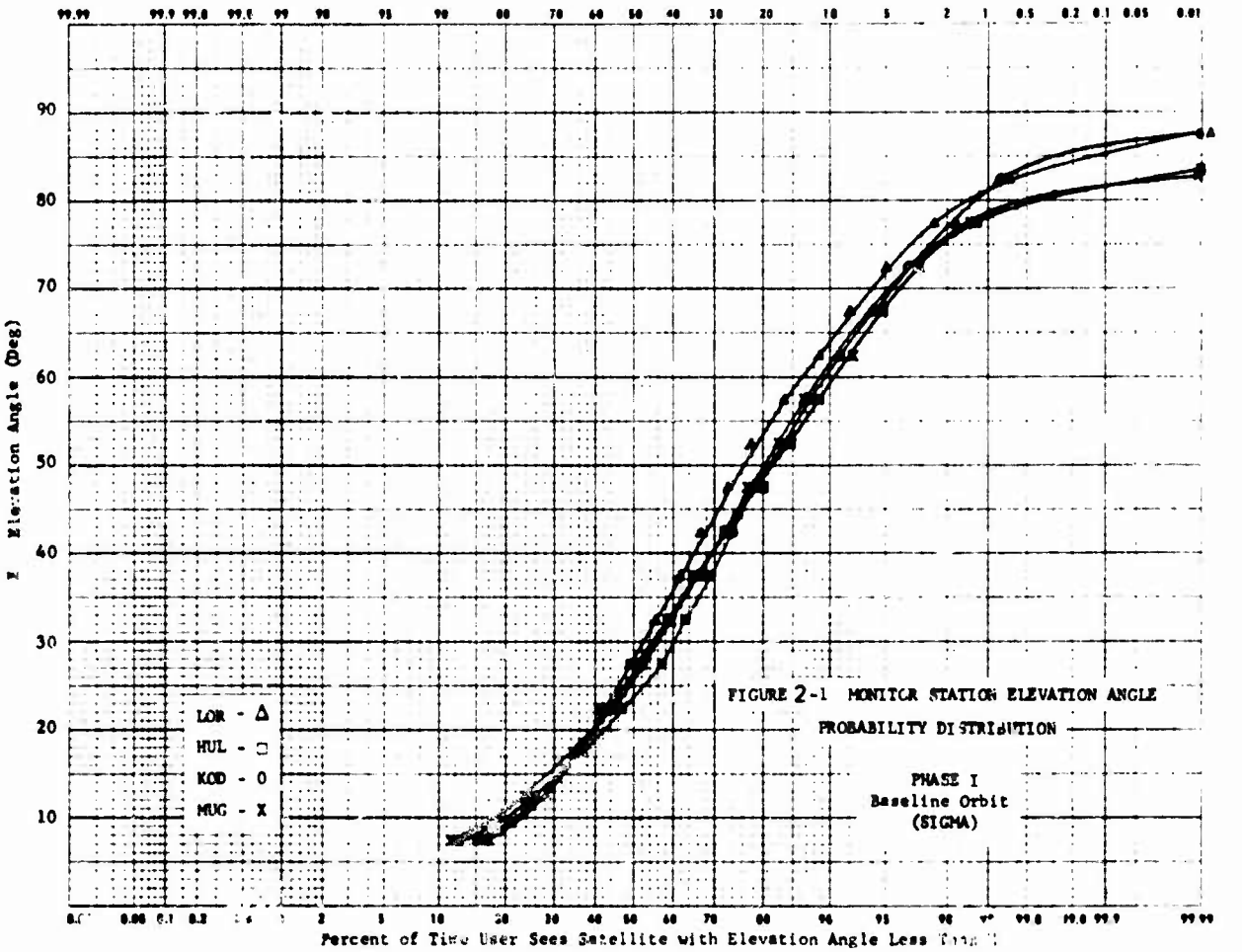


FIGURE 2-1 Monitor Station Elevation Angle Probability Distribution

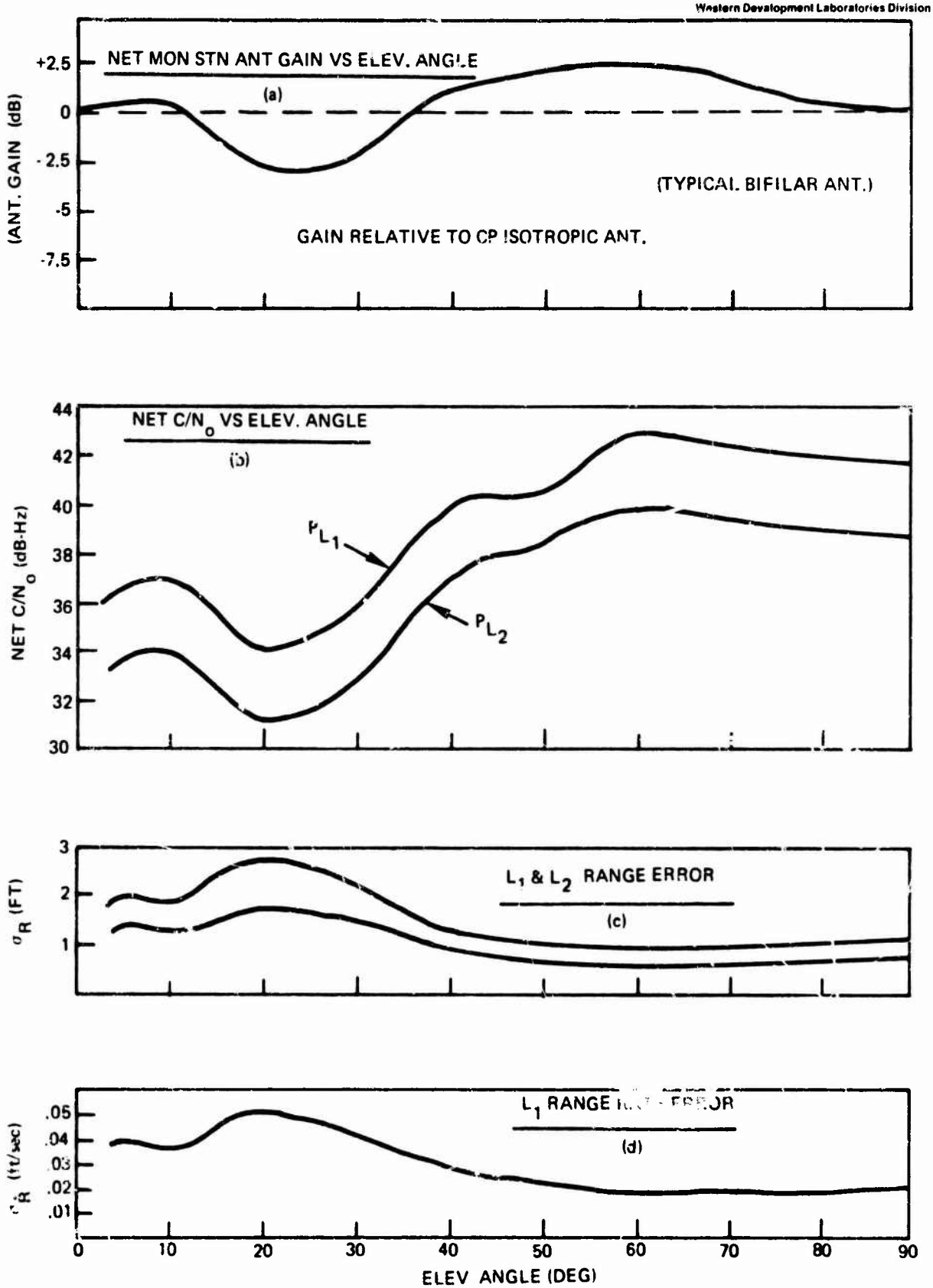


FIGURE 2-2 Monitor Station Receiver Error Data

The multipath analysis in Section 5 determines that the multipath error, $\bar{\ell}$, is given by the following:

$$\bar{\ell} = \frac{-\rho^2 \delta}{1 - \rho^2} \qquad \text{for} \qquad 0 \leq \delta < 50\ (1 - \rho)$$

where $$\rho^2 = \frac{S_{multipath}}{S_{direct}}$$

δ = 2 h Sin (Elevation Angle)

h = antenna height (feet)

It is assumed that the antenna directivity and the reflection characteristics of the area are such that $\rho^2 = 0.1$ (-10 dB). Figure 2-3 shows the multipath error as a function of elevation angle for antenna heights of 5, 10, and 20 feet. For an assumed antenna height of 10 feet and at 25° elevation angle the multipath error is 1 foot.

Adding the multipath and receiver noise errors yields L_1 and L_2 errors of 2.7 and 3.6 feet, respectively. Thus, conservative approximations of the expected uncorrelated error of each L_1 and L_2 15 second sample are 3 and 4 feet respectively. The value specified in the Monitor Station specification is 5 feet.

2.1.2 Ionospheric Correction Accuracy

The pseudorange observations on the L_1 and L_2 signals are forwarded to the MCS which will then perform an ionospheric correction on each 15 second sample. The ionospheric error correction algorithm is stated below in Table 2-1.

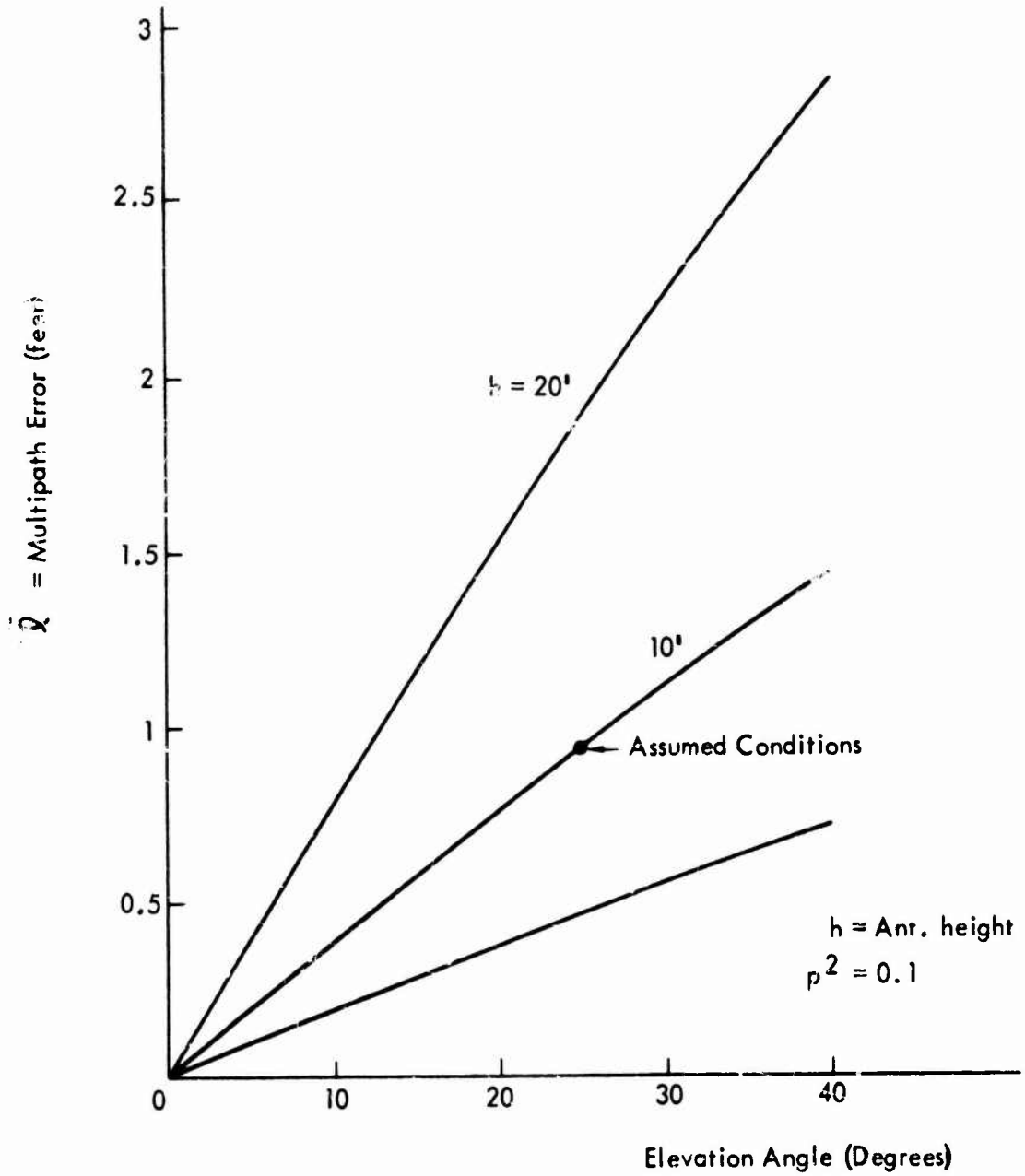


Figure 2-3 Monitor Station Multipath Error vs Elevation Angle & Antenna Height

TABLE 2-1

IONOSPHERIC ERROR CORRECTION ALGORITHM

(Used at MCS)

$$\Delta R_I = (R_{meas\ L_2} - R_{meas\ L_1})\ \rho^2 / (1 - \rho^2)$$

where ΔR_I = Range Correction Factor (ft)

$$\rho = \frac{f_2}{f_1} < 1$$

The accuracy of this correction is given by the following equation:

$$\sigma_{R+I} = \left[\left(\sigma_{L1}^2 + \rho^4 \sigma_{L2} \right)^{\frac{1}{2}} / (1 - \rho^2) \right]$$

where σ_{R+I} is the net error of the range measurement and includes both receiver noise and quantization errors and the ionospheric correction error. With σ_{L1} = 3 feet, σ_{L2} = 4 feet and ρ = 1240/1575, σ_{R+I} equals 10.3 feet.

2.1.3 Range Sample Smoothing

The MCS will then smooth the data for 15 minutes and generate one smoothed L_1 range every 15 minutes. The uncorrelated errors (noise, quantization, ionospheric correction, and multipath) will be reduced by a factor of approximately $\sqrt{N}$ where N is the number of samples and is 60 in this case. It is assumed that the error is reduced by only 5.2 instead of 7.8. This will reduce the uncorrelated L_1 error to 2.0 feet.

2.1.4 Tropospheric Correction Accuracy

The atmospheric data collected at the Monitor Station is also forwarded to the MCS along with the L_1 and L_2 data. The algorithm shown in Table 2-2 is used to generate a tropospheric correction factor which is applied to the smoothed L_1 range.

The error is conservatively estimated at 1.0 feet. This is the one-sigma error using the less exact user correction algorithm:

$$\Delta R_T = 7.8 \, \mathrm{Csc}E \, e^{\frac{-h}{23000}}$$

2.1.5 Net Accuracy of Range Data

The net error of each 15 minute sample is assumed to be the RSS of the smoothed L_1 range error and the tropospheric error. Thus, σ_{NET} is 2.2 feet. The corresponding accuracy required in the MCS specification is 5 feet. The system simulations have shown that the ephemeris accuracy required by the system specification, 12 feet, can be met with this 5 foot error.

To prevent timing errors from having a significant effect upon the accuracy of the system, the range rate accuracy should be on the order of 0.005 ft/sec. This can be met by using a Cesium standard which has a fractional frequency deviation of 10^{-12} over a 900 second (15 minute) averaging period.

Figure 2-4 summarizes the range error flow from the downlink signal to the ephemeris generating processor at the MCS.

TABLE 2-2

TROPOSPHERIC ERROR CORRECTION ALGORITHM

(MONITOR STATION RANGE MEASUREMENT)

$$e_s = 1.64 * 10^{0.0375\ (T-273)}$$

$$N = 77.6\ P/T + (3.73 * 10^5\ e_s * H/100 * T^2)$$

$$N_s = N \exp\ (0.043\ h)$$

$$\Delta R_s = \left|4.79 + 0.00972\ N_s\right|\ \mathrm{Csc}E - \left|0.00586\ (N_s - 360)^2 + 294\right| E^{-2.3}$$

$$K = \exp\left\{-\left|(6.07 * 10^{-5}\ N_s + 0.213)\ h + (0.077/N_s - 1.58 * 10^{-4})\ h^2\right|\right\}$$

$$\Delta R = K \Delta R_s$$

where:

ΔR = Range Correction Factor (ft)

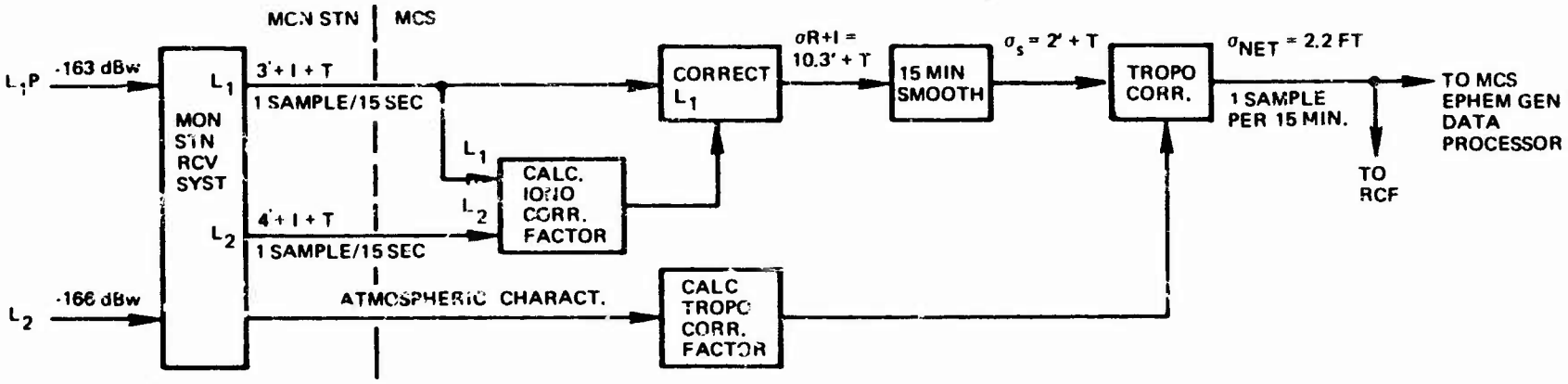
h = Receiver Altitude (ft)

E = Elevation angle to satellite (degrees)

P = Pressure (millibars)

T = Temperature (degrees Kelvin)

H = Relative Humidity (percent)



L_1 C/N_o = 36.2 dBHz

σ_R = 1.45'

L_2 C/N_o = 33.2 dBHz

σ_R = 2.45'

MULTIPATH ≈ 1.1'

$\therefore \sigma_{L_1} \approx$ 3' UNCORRELATED

$\sigma_{L_2} \approx$ 4' UNCORRELATED

ACCURACY OF L_1 AFTER IONO CORRECTION

$$\sigma_{R+I} = \left[(\sigma^2_{L_1} + P^4 \sigma^2_{L_2}) / (1-P^2) \right]^{1/2}$$

$$P^2 = \frac{f_2^2}{f_2^2} = \frac{1.24^2}{1.575^2} = 0.62$$

σ_{R+I} = 10.3 ft

SMOOTHING:

15 MIN/1 SAMPLE PER 15 SECONDS

NO. SAMPLES = 60

$$\sigma_s = \frac{\sigma_{R+I}}{\sqrt{60}} = 1.32 \text{ ft}$$

$\sigma_s \approx$ 2'

TROPO ERROR

σ_T = 0.91 ft

≈ 1'

$$\sigma_{NET} = (\sigma_s^2 + \sigma_T^2)^{1/2}$$

= 2.2 ft

Figure 2-4 Range Error Flow and Values

PHILCO
Philco-Ford Corporation
Western Development Laboratories Division

2.2 Net Ephemeris Error

Simulations were conducted which solved for user position, clock offset, and UERE under conditions of the current baseline for users distributed over the globe and in time from end of tracking. The one sigma UERE's are less than 9.2 feet; well within the twelve-foot error budgeted for ephemeris and satellite clock sources. This data is summarized in Table 2-3.

The covariance analysis which produced these results is based upon the following assumed conditions, which are summarized in Table 2-4.

a. Orbits. Sigma orbits as described in Report C9, Part II, Vol C

b. Perturbing Forces.

1. Zonal harmonics; J_2 through J_6
2. Tesseral harmonics; $J_{2,1}$ through $J_{5,5}$
3. Lunar and solar gravity
4. Solar radiation pressure

c. Tracking Stations and Observations. Simulated observation types are range and range rate with BOS, HUL, and VTS (see Report C9, Part II, Vol. C for locations) being the range rate stations, and KOD being the ranging station; these observations are assumed to have noise standard deviations of 5 feet and 0.005 feet/second, respectively; station location errors were assumed to be ten-feet spherical except that the KOD meridian is taken to be the reference meridian so that there is no uncertainty in the longitude of the ranging station.

d. User Simulation. The primary user is located at the White Sands Missile Range test site with error-free range observations taken at three hours after the tracking stations' observation span; other users are located at alternately northern and southern hemisphere locations, six hours apart around the globe.

TABLE 2-3
USER ONE-SIGMA POSITIONING ERRORS (in ft)
WITH DISTRIBUTED PROCESSING (Satellite Clocks Only)
CURRENT BASELINE

	Δt = 3		Δt = 9		Δt = 15		Δt = 21		Δt = 27	
	P	P + Q	P	P + Q	P	P + Q	P	P + Q	P	P + Q
LATITUDE	5.0	7.5	6.8	15.4	5.7	8.1	7.9	16.0	6.6	8.8
LONGITUDE	4.2	8.5	6.3	12.7	4.8	8.9	7.2	14.0	5.5	9.4
ALTITUDE	16.0	31.9	18.9	24.5	17.7	31.1	22.7	27.1	19.6	30.7
R BIAS	10.1	19.3	12.0	14.1	11.3	18.9	14.4	16.4	12.6	18.8
RSS	20.0	39.0	24.3	34.6	22.2	38.3	29.0	38.2	24.8	38.2
GDOP	4.25		4.25		4.25		4.25		4.25	
UERE	4.7	9.2	5.7	8.1	5.2	9.0	6.8	9.0	5.8	9.0

Δt = HOURS FROM END OF TRACKING SPAN

TABLE 2-4

BASELINE COVARIANCE ANALYSIS CONDITIONS
DISTRIBUTED PROCESSING CONCEPT

ORBIT	SIGMA
PERTURBATIONS	J_2 - J_6
	$J_{2,1}$ - $J_{5,5}$
	lunar and solar gravity
	solar radiation pressure
TRACKING	48 hours @ 15 minute intervals
	obs. type: range at KOD
	range rate at BOS, HUL, VTS
	noise sigma: range = 5 ft
	range rate = 0.005 ft/sec
PARAMETERS	Solution (P): 7 orbit states
	2 satellite clock states
	Consider (Q): 10-foot spherical sensor location error
USER CONDITIONS	ephemeris prediction times: 3, 9, 15, 21, 27 hours after obs. span
	locations: WSMR (Holloman), South Atlantic, Pakistan, Tasman Sea, WSMR, respectively
	observations: error free pseudorange

2.3 Effects of Baseline Variations

Covariance analyses of conditions on the distributed processing concept have been conducted under two baselines. The conditions described in 2.2 above are those of the current baseline in which KOD is the reference timing source (ranging station).

The earlier (January 30, 1974) baseline conditions are different only in that VTS was the ranging station and its longitude was the reference.

The values of UERE resulting from variations in these baselines are tabulated and described in Figure 2-5. Each following subsection refers to entries in Figure 2-5, column-by-column, left to right.
The effects investigated using the January 30 baseline were not repeated for the current baseline, but the results are expected to be qualitatively the same.

2.3.1 Distributed Processing

This column contains results of user positioning simulation under the current baseline and represents the reference for comparison of the variation results (except in the case of 2.3.10, as noted).

2.3.2 Satellite Clock Frequency Offset Uncertainty

To determine the effect of a varying frequency offset due to frequency instability, the TRACE parameter RDBI was entered as a "consider" parameter and was given the value (standard deviation) 0.001 ft/sec. The result was a slight and uniform increase in UERE for all users.

2.3.3 An Alternative Tracking Station

The high-latitude BOS tracking station was replaced in the tracking network by an alternative range rate station located at the GUAM site near the equator. It

CONDITIONS

USERS	TIME FROM END OF TRACKING (NRS)		CURRENT BASELINE (KOD RANGING)							1/30/74 BASELINE VTS RANGING)		
			① BASELINE DISTRIBUTED PROCESSING	② FREQUENCY OFFSET σ = 0.001'/SEC	③ GUAM REPLACES BOS	④ BOS EXCLUDED	⑤ REFERENCE LONGITUDE AT VTS	⑥ STATION LOCATION σ's = 0	⑦ RANGE RATE NOISE = 0.05'/SEC	⑧ DISTRIBUTED PROCESSING	⑨ SIMULTANEOUS MULTI-SATELLITE PROCESSING	⑩ GEOPOTENTIAL σ's INCLUDED
	3	WSMR	6.2	6.3	6.2	11.1	10.0	4.7	13.6	4.5	4.3	5.2
	9	SOUTH ATLANTIC	6.1	8.3	7.8	6.8	9.9	5.7	12.3	12.0	5.8	7.4
	15	PAKISTAN	6.5	6.1	9.1	11.9	9.8	5.2	13.9	5.2	4.7	9.2
	21	TASMAN SEA	8.0	6.1	8.4	9.4	9.7	6.7	13.1	13.1	6.1	9.1
	27	WSMR	6.9	6.1	9.1	11.1	9.9	5.6	14.0	6.0	5.3	7.5

Figure 2-5 User Equivalent Range Error (in feet)
For Baselines and Variations

was expected that this change would produce a smaller UERE at WSMR. Although sourthern hemisphere user errors did improve 6 to 8%, the error at WSMR did not change.

2.3.4 BOS Excluded from the Tracking Network

To determine the effect of a reduced tracking network, this simulation included observations from the HUL, KOD, and VTS stations only. As expected, the values of UERE increase significantly and become marginally acceptable for the northern hemisphere users.

2.3.5 Reference Longitude at VTS

Since the location of the VTS site is well known due to its proximity to the North American datum, it would be desirable to consider its longitude to be the reference longitude. The effect of this variation was determined by setting the KOD longitude uncertainty equal to ten feet and that of VTS to zero. Results indicate that the improvement due to station location accuracy is probably lost in the increased UERE.

2.3.6 Station Location Uncertainties

In the baseline simulation, station location uncertainties are the only considered (Q) parameters. To determine the effects of station location uncertainties, the solution parameter uncertainties were computed with Q-parameter effects omitted. The result shows station location to be a major contributor to UERE.

2.3.7 Range Rate Noise = .05 feet/second

To investigate the effect of crystal rather than cesium clocks at the monitor stations, the range rate noise was increased by an order of magnitude to .05 feet/second. Errors greater than the budgeted twelve feet are the result.

2.3.8 Distributed Processing (1/30/74 Baseline)

This earlier baseline assumed the ranging observations to be taken at VTS and its longitude to be the reference. Comparing these results with those of the

current baseline, one can see that 30% improvement in UERE of the southern hemisphere users is compensation for the 100% increase in the error of the northern hemisphere users, since the errors of the current baseline are more uniform and are all within the budgeted error.

2.3.9 Simultaneous Multi-Satellite Processing

These results were obtained by simulating the simultaneous solution for all ephemeris and vehicle-clock states under the baseline conditions with all stations observing range only. This concept produces obvious benefits in user accuracy but has the processing disadvantages discussed in Part II,Vol. C, Report C8. These results are used as a basis for judging the effects of geopotential uncertainties discussed below.

2.3.10 Geopotential Uncertainty Effects

The introduction of geopotential uncertainties into the simultaneous multi-satellite processing simulation produced the results displayed in this column. These uncertainties were taken to be 3% in the $J_{2,2}$ terms and 5% in the $J_{3,2}$ term and are based on experience with operational satellite systems. They are considered conservative for GPS. The results should be compared with those of 2.3.9 above.

2.4 Ephemeris Representation Errors

The representation to the user of the predicted satellite ephemeris is in the form of a function over time. The types of functions considered for this representation were orbital elements with time-varying corrections and sets of polynomials in time. The polynomials could utilize either inertial or earth fixed coordinates, and could be generated from the data by curve fitting, by differencing for interpolation, or by direct fit.

The performance criteria by which the alternatives were evaluated include: accuracy, data length, user computational requirements, and satellite requirements.

The 1σ fit error for an eighty minute fit interval is 4.0 feet for a fifth degree polynomial, and 0.4 feet for a sixth degree. Truncation error introduced by minimizing the data frame length introduces for three sixth degree polynomials an RSS error of 1.6 feet. If the ephemeris representation data frame size is increased by 42 bits the RSS truncation error may be reduced to 0.5 feet.

The fifth degree polynomial introduces an inadmissible amount of error; a sixth degree is required, and its contribution to the squared error is shown in Table 2.4 for different representation precisions. In the worst case, the impact on user position uncertainty is equal to the satellite position representation uncertainty.

TABLE 2-5

REPRESENTATION ERROR VS SIZE

Bits	Total Representation Error Squared
465	1.7
486	1.0
507	0.6

An additional 27 bits is required for a time word to show the zero point for the polynomials.

Thus, the recommended form of representation to the user is a set of three sixth order polynomials to be updated each hour, but valid for twenty minutes into the next hour. The satellite storage requirement for five days of such data is approximately 60,000 bits.

2.5 Relative Accuracy

For the purpose of this analysis, relative accuracy may be defined as the accuracy with which a stationary user is able to repeat previous estimates of his position and clock offset. The approach detailed in Report C9, Part II, Volume C was to consider the user to be at WSMR with observations twenty-four hours apart such that geometric dilution of precision is constant. The resulting errors in latitude, longitude, altitude, and range bias are 2.2, 1.9, 7.6, and 4.9 feet, respectively, with an RSS error 9.5 feet. This produces a relative UERE contribution by the Ephemeris of 2.24 feet after 24 hours.

3.0 IONOSPHERIC ERRORS

Three techniques are being considered for the correction of the ranging error caused by the ionosphere. These are (1) two frequency ranging system, (2) ionospheric modeling with coefficients transmitted via the navigation data link and (3) fixed bias correction. The first approach essentially subtracts an amount equal to $C(R_{L2} - R_{L1})\ (f_2^2 / f_1^2 - f_2^2)$ from R_{L1}. The net error is dependent solely upon the accuracy of the ranging measurements and is independent of the condition of the ionosphere. The following equation yields the error in the ionospheric correction.

$$\sigma_I = \frac{\omega}{\sqrt{n}} \left[\sigma_1^2 + \sigma_2^2 \right]^{\frac{1}{2}}$$

where σ_1 = rms measurement error pertaining to the higher frequency (L_1)

σ_2 = rms measurement error pertaining to the lower frequency (L_2)

$\omega = f_2^2 / (f_1^2 - f_2^2)$

n = No. of samples = 10

Figure 3-1 shows the relationship between the above net error and the C/N_o at the input to the receiver. Table 3-1 summarizes the one sigma ionospheric errors for the two frequency users assuming the received signal level, losses, system temperature, antenna patterns and elevation angle probability distributions given in Section 6, Receiver Noise and Quantization Errors.

TABLE 3-1

NET ERROR DUE TO IONOSPHERIC CORRECTION

(Two Frequency Receivers)

User Acceleration (g's)	0	1	2	4	7
Aircraft	1.1	1.2	1.3	1.4	1.6
Manpack and Ships	1.1	1.2	N/A	N/A	N/A

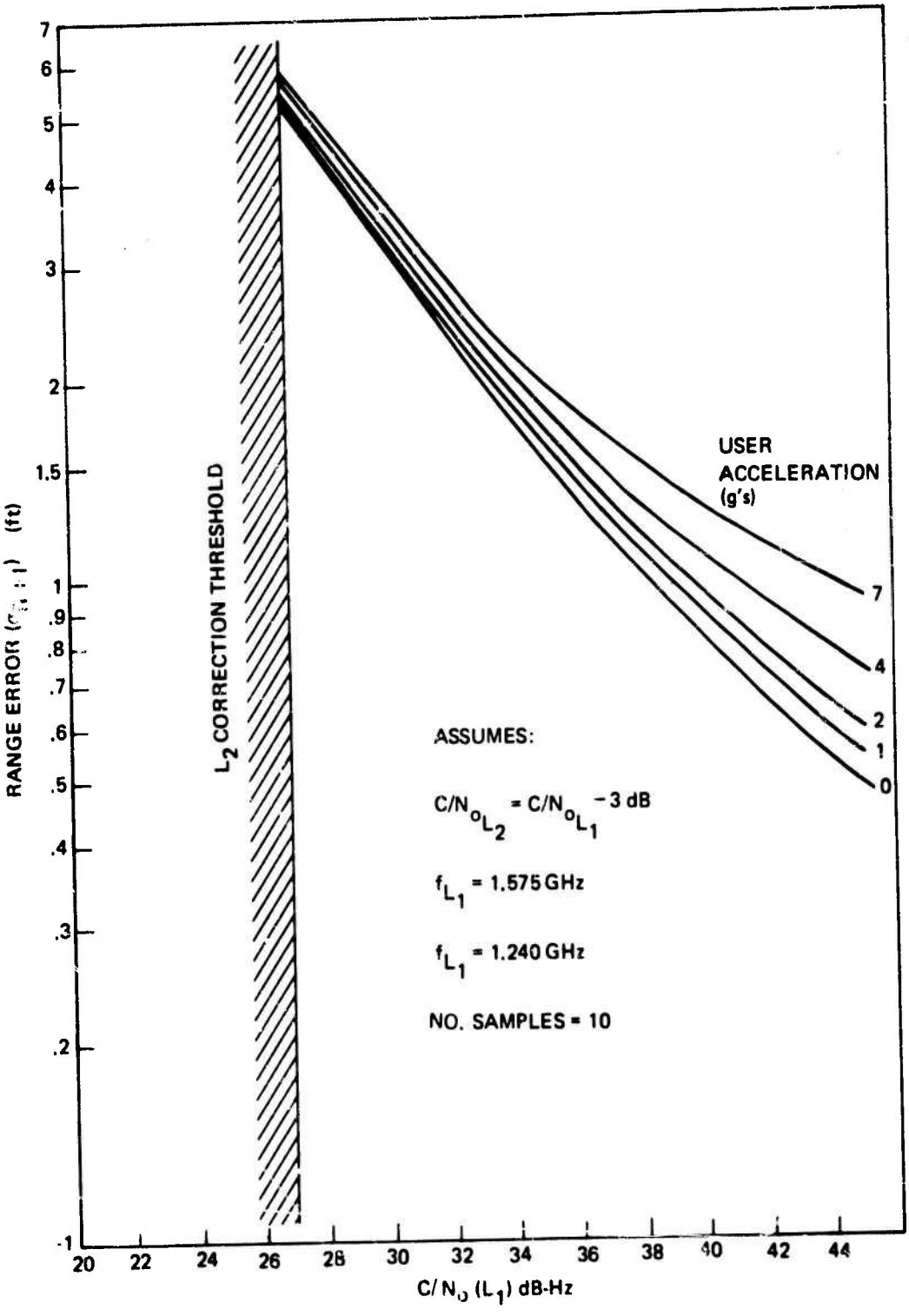


Figure 3-1 Ionospheric Correction Error (Two Frequency Technique)

The accuracy of the second technique, ionosphere modelling, is unknown at this time. Various approaches will be tested during Phase I to determine the accuracy obtainable. These will then be evaluated against the cost of implementing each to determine the optimum approach.

The third technique, fixed correction, is estimated to have a one sigma error of 30 feet.

3.1 Two Frequency Correction Accuracy

Range errors are caused by uncertainties in the speed of light through the ionosphere and in the path length. Numerous sources have agreed that the group delay velocity is inversely proportional to the square of the frequency.[1,2] Two independent analyses of the effects of the ionosphere on path length have concluded that this effect varies as the frequency to the -3.7 to -4th power. At 1.2 GHz (L_2), the increased path length is on the order of a few centimeters. It is therefore assumed that the ionospheric effect on measurements of pseudorange can be simply modelled as

$$\rho_T = \rho_i + \frac{K}{f_i^2}$$

where

ρ_T = true pseudorange

ρ_i = measured pseudorange

f_i = carrier frequency

K = a factor which depends upon the integrated electron density along the radio path when the measurement is taken.

We further assume that K is invariant between the user and a specific satellite during any 2 minute period.

---

[1] G. Millman, "Atmospheric Effects on VHF and UHF Propegation," Proc. of the IRE, Aug 1958

[2] L.J. Mitchler, "Preliminary TRW Active Ranging System Error Analysis", TRW Interoffice Correspondence, 7224.11-9, 9 April 1968

3.1.1 User's Ionospheric Correction Algorithm[3]

At regular intervals of time T, each channel of the user's receiver will make measurements of pseudorange against the L_2 signal ($\tilde{\rho}_2$). These will be compared with the last measurement made against the L_1 signal ($\tilde{\rho}_1$). The time interval between making these 2 measurements (τ) will be small (approximately 100 msec), but the geometric range between user and satellite will change significantly during that time, and will have to be taken into account. Thus,

$$\tilde{\rho}_2' = \tilde{\rho}_2 - \dot{R}\tau - 1/2\,\ddot{R}\,\tau^2 \tag{2}$$

where $\tilde{\rho}_2'$ is the L_2 pseudorange applicable at the same time that $\tilde{\rho}_1$ was measured, and $\dot{R}$ and $\ddot{R}$ are the navigation filter estimates of the geometric range rate and range acceleration. The error incurred in using this simple extrapolation is

$$\tau\Delta\dot{R} + \frac{\tau^2}{2}\Delta\ddot{R} \tag{3}$$

where $\Delta\dot{R}$ is the error in the estimate of $\dot{R}$, etc.

In practice, $\tau < 0.1$ seconds, $\Delta\dot{R} < 1$ fps and $\Delta\ddot{R} < 10$ ft/sec^2, so that the extrapolation error is insignificant.

The difference measurement, $\Delta\rho$, is then formed,

$$\Delta\rho = \tilde{\rho}_2' - \tilde{\rho}_1 \tag{4}$$

Then, n such difference measurements will be averaged

$$\overline{\Delta\rho} = \frac{1}{n}\sum^{n} \Delta\rho \tag{5}$$

---

[3] M. Deckett, "Description and Performance of the L_1/L_2 Ionospheric Correction Algorithm", TRW Interoffice Correspondence, ENSDP-MD-221, 21 Dec 1973.

This term is linearly related to the estimate of K applicable at this time,

$$K = \overline{(\Delta\rho)} \Big/ \left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right) \tag{6}$$

Thus, the ionospheric correction term, R_I, is

$$R_I = \frac{K}{f_1^2} \tag{7}$$

$$= \omega \overline{\Delta\rho}$$

where

$$\frac{1}{\omega} = \left(\frac{f_1}{f_2}\right)^2 - 1 \tag{8}$$

This term, R_I, will be subtracted from every measurement ρ_1 in the following period T. At the end of that period, a new ρ_2 measurement will be made, which will be extrapolated to $\tilde{\rho}_2'$ and will generate a new $\Delta\rho$ value. The average will then be computed again.

To reiterate, the equations of the algorithm are,

$$\tilde{\rho}_2' = \rho_2 - \dot{R}_T - 1/2\ \ddot{R}\ \tau^2 \tag{9}$$

$$\Delta\rho = \tilde{\rho}_2' - \tilde{\rho}_1 \tag{10}$$

$$\Delta\rho = \frac{1}{n} \sum^{n} \Delta\rho \tag{11}$$

$$R_I = \overline{\omega\Delta\rho} \tag{12}$$

$$\rho = \tilde{\rho}_1 - R_I \tag{13}$$

3.1.2 Accuracy of the User's Two Frequency Correction

As previously stated, errors caused by the invalidity of the model are assumed to be negligible (<< 1 foot). Also, with the expected time interval between ρ_1 and ρ_2 measurements, it was shown that the extrapolation process will not cause significant error.

The term ω will be a program constant, based on the nominal values of f_1 and f_2. It can be easily shown that variations in f_1 and f_2 from their nominal values will cause only insignificant errors in the ionospheric correction.

The only significant error sources are the measurement noise errors. These are uncorrelated with each other, and uncorrelated serially.

Thus $$\sigma^2(\Delta\rho) = \sigma^2(\tilde{\rho}_1) + \sigma^2(\tilde{\rho}_2)$$

and $$\sigma^2\overline{(\Delta\rho)} = \frac{1}{n}\sigma^2(\Delta\rho)$$

$$\sigma^2(R_I) = \frac{\omega^2}{n}\left[\sigma^2(\rho_1) + \sigma^2(\rho_2)\right]$$

The rms accuracy of the L_1 and L_2 measurements, $\sigma(\rho_1)$ and $\sigma(\rho_2)$ is shown in Table 3-2 for the 0 dB user antenna at the average elevation angle of 25°. Using these values in the equation above yields the correction error in Table 3-2. L_1 and L_2 are assumed to be 1575 and 1240 MHz, respectively. Pseudo range differences will be measured every 10 seconds for 10 measurements, thus n = 10.

TABLE 3-2

1σ Two Frequency Ionospheric Correction Error

(0 dB Antenna, 25° Elevation Angle)

User Acceleration (g's)	0	1	2	4	7
L_1 Range Error (ft) $\delta(\rho_1)$	1.25	1.35	1.45	1.60	1.9
L_2 Range Error (ft) $\delta(\rho_2)$	1.8	1.9	2.0	2.1	2.4
Net Correction Error	1.1	1.2	1.3	1.4	1.6

3.2 Fixed Bias Correction Accuracy

Two approaches were taken to estimate the accuracy of the fixed bias correction technique. These were based on a Monte Carlo simulation and Total Electron Content measurements, respectively.

3.2.1 Monte Carlo Simulation[4]

The ionospheric range error may be expressed

$$\Delta R_{ft} = K \frac{Q(E)\ N_T\ (t_d,\ t_{ss},\ t_y,\ L)\ (e/m^2)}{f^2} \qquad (1)$$

where $K = 1.3 \cdot 10^{-16}\ \frac{ft\ G^2}{e/m^2}$

$Q(E)$ = obliquity function of elevation angle

$Q(90^o) = 1$

$Q(20^o) = 2.5$

f = carrier frequency, GHz

N_T = total vertical electron columnar content, electrons/m^2

a function of:

t_d = time of day

t_y = time (season) of year

t_{ss} = time in sunspot cycle

L = latitude

---

[4] A. J. Mallinckrodt, Distribution Function of Ionospheric Errors, TRW Interoffice Correspondence, 7222.2-419, 23 February 1968

Peskoff estimates the maximum columnar content at $20 \cdot 10^{17}$ electrons/m^2. Now if we consider N_T to be a product-separable function of its various independent arguments we may write[5]

$$N_T = C_o\, f_1(t_d)\, f_2(t_y)\, f_3(t_{ss})\, f_4(L) \tag{2}$$

As a crude approximation we propose to represent each of the f_i as an independent random variate uniformly distributed over the range.

$$1 \leq f_i \leq R_i \tag{3}$$

From Peskoff's results and others we estimate the following ranges of variation

$$\begin{aligned} R_1 &\text{ (diurnal)} = 6 \\ R_2 &\text{ (seasonal)} = 2 \\ R_3 &\text{ (sunspot)} = 6 \\ R_4 &\text{ (latitudinal)} = 3 \end{aligned} \tag{4}$$

The corresponding total range from maximum to minimum is 216, so $C_o = 20 \cdot 10^{17}/216 = .92 \cdot 10^{16}$

The resulting overall distribution of N_T was found by a Monte Carlo technique, generating a series of random numbers N_T according to this distribution (eqs 2, 3, 4) and analyzing the resulting sample by the GE library program DESTAT***.

The result is shown in Figure 3-2, from which we may read the following pertinent points. The third column results from Equation 1 with $Q(20^o) = 2.5$

Fractile	N_T not exceeded Fractile of the time	ΔR_{ft} for E=20°
1.0 (maximum)	2×10^{18}	254
.9	6.5×10^{17}	83
.5 (median)	2.5×10^{17}	32

[5] A. Peskoff, "Range Errors and Other Effects of the Ionosphere on L-Band Propegation," TRW Interoffice Memorandum 3221.1-354, 22 Feb 1968.

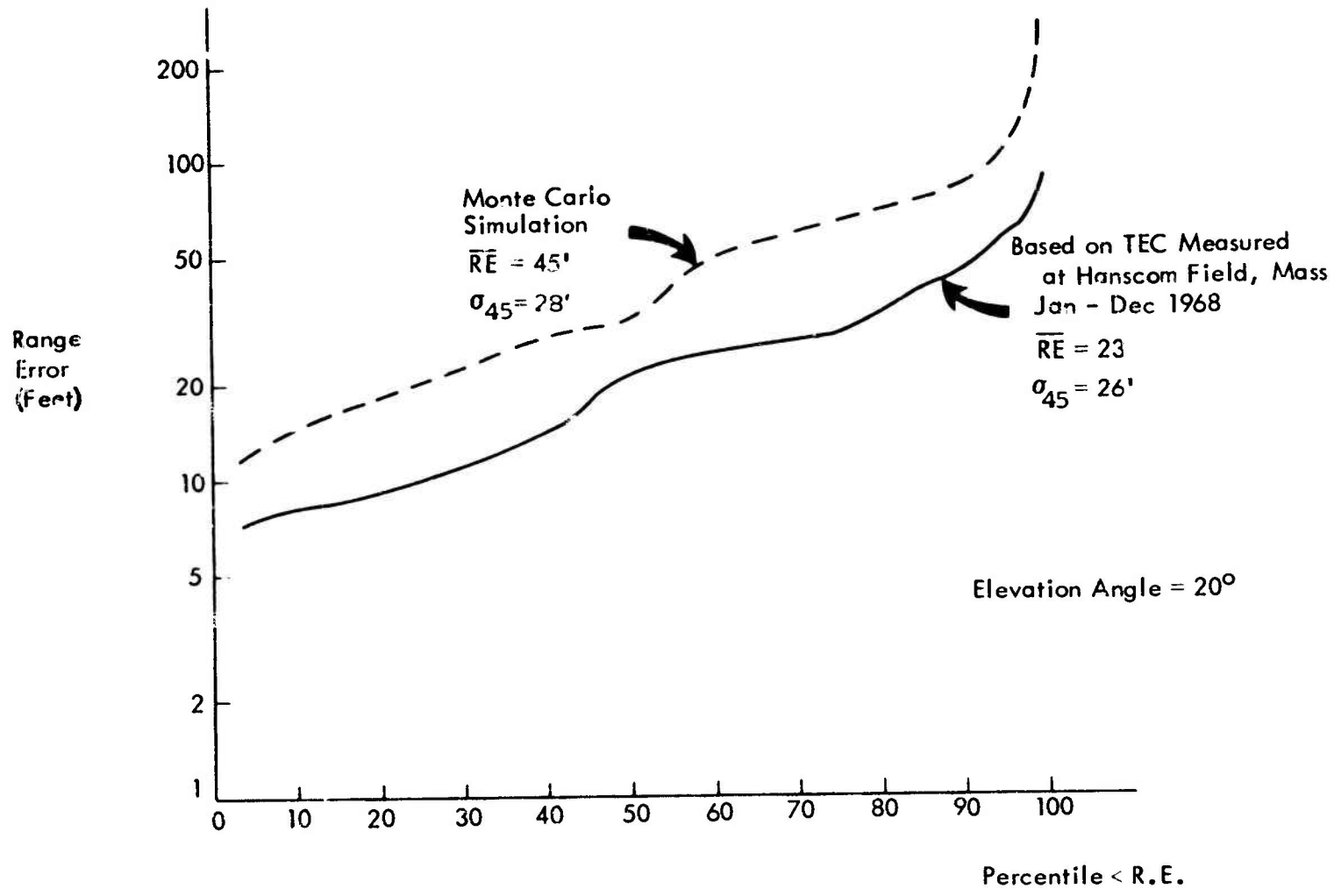


Figure 3-2 Ionospheric Error Statistics

FHILCO
Philco-Ford Corporation
Western Development Laboratories Division

The striking point of this is that since N_T arises as the product of a number of independent variables, values near the simultaneous maximum are highly improbable.

3.2.2 Total Electron Content Measurements

Approximately one year's data was collected at Hamilton, Massachusetts[6] during 1967 and 1968 using ATS-3.

Applying Equation (1) where $Q(20^{\circ}) = 2.5$ and $f = 1.575$ GHz yields the following results:

Fractile	N_T not exceeded Fractile of the time	ΔR_{ft} For $E = 20^{\circ}$
1.0 (maximum)	7×10^{17}	92
.9	3.4×10^{17}	45
.5 (median)	1.6×10^{17}	21

This is also shown on Figure 3.2.

3.2.3 Summary of Fixed Bias Analysis

It can be seen from Figure 3-2 that the slope of the two curves given by the different approaches is quite similar. The mean of the simulation was 45 feet while the mean of the measured data was 23 feet. If a fixed correction of 45 feet were applied in both cases, the standard deviations of the resulting errors are 28 feet and 26 feet, respectively. Thus, it appears that the fixed bias correction technique would provide a one sigma accuracy of 30 feet.

---

[6] J. A. Klobuchar, Total Electron Content Predictions for the Systems Engineer, Symposium on the Application of Atmospheric Studies to Satellite Transmission, p. 31-34.

3.3 Relative Accuracy

The ionospheric error changes with time because of two effects; (1) the condition of the ionosphere changes and (2) the path length through the ionosphere changes because of the satellite's motion.

1. Ionosphere Changes - Typical changes during the passage of the terminator are a linear increase from 10 Hexams to 50 Hexams * between 6AM and 12AM. This is an average of 0.11 Hexams per minute or 0.15 ft/min under the following conditions:

   Frequency - 1.575 GHz
   Elev. Angle - 10°
   H - 400km

2. Path Length Changes - The maximum rate of change of ionospheric path length with time occurs at approximately 15° elevation angle. During the 30 minutes the satellite passes from 21° elevation angle to 10° elevation angle, the ionospheric effect on path length increases from 27.8 feet to 35.4 feet. This is assuming the ionosphere has a TEC of 25 Hexams, H is 400km and the frequency is 1.575 GHz. This is an average of 0.25 ft/minute.

Since these errors can be correlated, the net effect is an error which can increase at the rate of 0.40 feet/minute. Thus, under typical adverse conditions, two frequency sequential measurements should not be averaged over a period longer than 2 minutes if the error contributions of ionospheric condition changes and path length changes are to be held to less than 1 foot.

The fixed bias correction technique user receiver will have a relative range error which varies as shown in Table 3-3.

TABLE 3-3

Relative Ranging Error of Fixed Correction User

Time since original meas. (minutes)	5	10	20	30	60
Relative Ranging Error (ft)	2	4	8	12	24

*Hexam $\triangleq$ 1×10^{16} (electrons/m^2) column

4.0 TROPOSPHERIC ERRORS

The lowest layer of the atmosphere, the troposphere, extends from the earth's surface up to approximately 30 Km. Within this region the dielectric constant varies slightly and is dependent upon humidity, temperature and pressure. We have assumed the CRPL* Exponential Reference Atmosphere (1958) as our model and have examined the effect of varying certain parameters of that model. The model is used to produce values of refractivity as a function of altitude with equation parameters of sea level pressure, sea level temperature, user relative humidity, and temperature lapse rate. It is fully recognized that the CRPL atmosphere is only a statistical entity and, as such, is liable to be wrong at any particular point in space and time. The alternative to using the CRPL model requires the user to make atmospheric measurements (temperature, pressure, humidity) along the signal path, which is clearly impractical outside of a research experiment. Even if the user makes ground measurements at his location, he must assume some model to predict refractivity along the signal path. The alternatives to using a statistical model have not been addressed in this study. The following treatment is an error analysis (which is inherently statistical in nature) and since we are concerned with errors on the order of "1-sigma" variations, a treatment using the CRPL model is the most practical approach.

Thus, a radio signal propagating through the troposphere experiences a transit time delay which can be derived by integrating the refractivity from the position of the user along its path (assuming a model atmosphere) to the "top" of the troposphere.

To compensate for this time delay, a GPS user will employ a simplified and approximate algorithm which is only a function of altitude and elevation angle.

---

* Central Radio Propagation Laboratory, National Bureau of Standards (now the Institute for Telecommunication Sciences)

The purpose of this section is to first compare the time delay as calculated by this simplified algorithm to the time delay as calculated by a detailed numerical integration of refractivity based on the CRPL model atmosphere; and second, to examine the excursions of this time delay due to 1-sigma fluctuations in the model atmosphere parameters.

The user will employ the following algorithm to determine troposphere range correction (in feet):

$$\rho_T = K_T \, \mathrm{Csc}\, E \; e^{\frac{-h}{23000}}$$

where E is the elevation angle to the satellite, K_T is a constant (approximately 7.8 ft) and h is the receiver altitude in feet. This correction factor must be recalculated for each range measurement, as satellite motion can cause the tropospheric effect to change at a rate of 7 feet per minute.

Absolute Range - The error in the correction for the troposphere is a function of elevation angle and receiver altitude as well as variations from the assumed standard atmosphere. Figure 4-1 shows the one sigma error probability distribution for users at sea level and 10,000 feet. If one assumes a Phase I satellite elevation angle distribution (shown in Figure 4-2), the average elevation angle experienced by the user is 28 degrees (see Section 4.6).

Figure 4-3 presents a crossplot of this information, showing the percentage of users with errors less than a given value. Table 4-1 summarizes the one sigma tropospheric range errors experienced by "average" Phase I users at sea level and 10,000 feet due to atmospheric fluctuations.

TABLE 4-1
1 Sigma Troposphere Range Error

	1 Sigma User Error (ft)	
	Sea Level	10,000 ft
Phase I:	0.91	0.56

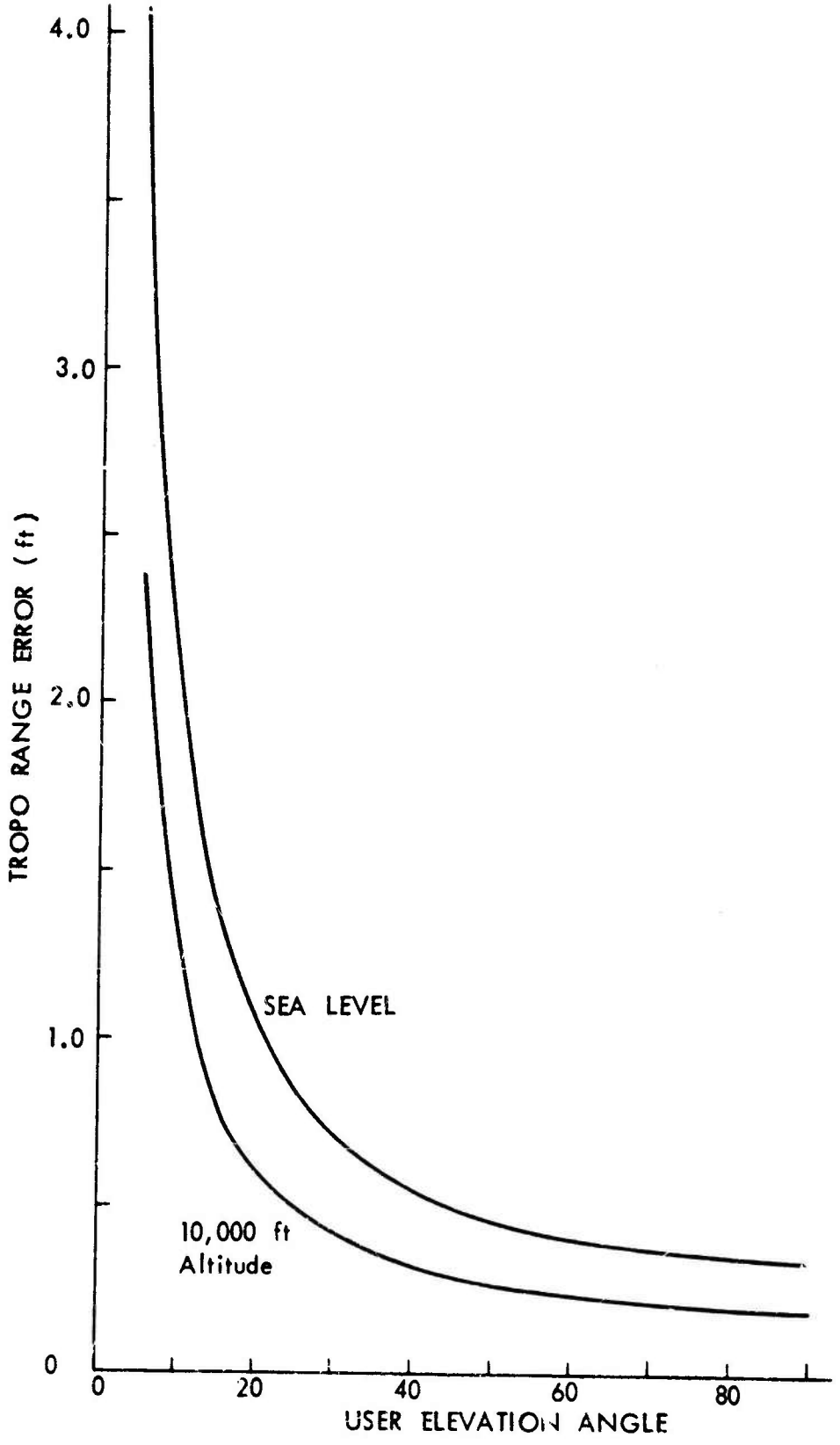


FIGURE 4 - 1 One Sigma Error Due To Deviations From Assumed Standard Atmosphere

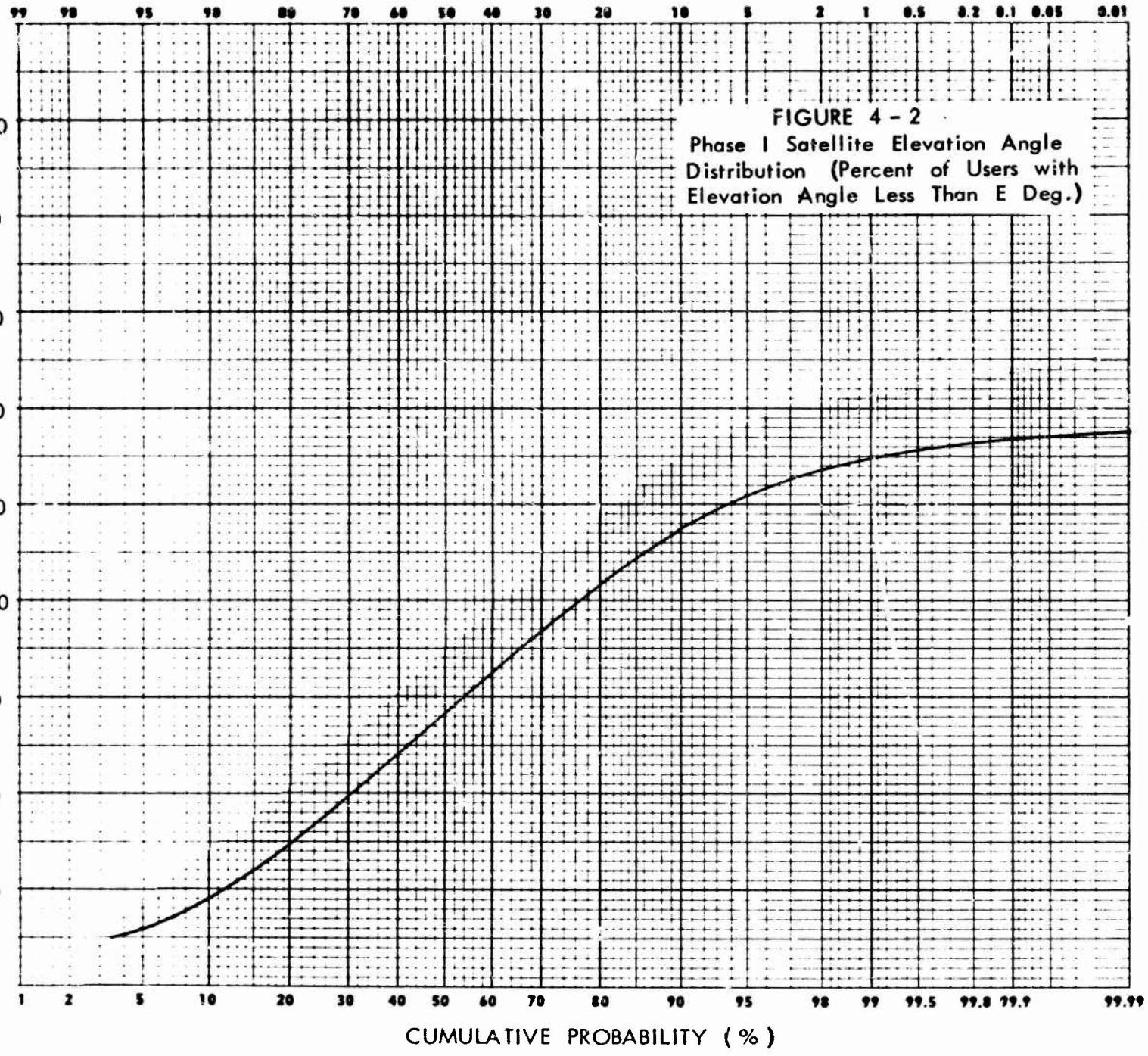
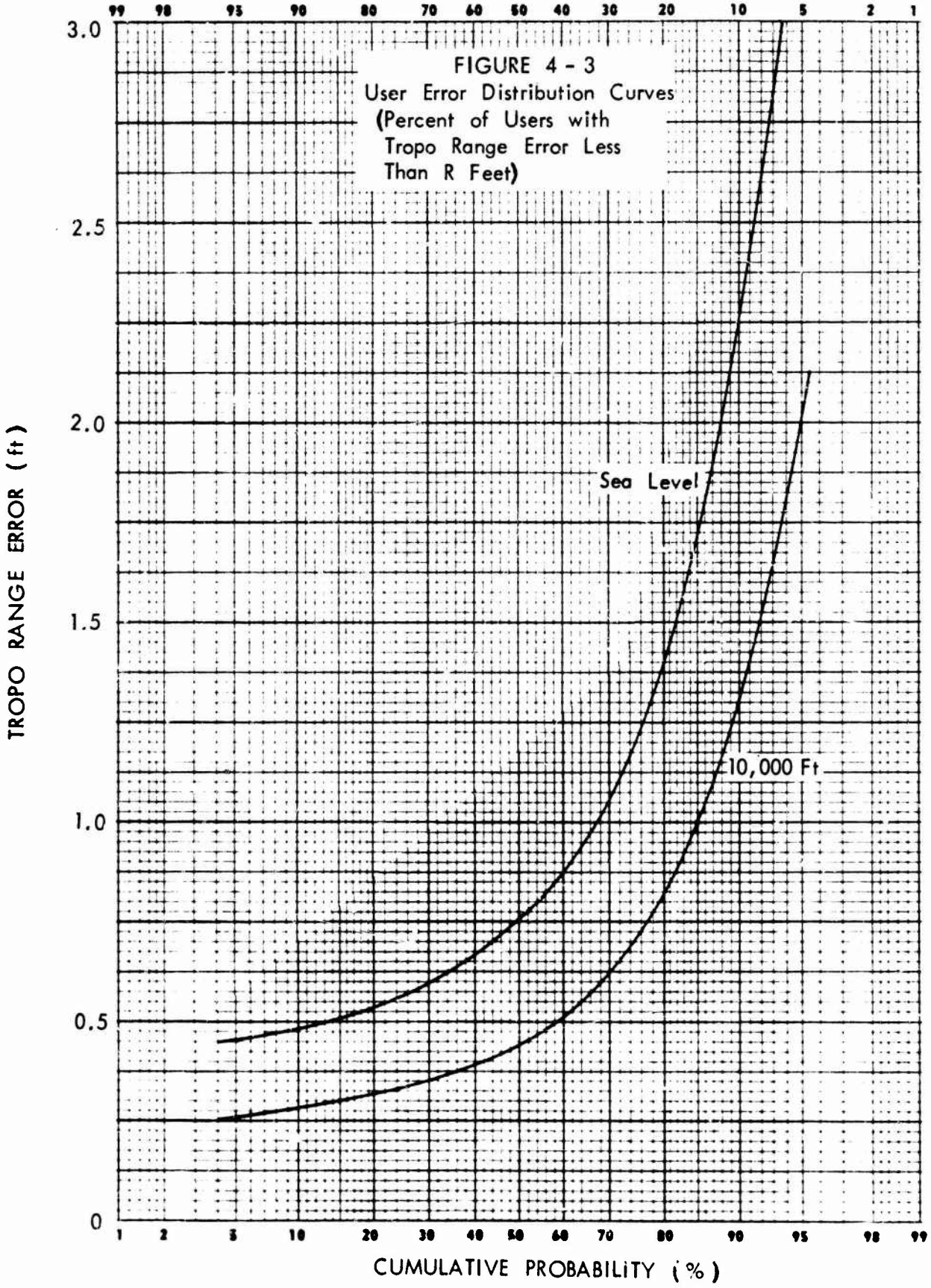


FIGURE 4 - 2
Phase I Satellite Elevation Angle Distribution (Percent of Users with Elevation Angle Less Than E Deg.)


FIGURE 4 - 3
User Error Distribution Curves
(Percent of Users with
Tropo Range Error Less
Than R Feet)
Sea Level
10,000 Ft
TROPO RANGE ERROR (ft)
CUMULATIVE PROBABILITY (%)
3.0
2.5
2.0
1.5
1.0
0.5
0
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1
1 2 5 10 20 30 40 50 60 70 80 90 95 98 99

Relative Range - The errors can be expected to be highly correlated over a limited geographic area over a period of approximately one hour. Data on the time-rate-of-change of lapse rate, pressure, temperature and relative humidity[1] during storms and the most severe periods of the day indicate the following:

1 - sigma fluctuations (as defined in this section) can occur over a 5.8 to 8.3 hour period.

The RSS worst-cast total rate-of-change of all these fluctuations amounts to 0.56 ft/hr at sea level at 5 deg elevation.

Expressed as a percentage of the full tropospheric range correction at these conditions (89.55 ft), these time fluctuations amount to 0.6% per hour.

While these atmospheric parameters will not change by more than about 1 sigma during a 6 hour period, the movement of the satellite will cause some change in the path conditions due to area gradients. It is assumed that this will essentially double the rate of change of this error. Thus, the net increase in relative tropospheric error is 2 x 0.91 feet per hour (0.03 feet/minute) or 0.6 feet after 20 minutes. It must be pointed out that the user equipment must use the algorithm stated above on every measurement in order to achieve this accuracy.

4.1 Tropospheric Error Analysis

Time delays in the troposphere are a function of the index of refraction. Unlike the ionosphere, the troposphere is largely non-conducting and the index of refraction is not a function of the transmission carrier frequency. The refractive index at any altitude within the troposphere can be expressed[2] as:

$$n = 1 + \frac{C}{T}\left(P + \frac{be}{T}\right) 10^{-6}$$

where

$C = 77.6^{\circ}$ K/MB

$b = 4810^{\circ}$K

P = pressure (MB)

e = partial pressure of water vapor (MB)

T = temperature $^{\circ}$K

It is possible to express the above equation as a function of altitude up to 10 KM for a standard atmosphere: by substituting (Reference 3)

$$P = P_o \left(\frac{T}{T_o}\right)^{\frac{1}{\alpha R}}$$

$$T = T_o - \alpha h$$

$$e = e_o (1 - 0.064h)^6$$

To get

$$n = 1 + \left(\frac{C\ P_o}{T_o} \left(1 - \frac{\alpha h}{T_o}\right)^{\left(\frac{1}{\alpha R} - 1\right)} + \frac{Cbe_o\ (1 - .064h)^6}{(T_o - \alpha h)^2} \right) 10^{-6}$$

where

α = Temperature Lapse Rate, nominally 6.5 $\frac{^oK}{KM}$

R = Gas Constant = 0.0292 $\frac{KM}{^oK}$

P_o = Sea level pressure nominally 1013 MB

T_o = Sea level temperature nominally 288 ^oK

h = altitude - KM

Above 10 KM, the refractive index is assumed to decay exponentially as

$$n = N_o\ e^{-\frac{h}{7.62}}$$

where h is the altitude in KM, and N_o is the refractive index at the earth's surface.

The range error is:

$$\Delta R = \int_s dS - \int_s \frac{dS}{n} \approx \int_s (n-1)\ dS$$

Since the path length change (due to ray bending) is small compared to the transit time delay, we can assume a linear path and numerically integrate this expression. Figure 4-4 shows the result of a 200-layer trapezoidal integration and compares it with the simplified algorithm presented above (Section 4.0).

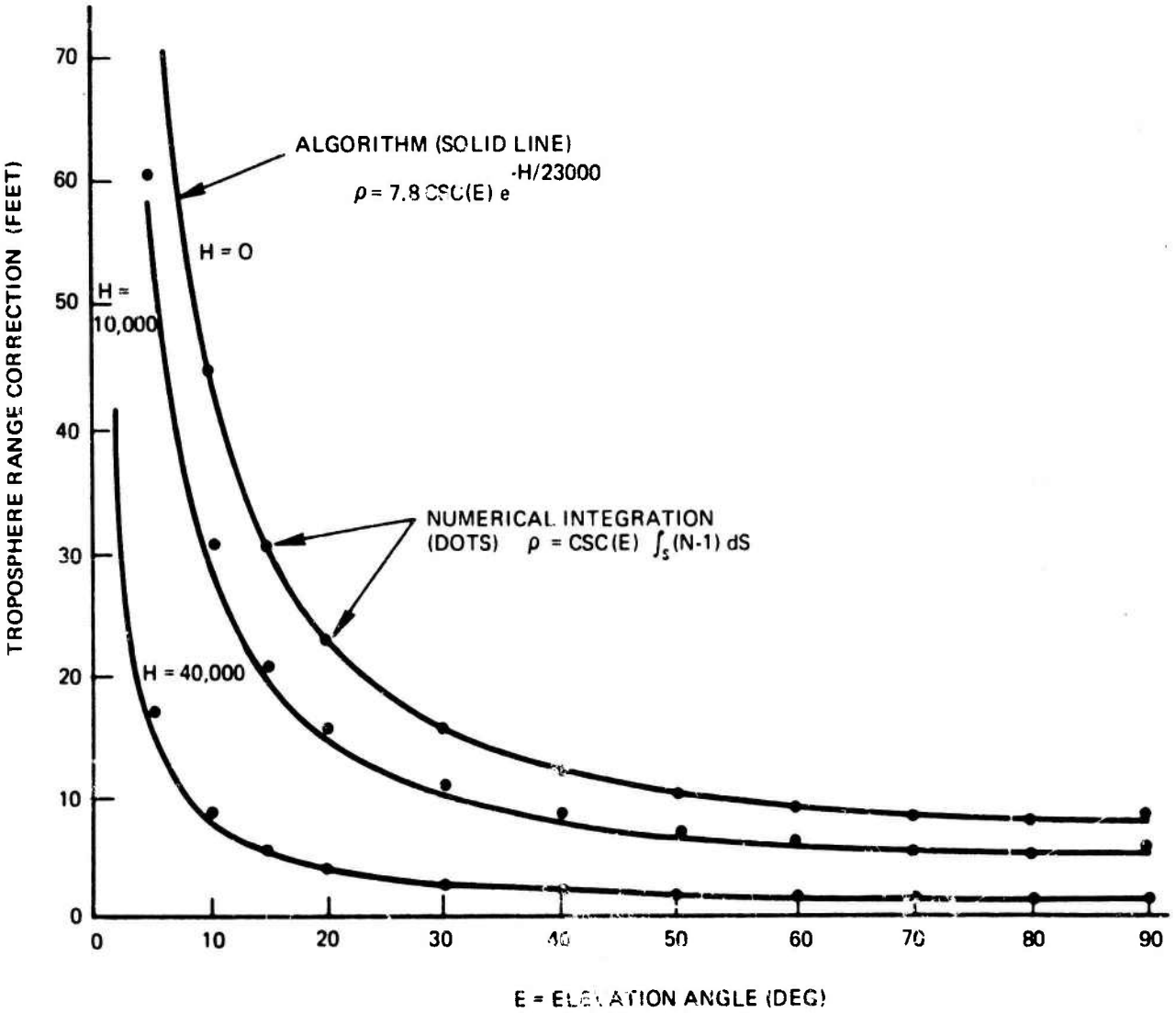


FIGURE 4-4 Tropo Range Correction
(Comparison of Algorithm
With Numerical Integration)

4.2 Humidity Errors

Relative humidity is the ratio e/e_s, expressed as a percentage, at the location of the user. The symbol, e, refers to the partial pressure of water vapor at the altitude of interest (see Sec. 4.1), and e_s refers to the saturated water vapor pressure which is a function of temperature. Using the equation for temperature as a function of altitude (Sec. 4.1) we can thus obtain e_s as a function of altitude.[4]

Altitude (ft)	Temp (^oK)	e_s (mB)
0	288	17.0
10,000	268	4.2
40,000	208.7	0.007

Assuming that the 0% and 100% humidity cases are 2-sigma values, and that humidity is normally distributed about 50%, then the 1-sigma humidities are 50 $\pm$ 25%. Table 4-2 presents the nominal (50%) values of the range correction and the 1- and 2-sigma deviation magnitudes. The tabulated values assume a cosecant dependence of range with elevation angle and are the result of a 200 layer numerical integration between the altitude of the user and 30 Km (98,400 ft) which is assumed to be the upper boundary of the troposphere. (One test case was run to observe the effect of a 40 Km troposphere boundary, with the result that the range error correction increased by 1.46 percent.)

4.3 Sea Level Pressure Errors

Table 4-3 shows values of the range correction at the assumed nominal sea level pressure of 1013 mB, and the effect of a 3-sigma variation of $\pm$ 57 mB about this value. Once again, pressures are assumed normally distributed about the nominal value and range path lengths have a cosecant dependence on elevation angle. The table presents 1- and 3-sigma deviations as a function of elevation angle at sea level, 10,000 ft, and 40,000 ft.

TABLE 4-2

HUMIDITY ERRORS (FT)

NOMINAL HUMIDITY = 50%

ALTITUDE = 0 FT

ELEV ANGLE	±50% (±2 SIGMA)	±25% (±1 SIGMA)	NOMINAL
5	6.09	3.04	89.55
10	3.05	1.53	44.95
20	1.55	.78	22.82
30	1.06	.53	15.61
40	.83	.41	12.14
50	.69	.35	10.19
60	.61	.31	9.01
70	.56	.28	8.31
80	.54	.27	7.93
90	.53	.27	7.80

ALTITUDE = 10000 FT

ELEV ANGLE	±50% (±2 SIGMA)	±25% (±1 SIGMA)	NOMINAL
5	3.51	1.75	60.30
10	1.76	.88	30.26
20	.89	.45	15.37
30	.61	.31	10.51
40	.48	.24	8.18
50	.40	.20	6.86
60	.35	.18	6.07
70	.33	.16	5.59
80	.31	.16	5.34
90	.31	.15	5.26

ALTITUDE = 40000 FT

ELEV ANGLE	±50% (±2 SIGMA)	±25% (±1 SIGMA)	NOMINAL
5	1.95	.97	16.68
10	.98	.49	8.37
20	.50	.25	4.25
30	.34	.17	2.91
40	.26	.13	2.26
50	.22	.11	1.90
60	.20	.10	1.68
70	.18	.09	1.55
80	.17	.09	1.48
90	.17	.08	1.46

ready

TABLE 4-3

SEA LEVEL PRESSURE (Ft)

NOMINAL PRESSURE = 1013 MB

ALTITUDE = 0 FT

ELEV ANGLE	±57 MB (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	4.65	1.55	89.55
10	2.33	.78	44.95
20	1.18	.39	22.82
30	.81	.27	15.61
40	.63	.21	12.14
50	.53	.18	10.19
60	.47	.16	9.01
70	.43	.14	8.31
80	.41	.14	7.93
90	.41	.14	7.80

ALTITUDE = 10000 FT

ELEV ANGLE	±57 MB (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	3.31	1.10	60.30
10	1.66	.55	30.26
20	.84	.28	15.37
30	.58	.19	10.51
40	.45	.15	8.18
50	.38	.13	6.86
60	.33	.11	6.07
70	.31	.10	5.59
80	.29	.10	5.34
90	.29	.10	5.26

ALTITUDE = 40000 FT

ELEV ANGLE	±57 MB (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	1.00	.33	16.68
10	.50	.17	8.37
20	.25	.08	4.25
30	.17	.06	2.91
40	.14	.05	2.26
50	.11	.04	1.90
60	.10	.03	1.68
70	.09	.03	1.55
80	.09	.03	1.48
90	.09	.03	1.45

ready
*

4.4 Sea Level Temperature Errors

Table 4-4 shows a similar effect with the variation of sea level temperatures about an assumed nominal sea level temperature of 288 deg. K. The 3-sigma variation is assumed to be $\pm$ 26 deg.

4.5 Temperature Lapse Rate Errors

Table 4-5 shows the effect of a $\pm$ 1.7 deg/Km 3-sigma deviation of the lapse rate constant from the assumed nominal value of 6.5 deg/Km. Once again, the nominal value, the 3-sigma, and the 1-sigma deviations are tabulated versus user elevation angle for sea level, 10,000 ft, and 40,000 ft user altitude.

TABLE 4-4

SEA LEVEL TEMPERATURE(Ft)

Nominal Temperature = 288 Deg K

ALTITUDE = 0 FT

ELEV ANGLE	±26 Deg (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	6.06	2.02	89.55
10	3.04	1.01	44.95
20	1.54	.51	22.82
30	1.06	.35	15.61
40	.82	.27	12.14
50	.69	.23	10.19
60	.61	.20	9.01
70	.56	.19	8.31
80	.54	.18	7.93
90	.53	.18	7.80

ALTITUDE = 10000 FT

ELEV ANGLE	±26 Deg (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	3.46	1.15	60.30
10	1.74	.58	30.26
20	.88	.29	15.37
30	.60	.20	10.51
40	.47	.16	8.18
50	.39	.13	6.86
60	.35	.12	6.07
70	.32	.11	5.59
80	.31	.10	5.34
90	.30	.10	5.26

ALTITUDE = 40000 FT

ELEV ANGLE	±26 Deg (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	1.67	.56	16.68
10	.84	.28	8.37
20	.42	.14	4.25
30	.29	.10	2.91
40	.23	.08	2.26
50	.19	.06	1.90
60	.17	.06	1.68
70	.15	.05	1.55
80	.15	.05	1.48
90	.15	.05	1.45

ready

TABLE 4-5

TEMPERATURE LAPSE RATE

Nominal Lapse Rate = 6.5 Deg/Km

ALTITUDE = 0 FT

ELEV ANGLE	±1.7 o/km (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	1.19	.40	89.55
10	.60	.20	44.95
20	.30	.10	22.82
30	.21	.07	15.61
40	.16	.05	12.14
50	.14	.05	10.19
60	.12	.04	9.01
70	.11	.04	8.31
80	.11	.04	7.93
90	.10	.03	7.80

ALTITUDE = 10000 FT

ELEV ANGLE	±1.7 o/km (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	.95	.32	60.30
10	.47	.16	30.26
20	.24	.08	15.37
30	.16	.05	10.51
40	.13	.04	8.18
50	.11	.04	6.86
60	.10	.03	6.07
70	.09	.03	5.59
80	.08	.03	5.34
90	.08	.03	5.26

ALTITUDE = 40000 FT

ELEV ANGLE	±1.7 o/km (±3 SIGMA)	(±1 SIGMA)	NOMINAL
5	.36	.12	16.68
10	.18	.06	8.37
20	.09	.03	4.25
30	.06	.02	2.91
40	.05	.02	2.26
50	.04	.01	1.90
60	.04	.01	1.68
70	.03	.01	1.55
80	.03	.01	1.48
90	.03	.01	1.45

4.6 Summary of Errors

Table 4-6 shows the summary of 1-sigma troposphere parameter errors tabulated at elevation angles of 90 deg and 10 deg. The RSS values of each column are shown as the last entry in the column. Figure 4-1 (above) follows directly from the RSS totals at 90 deg elevation by introducing a cosecant factor to show approximate elevation angle behavior.

Figures 4-5 and 4-6 are differential versions of Figures 4-2 and 4-3. Figure 4-5 shows the normalized distribution function of the Phase I user elevation angles. The most probable elevation angle is seen to be 27.5 deg. Figure 4-6 presents the sea level and 10,000 ft distribution functions of user error, clearly showing the peaks of most-probable error.

TABLE 4-6

1-SIGMA TROPOSPHERE ERRORS (FT)
AT 90 AND 10 DEG. ELEVATION ANGLES

PARAMETER	USER AT SEA LEVEL		USER AT 10,000 Ft Alt	
	E = 90°	E = 10°	E = 90°	E = 10°
Relative Humidity ($\pm 2\sigma = \pm 50\%$)	0.27	1.53	0.15	0.88
Sea Level Pressure ($\pm 3\sigma = \pm 57$ mB)	0.14	0.78	0.10	0.55
Sea Level Temperature ($\pm 3\sigma = \pm 26$ deg)	0.18	1.01	0.10	0.58
Temp. Lapse Rate ($\pm 3\sigma = \pm 1.7$ deg/Km)	0.03	0.20	0.03	0.16
RSS Total:	0.355	2.002	0.208	1.200

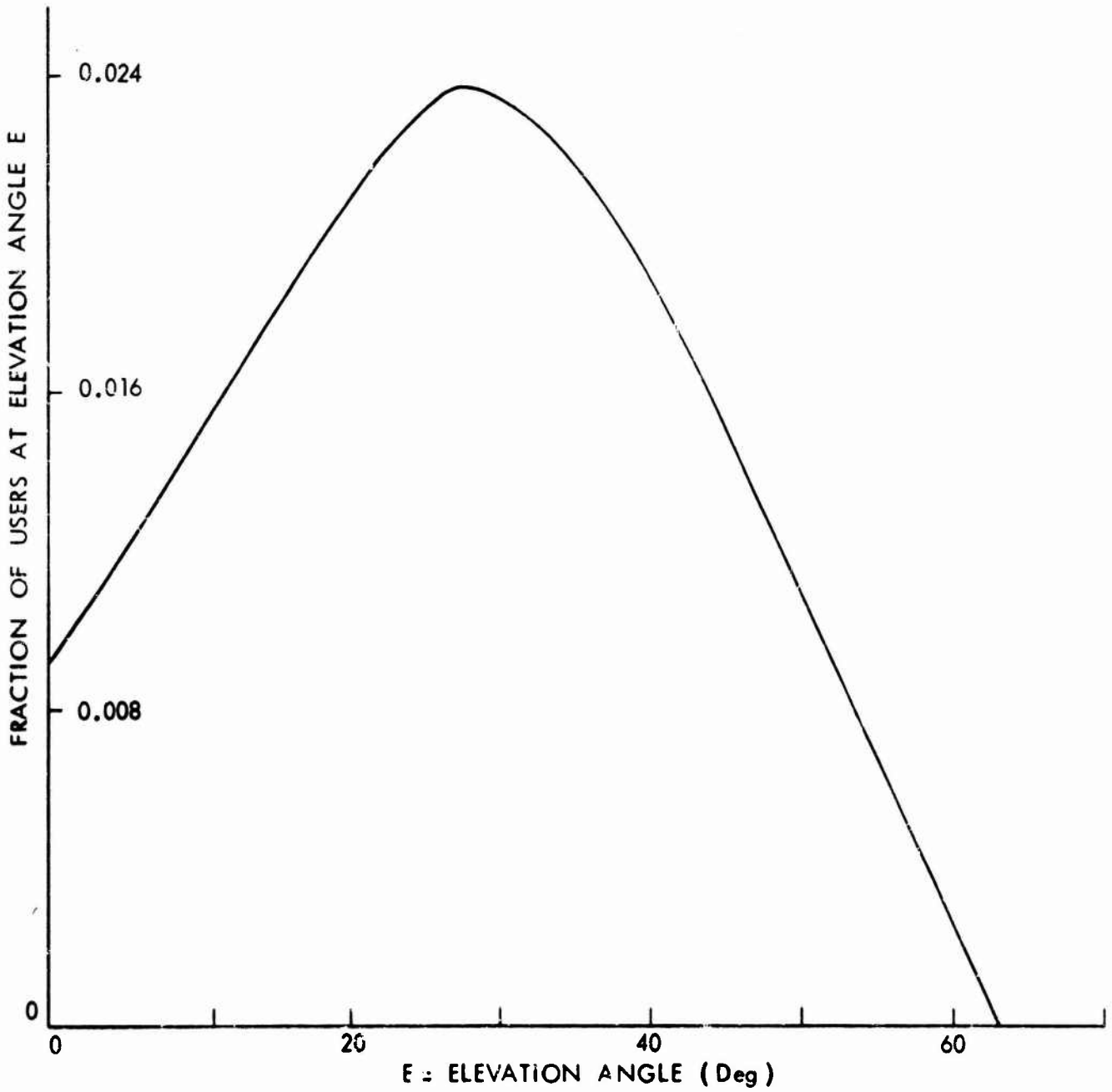


FIGURE 4-5 Normalized Distribution Of User Elevation Angles

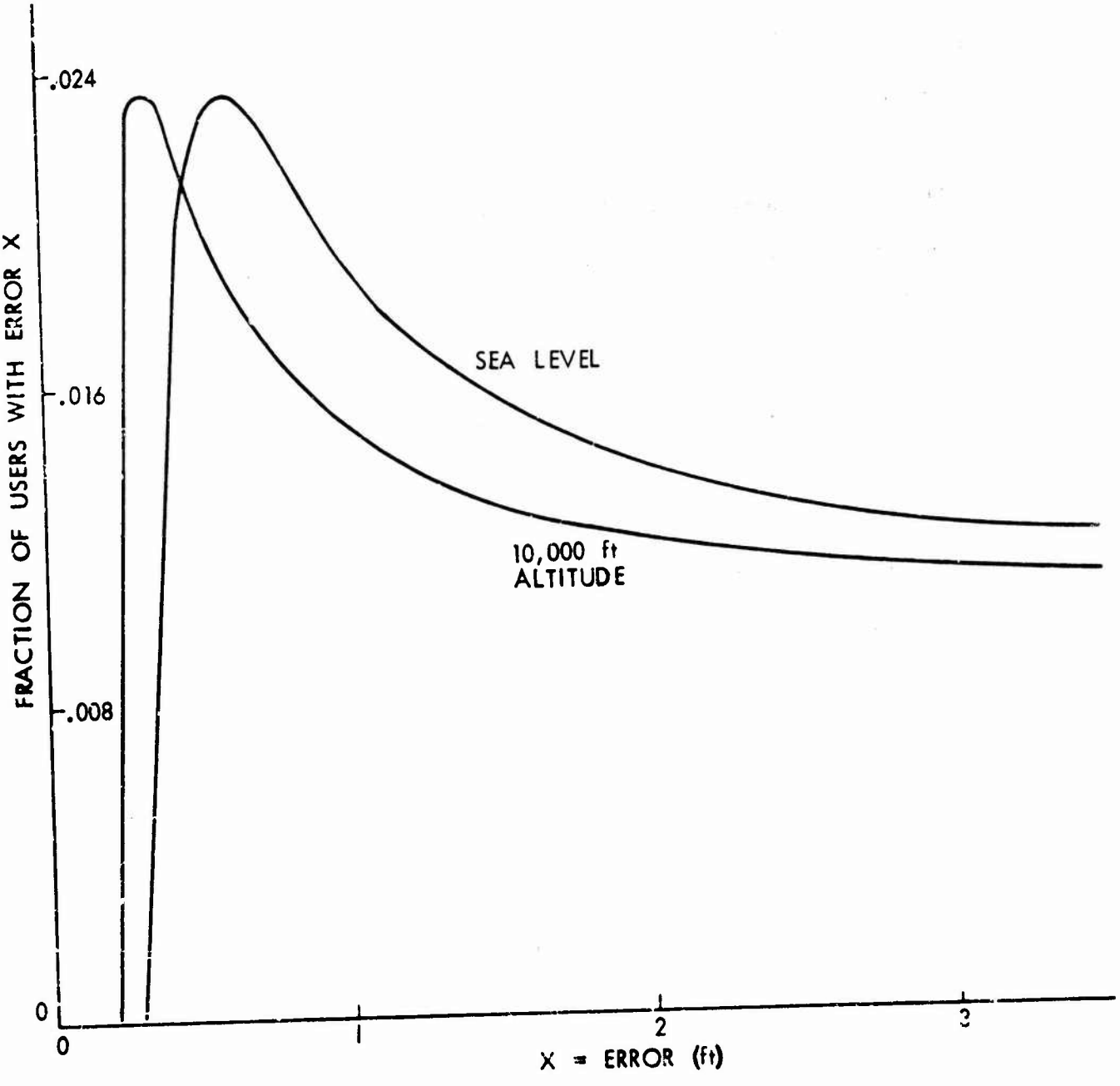


FIGURE 4-6 Distribution of Phase I Users With Error Due To Atmosphere Fluctuations

REFERENCES

1. Arthur N. Strahler, Intorduction to Physical Geography, John Wiley and Sons, Inc., July 1965.

2. Millman, G. H., Atmospheric and Extraterrestial Effects on Radio Wave Propagation, General Electric Technical Information Series, R61EMH29, June 1961.

3. L. J. Mitchler, Preliminary TRW Active Ranging System Error Analysis, No. 7224.11-9, 9 April 1968.

4. Handbook of Chemistry and Physics (43rd Edition) Vapor Pressure of Water, Pp 2361-68.

5.0 MULTIPATH ERROR

This section analyzes the effect of specular multipath signals upon the operation of the Delay Lock Loop in the user receiver. Section 5.1 describes the conditions under which this multipath will take place and Section 5.2 determines the error caused by this type of reflection.

5.1 Characteristics of the Multipath Signal

A multipath signal results when the desired signal arrives via the direct path and a reflected path. The reflected path can result from ground reflections or vehicle body reflections, such as reflections from the wing of an aircraft. The reflected signal is composed of a "specular" part and a "diffuse" part. Which type of reflection dominates depends primarily on the roughness of the reflecting surface compared to the wavelength of the signal. The Rayleigh criteria is often used to differentiate between the two extremes. According to the Rayleight criteria*, the surface is considered smooth if:

$$\frac{\sigma_\omega \sin E}{\lambda} < \frac{1}{8}$$

where σ_ω = rms surface height variation

E = elevation angle

λ = RF signal wavelength

Figure 5-1 shows the maximum rms surface height for specular reflection as a function of elevation angle assuming λ equals 0.625 feet (f = 1.574 GHz). It can be seen that a surface roughness greater than one foot produces diffuse reflections at the elevation angles of interest (>5 degrees). Thus, almost all reflections resulting from the ground will be predominantly diffuse in nature. Diffuse type reflections tend to spread the reflected power over a wider bandwidth than the original signal and, therefore, are less troublesome than specular type reflections. For the P-signal, all reflections resulting from path differences greater than 150 feet will be reduced by at least 60 dB. For the C/A-signal, all reflections resulting from path differences greater than 1500 feet will be reduced by 30 dB. Thus ground reflections should not cause any loss in P-signal performance, and will probably have a negligible effect upon C/A-signal performance. However, reflections off the body of an aircraft or a smooth water surface will normally be specular in nature and have differences less than 150 feet. The effect of these type reflections on delay lock loop performance are considered in the next section.

* Beckmann and Spizzichino, "The Scattering of Electromagnetic Waves from Rough Surfaces," pg. 9, Pergamon Press, 1963

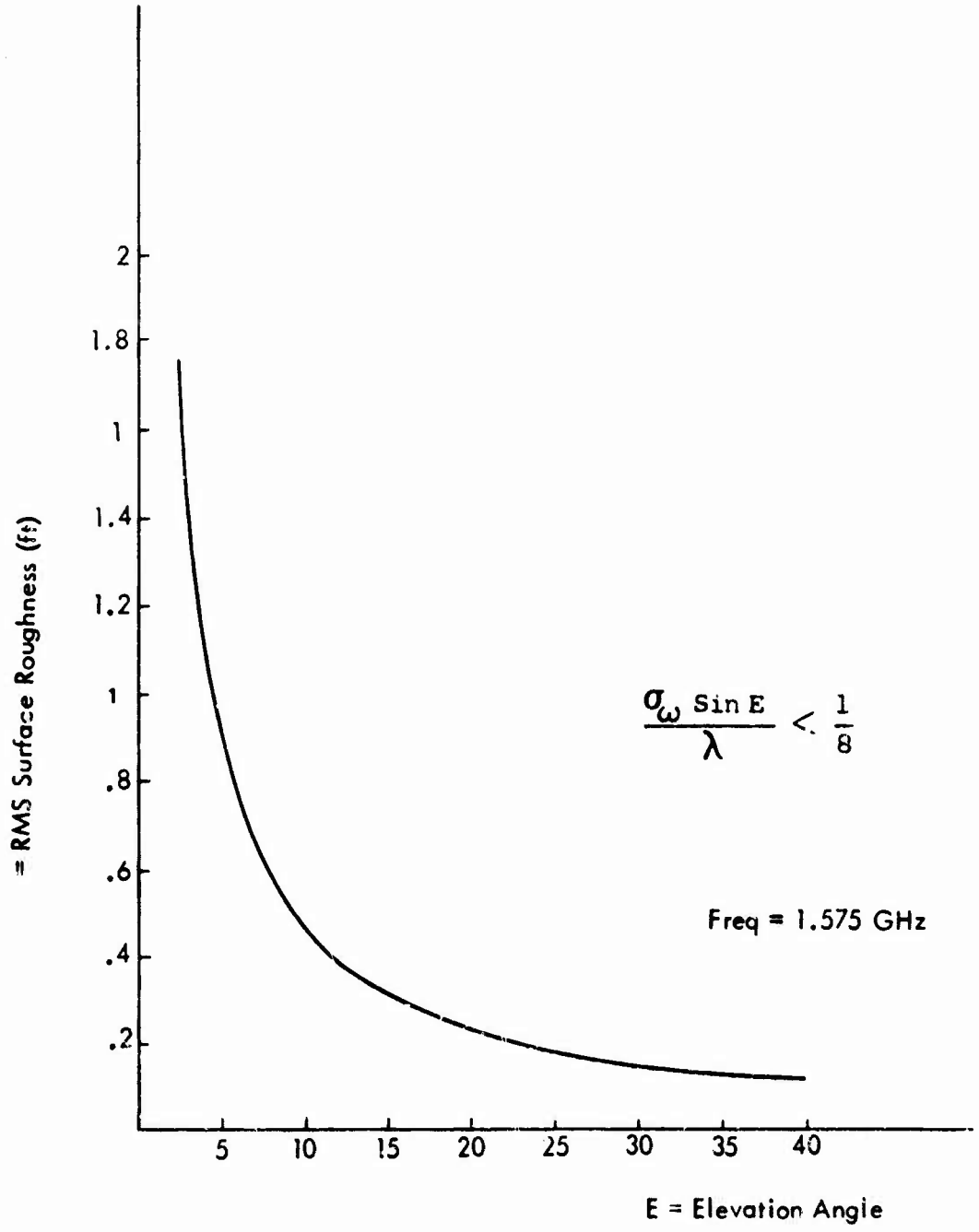


Figure 5-1

Maximum Surface Roughness for Specular Reflection vs Elevation Angle

5.2 Delay Lock Loop Effects

The purpose of this analysis is to develop a model of the multipath errors in GPS measurements of pseudorange and pseudorange rate. The errors are not modeled for one specific receiver, but generally for any single channel, timeshared, GPS receiver. Where necessary, this has been done by looking at worst case situations. The majority of the following analysis was initially conducted for the 621B program. (1)

5.2.1 Summary of Results

5.2.1.1 Pseudorange Error

Except when the aircraft is near the ground, the multipath error in pseudorange is distributed about zero mean with a standard deviation of 2 feet. It is caused by multipath reflections from the aircraft body.

During an ILS approach, an additional, larger, and more deterministic multipath error is present due to multipath reflections from the ground. Its value is $\bar{\ell}$, where

$$\bar{\ell} = -\delta \quad \text{, for } 0 < \delta < 14.6$$

$$\bar{\ell} = .3366 - 19.5 \quad \text{, for } 14.6 < \delta < 85.5$$

$$\bar{\ell} = .143(150 - \delta) \quad \text{, for } 85.5 < \delta < 150$$

and

$$\bar{\ell} = 0 \quad \text{, for } 150 < \delta$$

The parameter δ is the multipath distance, and is given by

$$\delta = 2h \sin \theta \quad \text{, for a satellite transmitter}$$

$$\delta = \frac{2hb}{d} \quad \text{, for a beacon transmitter}$$

where h = aircraft altitude

θ = satellite elevation angle

b = beacon antenna height (assume b = 10 ft.)

and d = horizontal distance, beacon to aircraft.

A plot of $\bar{\ell}$ versus δ is shown in Figure 5-2.

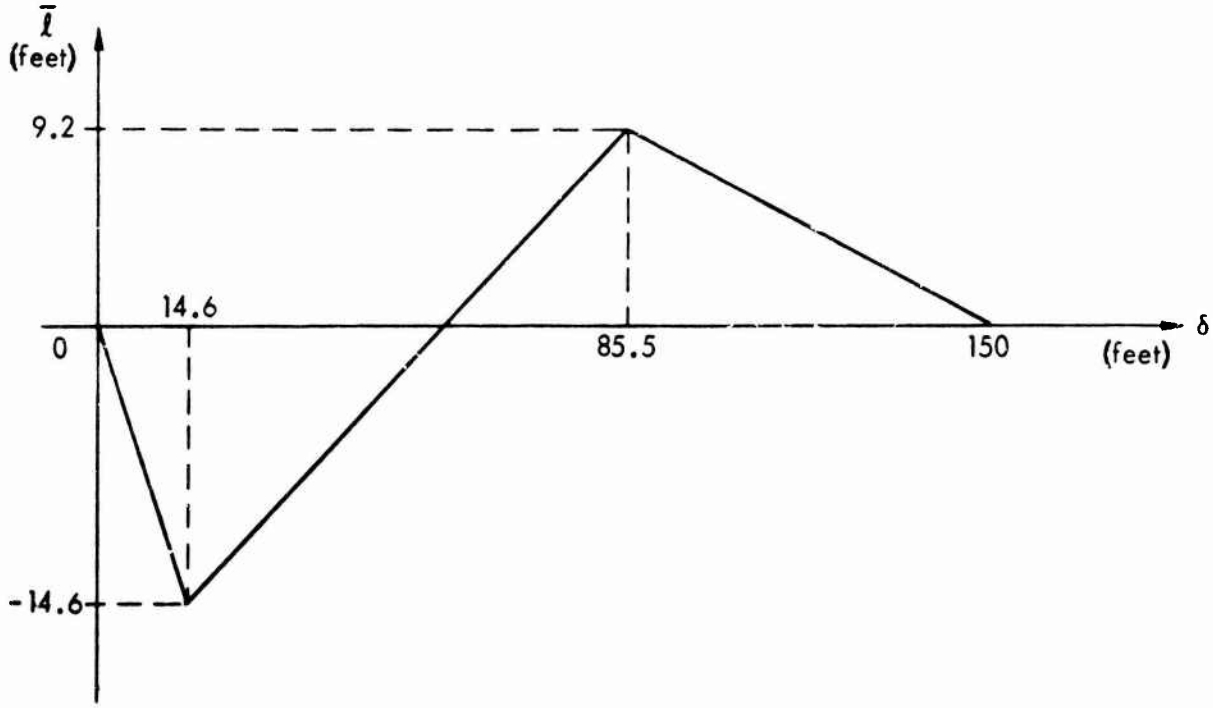


Figure 5-2 Range Error During an ILS Approach

5.2.1.2 Pseudorange Rate Error

The multipath error in pseudorange rate due to ground reflections is not significantly different from that due to aircraft body reflections.

This error is statistically distributed about zero mean, and has a standard deviation given by

$$\sigma(\dot{R}) = \frac{0.01}{T} \text{ fps}$$

where T = range rate counter sample interval, in seconds.

5.2.2 RANGE ERROR

5.2.2.1 Discriminator Characteristic

The generalized discriminator characteristic curve of a tau-dithered delay lock loop, such as is used in GPS time shared receivers, is shown in Figure 5-3.

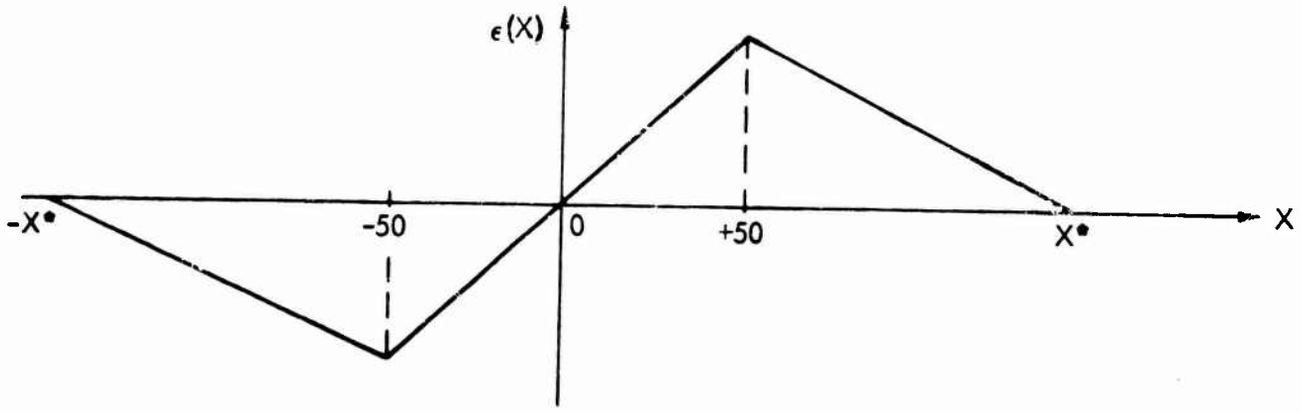


Figure 5-3 Generalized Discriminator Curve

In this curve, $\epsilon(X)$ is the discriminator output for a phase difference between the received code and the replica code equivalent to X feet. The error monotonically increases in the range ± 50 feet, and is zero outside the range $\pm X^*$. The value of X^* depends upon the exact design of the delay lock loop. From the point of view of multipath-induced errors, it is conservative to consider designs with large values of X^*, and for this reason we will assume $X^* = 150$ feet. The deviation of the discriminator curve with X^* of 150 feet is shown in Reference 5. It should be noted that References 3 and 4 both assume delay lock loops with $X^* = 100$ feet, and hence end up with different results. (However, this is only one of the differences

between this treatment and the earlier ones, so it does not account for all the discrepancies in the numerical results.)

Hence, we will use for the discriminator curve,

when $X < -150$, $\epsilon(X) = 0$

$-150 < X < -50$, $\epsilon(X) = -\frac{X}{100} - \frac{3}{2}$

$-50 < X < +50$, $\epsilon(X) = \frac{X}{50}$

$+50 < X < +150$, $\epsilon(X) = -\frac{X}{100} + \frac{3}{2}$

$+150 < X$, $\epsilon(X) = 0$

5.2.2.2 Error Equation

We will assume that there are two signals in the receiver, the desired, direct signal, and one, significantly-strong multipath signal, which is delayed by an amount equivalent to a path distance difference of δ. Further, we will assume a reflection coefficient such that the ratio of peak voltages of the multipath signal to the desired signal is ρ. (That is, the voltage reflection coefficient is ρ. The reflection coefficient in power terms is ρ^2.)

Now, each of the two signals gives rise to an error output from the discriminator, and the delay lock loop drives the sum-error to zero. Let the final offset of the settled phase from the desired signal phase be equivalent to a distance ℓ. Then, the individual discriminator outputs would be

$\epsilon(\ell)$, from the desired signal

and $\rho \cos\left(\frac{2\pi\delta}{\lambda}\right) \epsilon(\ell - \delta)$, from the multipath signal

where λ = carrier wavelength

Since the loop drives the sum error to zero, the value of ℓ, which is the multipath-induced error in pseudorange, is given by,

$$\Sigma(\ell) = 0$$

where $$\Sigma(\ell) = \epsilon(\ell) + \rho \cos\left(\frac{2\pi\delta}{\lambda}\right) \epsilon(\ell - \delta)$$

Now, since $0 < \rho < 1$,

it is seen that $\Sigma(-50) < 0$

and $\Sigma(+50) > 0$.

Hence, there is a solution of $\Sigma(\ell) = 0$ in the range $\ell[-50, +50]$.

Thus, the equation for ℓ can be written as

$$\frac{\ell}{50} + \rho \cos\left(\frac{2\pi\delta}{\lambda}\right) \epsilon(\ell - \delta) = 0$$

For further consideration, we must consider the various ranges of the argument, $(\ell - \delta)$.

Suppose $\ell - \delta < -150$

Then $\epsilon(\ell - \delta) = 0$

from which, if $\delta > 150$, $\ell = 0$

Suppose $-150 < \ell - \delta < -50$

Then $$\epsilon(\ell - \delta) = -\frac{3}{2} - \frac{\ell - \delta}{100}$$

and $$\frac{\ell}{50} + \rho\left[-\frac{3}{2} - \left(\frac{\ell - \delta}{100}\right)\right] \cos\left(\frac{2\pi\delta}{\lambda}\right) = 0$$

or $$\ell = \frac{(150 - \delta)\,\rho \cos\left(\frac{2\pi\delta}{\lambda}\right)}{2 - \rho \cos\left(\frac{2\pi\delta}{\lambda}\right)}$$

Because of the transcendental nature of the equations, we cannot proceed further with an exact analysis. Instead, noting that ℓ oscillates with δ quite rapidly compared with the rate of change of its peaks with δ, we can make the approximation that

$$\ell_{max} = \frac{(150 - \delta)\rho}{2 - \rho}$$

and

$$\ell_{min} = -\frac{(150 - \delta)\rho}{2 + \rho}$$

Proceeding with algebra, it is easy to show that

$$\text{upper limit of } \ell = \frac{(150 - \delta)\rho}{2 - \rho}$$

whenever $50(1 + \rho) < \delta < 150$

and

$$\text{lower limit of } \ell = \frac{-(150 - \delta)\rho}{2 + \rho}$$

whenever $50(1 - \rho) < \delta < 150.$

Very similarly, it can be shown that

$$\text{upper limit of } \ell = \frac{\rho\delta}{1 + \rho}$$

whenever $0 < \delta < 50(1 + \rho)$

and

$$\text{lower limit of } \ell = -\left(\frac{\rho\delta}{1 - \rho}\right)$$

whenever $0 < \delta < 50(1 - \rho).$

Hence, ℓ oscillates within the quadrilateral-shaped envelope shown in Figure 5-4.

One feature of this result is that it implies that, if $\rho = 1$, the error can oscillate between zero and -50 feet for near zero values of δ. This is, of course, contrary to common sense which tells us that if there is no multipath (i.e., if $\delta = 0$), then there is no multipath error (i.e., $\ell = 0$). Evidently, the simplifying assumptions do not work for $\rho = 1$, and hence these results cannot be used with this value.

5.2.2.3 The Value of the Multipath Delay

We differentiate here between multipath caused by reflections from the aircraft structure, and multipath caused by ground reflections. Because of the very considerable coding gain, we need concern ourselves only with multipath signals whose delay is less than 150 feet. Hence, ground reflection multipath is only significant during an ILS approach.

In all other flight modes, we have to assume some kind of model for ρ and δ. We here assume that,

(i) there is only one multipath signal of significance, and all others are so much weaker they can be ignored, and

(ii) the value of δ for this one signal is uniformly distributed between zero and 150 feet.

Based on these assumptions, we can model the statistical behavior of the error. This is done in the next section (5.2.2.4).

During an ILS approach, the ground reflected multipath is most significant. The variation in δ will be of two kinds. Firstly, there will be a random, noise like variation of the order of a wavelength or two. Additionally, the mean value of the delay will systematically vary in accordance with the steadily changing geometry.

It is very simple to show that, for a satellite transmission, the delay is given by

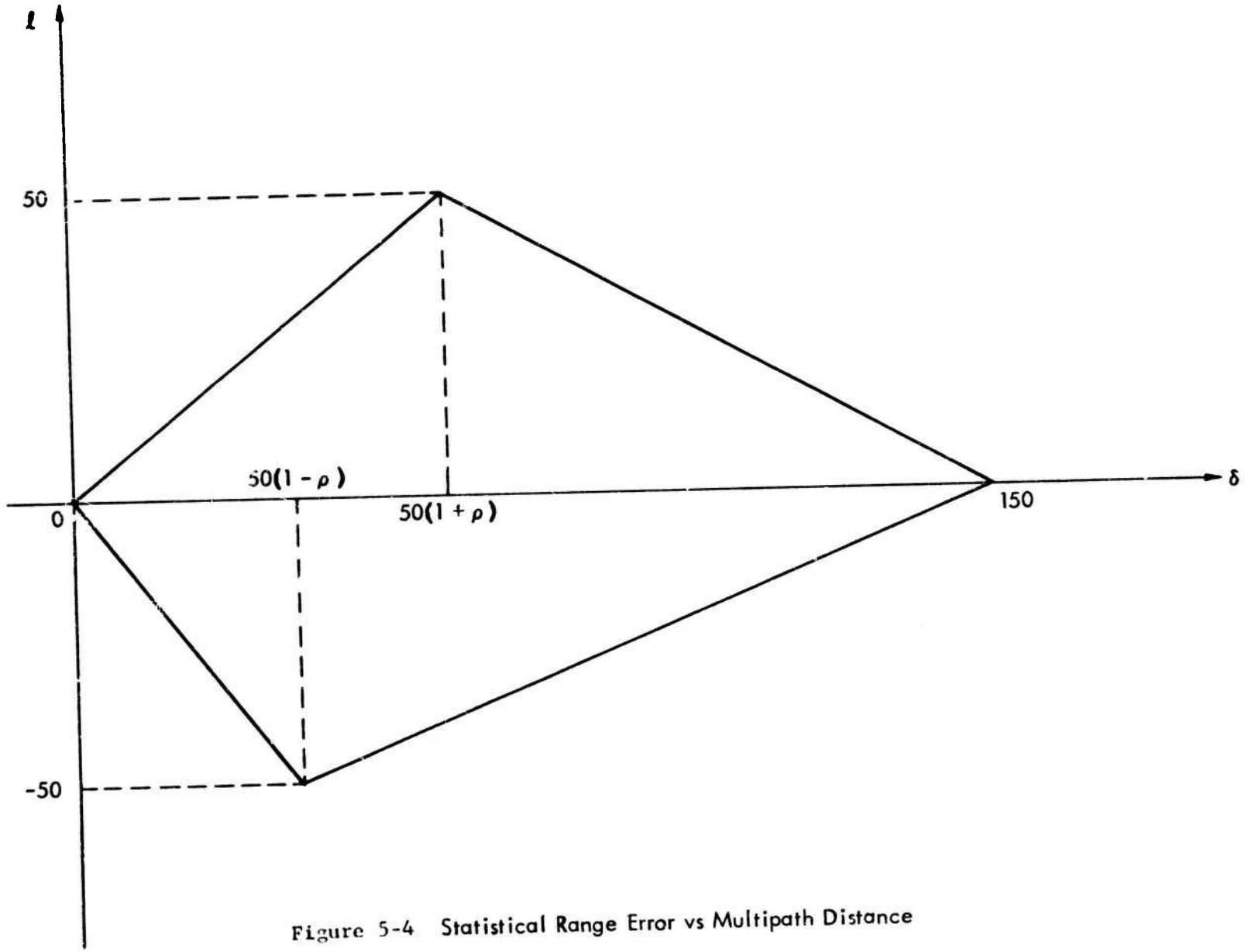


Figure 5-4 Statistical Range Error vs Multipath Distance

$$\delta = 2h \sin \theta$$

where h = aircraft altitude

and θ = elevation angle to the satellite.

Also, for the transmission from a ground beacon

$$\delta = \frac{2hb}{d}$$

where b = height of beacon antenna

and d = horizontal distance between aircraft and beacon.

For modeling errors in a simulation, we will assume that the slight variations of δ about the above systematic values has the result of averaging the error between the upper and lower bounds of Figure 5-4. Thus, the error to be used from ground reflected multipath is $\bar{\epsilon}$, where

$$\bar{\epsilon} = -\frac{\rho^2 \delta}{1-\rho^2} \quad , \text{for } 0 < \delta < 50(1-\rho)$$

$$\bar{\epsilon} = \frac{\rho(3\delta + 2\rho\delta - 150\rho - 150)}{2(1+\rho)(2+\rho)} \quad , \text{for } 50(1-\rho) < \delta < 50(1+\rho)$$

$$\bar{\epsilon} = \frac{\rho^2(\delta - 150)}{\rho^2 - 4} \quad , \text{for } 50(1+\rho) < \delta < 150$$

and $\bar{\epsilon} = 0$, for $\delta < 150$.

These relationships are shown in Figure 5-5.

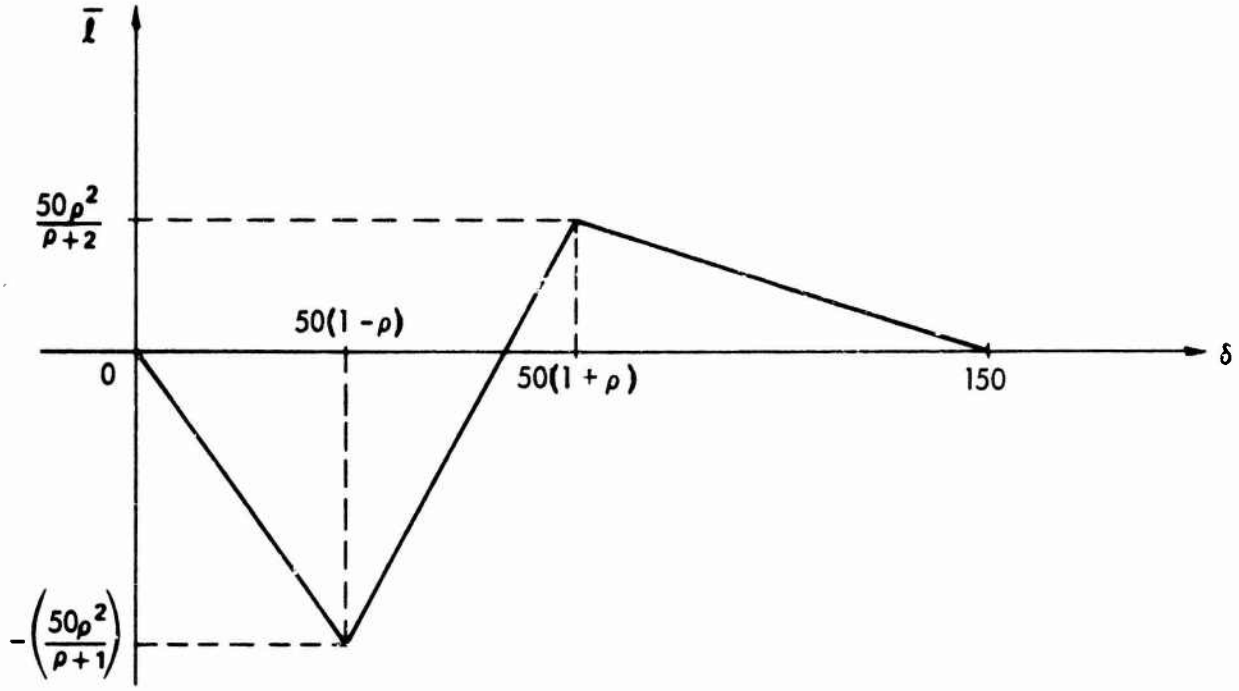


Figure 5-5 Deterministic Range Error Versus Multipath Distance

5.2.2.4 Statistical Description of the Error

With the simplifying assumption that δ is distributed uniformly between zero and 150 feet, even though we cannot analytically describe the fine structure of the variation of ℓ with δ within the envelope, we can see that it is reasonable to take, as the probability density function of ℓ, a triangular distribution, as shown in Figure 5-6. This is based on the fact that the length of an intercept of a line of constant ℓ by the envelope in Figure 3 is proportioned to $(50\rho - |\ell|)$.

It can then easily be shown that ℓ has a standard deviation given by

$$\sigma_\ell = \frac{50\rho}{\sqrt{6}} \quad \text{feet}$$

$$\sim 20\rho \quad \text{feet}$$

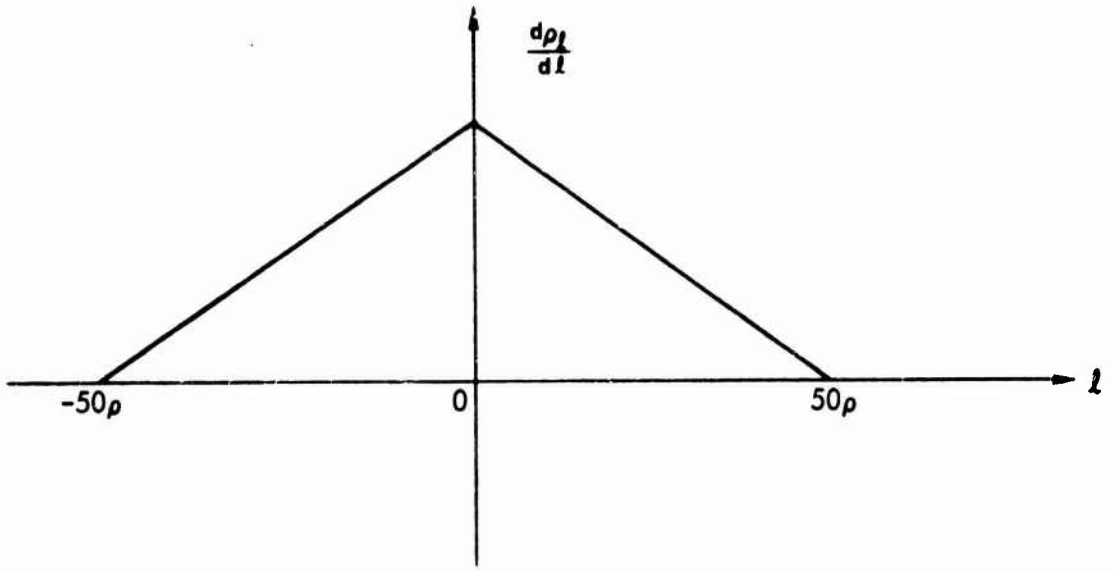


Figure 5-6 Probability Density of Statistical Range Error

For ground reflected multipath, we will arbitrarly assume that half the power is reflected, so that

$$\rho^2 = 0.5$$

$$\text{or } \rho = .707.$$

Then, with δ calculated from the geometry in accordance with the equations shown in Section 5.2.2.3, the value of the error is given by,

$\bar{l} = -\delta$, for $0 < \delta < 14.6$

$\bar{l} = -19.5 + .336\,\delta$, for $14.6 < \delta < 85.5$

$\bar{l} = .143(150 - \delta)$, for $85.5 < \delta < 150$

$\bar{l} = 0$, for $150 < \delta$

5.2.2.5 Numerical Results

For the case of multipath reflections from the aircraft structure (always present), we assume

$$\rho = 0.1$$

Thus, from the formula in Section 5.2.2.4.

$$\sigma_l = 2 \quad \text{feet}$$

Because the error is proportional to chip width and the C/A signal chip width is 10 times the width of the P signal chip, the equivalent multipath error for the C/A signal is 20 feet. Figure 5-7 shows the variation in average range rate error due to multipath for various P signal user conditions.

5.2.3 Range Rate Error

5.2.3.1 Phase Error In Reconstructed Carrier

We again assume the desired, direct signal and one multipath signal. It can be shown (see, for example, Reference 3), that the phase of the resultant signal is $\phi(t)$, where

$$\phi(t) = \psi_1(t) + \tan^{-1}\left(\frac{\rho \sin \psi}{1 + \rho \cos \psi}\right)$$

where $\psi_1(t)$ = phase of direct signal

$$\psi = \psi_2(t) - \psi_1(t)$$

and $\psi_2(t)$ = phase of multipath signal

Now, ψ_2 differs from ψ_1 in three ways. The first is because the multipath signal had to travel an additional path distance δ. The second

USER VEHICLE	ANTENNA HEIGHT	UNDESIRED TO DESIRED SIGNAL RATIO	RANGE ERROR
Aircraft	75'	-20 dB	2'
Monitor Station	10'	-10 dB	1.1'
All Other	0 - 75'	-3 dB (water)	See Below
		-10 dB (land)	

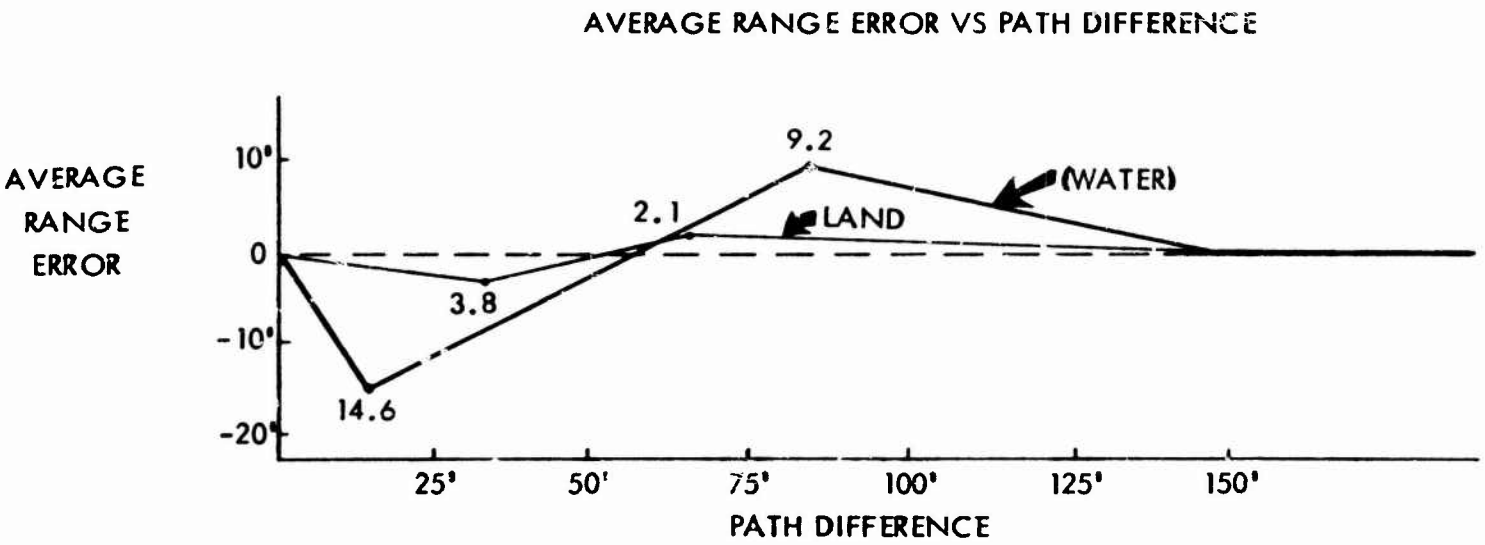


Figure 5-7 P Signal Multipath Errors in Delay Lock Loop

is that it may be 180 degrees out of phase from ψ_1 because it is one or more chips late. However, this second cause can be discounted here, because the receiver operates on reconstructed carrier, and if the signal is a chip or more out of phase, it will not be reconstructed. The third is because the signal phase is changed on reflection.

Hence, the multipath-induced phase error is due to change of phase along path distance δ and at reflection only. That is,

$$\Delta\phi = \phi(t) - \psi_1(t)$$

$$= \tan^{-1}\left(\frac{\rho \sin \psi}{1 + \rho \cos \psi}\right)$$

where $\psi = \frac{2\pi\delta}{\lambda} + \mu$

and μ = phase change at reflection.

5.2.3.2 Relationship Between Velocity Error and Phase Error

All receivers likely to be used in the GPS application measure velocity by determining the count increase in a known period of time, T, of a counter being clocked by a signal whose frequency is linearly related to the received signal frequency.

Hence, the count increase is linearly related to the integrated phase during the counter sampling period, T,

$$\int_{t=0}^{T} d[\phi(t)]$$

The algorithm for computing velocity from count increase will use the same linear relationship, so that, as for as errors are concerned, it can easily be shown tha

$$\Delta V = \frac{\lambda}{2\pi T} \Delta\left\{\int_{t=0}^{T} d[\psi(t)]\right\}$$

where ΔV = error in computed velocity.

Now, the error operator, Δ, can be transposed from without the integrating operator to within the differentiating operator. That is,

$$\Delta V = \frac{\lambda}{2\pi T}\int_{t=0}^{T} d[\Delta\phi(t)]$$

But $$\Delta\phi(t) = \tan^{-1}\left(\frac{\rho \sin \psi}{1 + \rho \cos \psi}\right)$$

Hence $$d[\Delta\phi(t)] = \dot{\psi}\left(\frac{\rho^2 + \rho \cos \psi}{1 + 2\rho \cos \psi + \rho^2}\right) dt$$

and $$\Delta V = \frac{\rho\lambda}{2\pi T}\int_{0}^{T} \dot{\psi}\left(\frac{\rho + \cos \psi}{1 + 2\rho \cos \psi + \rho^2}\right) dt$$

where $$\dot{\psi} = \frac{2\pi}{\lambda}\dot{\delta}$$

5.2.3.3 Approximation of the Velocity Error Expression

It was found that the analysis rapidly becomes unmanageable unless the expression for velocity error, as determined in Section 5.2.3.2, is simplified. To do this, we will assume that

$$\rho \ll 1.$$

Then $$\frac{\rho + \cos \psi}{1 + 2\rho \cos \psi + \rho^2} \sim \cos \psi - \rho \cos 2\psi$$

Thus $\Delta V = \frac{\rho\lambda}{2\pi T}\int_{t=0}^{T} (\cos \psi - \rho \cos 2\psi)\, d\psi$

$$= \frac{\rho\lambda}{2\pi T} \left[\sin \psi - \frac{\rho}{2} \sin 2\psi\right]_{t=0}^{T}$$

Now, when $t = 0$, $\psi = \psi_o$, say.

Then, when $t = T$, $\psi = \psi_o + \frac{2\pi T}{\lambda}\ \dot{\delta}$

Where $\dot{\delta}$ = rate of change of multipath differential distance.

Hence,

$$\Delta V = \frac{\rho\lambda}{2\pi T} \left[\sin \left(\psi_o + \frac{2\pi T}{\lambda}\ \dot{\delta}\right) - \sin (\psi_o) - \frac{\rho}{2} \sin \left(2\psi_o + \frac{4\pi T}{\lambda}\ \dot{\delta}\right) + \frac{\rho}{2} \sin (2\psi_o)\right]$$

5.2.3.4 Statistical Description of the Error

Since $\dot{\psi} = \frac{2\pi\dot{\delta}}{\lambda}$, the statistical distribution of ψ depends directly on that of $\dot{\delta}$. We assume that $\dot{\delta}$ is normally distributed about zero, with a standard deviation of $\sigma(\dot{\delta})$. Then, it can be shown that

$$\sigma_V^2 = \left(\frac{\rho\lambda}{2\pi T}\right)^2 \left[1 - e^{-\frac{1}{2}\sigma_x^2} + \frac{\rho^2}{4}\left\{1 - e^{-2\sigma_x^2}\right\}\right]$$

Where $\sigma_x = \frac{2\pi T}{\lambda}\, \sigma(\dot{\delta})$

Even with quite small values of $\sigma(\dot{\delta})$, this expression rapidly approaches its asymptotic value. Hence, we will say,

$$\sigma_v^2 = \left(\frac{\rho\lambda}{2\pi T}\right)^2 \left(1 + \frac{\rho^2}{4}\right)$$

$$= \left(\frac{0.1\rho}{T}\right)^2 \left(1 + \frac{\rho^2}{4}\right)$$

With $\rho = 0.1$,

$$\sigma_v = \frac{.01}{T} \quad \text{fps.}$$

5.2.3.5 Range Rate Error During an ILS Approach

The rapid variation of δ which causes the range rate error occurs principally from aircraft structure reflections, and not from the more deterministic ground reflections. Accordingly, there is no difference between range rate errors in flight and those during approach, as was exhibited for range error.

REFERENCES

1. "Final Report of the System 621B Timeshared User Equipment Study," SAMSO-TR-72-282, Volume II, Appendix 7.

2. "System 621B Receiver Test/Demonstration, Final Technical Report, Volume I (U)," TRW, April 1970. Secret. TRW #12397-6032-R3-01, SAMSO #TR-70-126-Vol. I.

3. "System 621B User Equipment Definition and Equipment Program, Phase I, Volume IV," Grumman/Hazeltine, Dec 1970. GAC #621B-FR1, SAMSO #TR-70-440.

4. "C-Band Signal and Reception Techniques Study, Final Report," Philco-Ford, Jan 1971, Confidential. SAMSO #TR-70-434.

5. "Effects of Multipath on Coherent and Noncoherent PRN Ranging Receiver," Aerospace, 73 May 15, L.L Hagerman, TOR-0073(3020-03)-3.

6. "Simulation of the INI Receiver Control Loop," TRW Memo 7121.17-20, 15 May 1972, J.C. Wilcox.

6.0 RECEIVER NOISE AND QUANTIZATION

This section of the error budget analysis report addresses the errors in the range and range rate caused by noise and quantization. Table 6-1 summarizes the anticipated one sigma error for the monitor station and various types of users based upon the assumptions stated later in this section.

TABLE 6-1

1σ Ranging Error Due to Noise and Quantization (Ft.)

User Type \ User Acceler. (g's)		0	1	2	4	7
Aircraft	P-Code L_1	1.25	1.35	1.45	1.6	1.9
	P-Code L_2	1.8	1.9	2.0	2.1	2.4
	C-Code	8.0	8.2	8.4	8.7	9.0
Ships and Manpack	P-Code L_1	1.25	1.35	N/A	N/A	N/A
	P-Code L_2	1.8	1.9	N/A	N/A	N/A
	C-Code L_1	8.0	8.2	N/A	N/A	N/A

Table 6-2 gives the range rate errors for the various types of users.

TABLE 6-2

1σ Range Rate Error Due to Noise and Quantization (Ft./sec.)

User Type \ User Acceler. (g's)		0	1	2	4	7
Aircraft	P-Code L_1	.036	.51	1.0	1.9	3.3
	P-Code L_2	.05	.52	1.0	1.9	3.3
	C-Code L_1	.026	.50	1.0	1.9	3.3
Ships and Manpack	P-Code L_1	.036	.51	N/A	N/A	N/A
	P-Code L_2	.05	.52	N/A	N/A	N/A
	C-Code L_1	.026	.50	N/A	N/A	N/A

The relative accuracy is approximately the same as the absolute accuracy as given above. Even though the major source of range rate errors under high acceleration conditions is a bias in the loop, these conditions can be expected to change significantly within a minute. Thus, this error takes on random characteristics and no significant improvement can be expected from the standpoint of relative accuracy.

Section 6.1 computes the receiver range and range rate errors due to receiver noise, quantization and instrumentation, and receiver acceration as a function of C/N_o at the input to the receiver. Section 6.2 establishes a reference baseline link for purposes of error analysis and presents satellite elevation angle statistics. Section 6.3 presents antenna characteristics for each type of receiver and computes the error probability distributions that can be expected for each.

6.1 Pseudo-Range and Range-Rate Tracking Errors

In this section the user receiver range and range rate measurement errors due to receiver noise, quantization, instrumentation, and due to user accelerations are discussed. The receiver noise, quantization, and instrumentation errors are random and uncorrelated from measurement to measurement. The acceleration errors are bias errors.

6.1.1 Range Errors

The total random range error is the rms sum of the noise, quantization, and instrumentation errors, as follows:

$$\sigma_T(R) = \left[\sigma_n^2(R) + \sigma_q^2(R) + \sigma_I^2(R)\right]^{\frac{1}{2}} \qquad (1)$$

The noise error in turn has two sources: phase jitter in the carrier (Costas) tracking loop and phase jitter in the delay lock loop (DLL). Phase jitter in the Costas tracking loop is equal to[1]

$$\sigma_{\phi 1} = \left[\frac{B_1}{(C/N_o)_{C1}}\left(1 + \frac{W}{2(C/N_o)_{C1}}\right)\right]^{\frac{1}{2}} \qquad (2)$$

where $\sigma_{\phi 1}$ = rms phase jitter in carrier loop

B_1 = one-sided loop bandwidth

W = noise bandwidth of RC data filter in the loop

(C/N_o) C_1 = signal-to-noise density ratio in the carrierloop

Phase jitter in the DLL is equal to [2]

$$\sigma_{\phi 2} = 2\pi \left[\frac{B_2}{(C/N_o)_{DLL}} \left(1 + \frac{2B_{IF}}{(C/N_o)_{DLL}} \right) \right]^{\frac{1}{2}} \quad (3)$$

where $\sigma_{\phi 2}$ = rms phase jitter in DLL

B_2 = one-sided loop bandwidth

B_{IF} = predetection IF bandwidth

$(C/N_o)_{DLL}$ = signal-to-noise density ratio in DLL

Range error is related to the phase jitter as follows:

$$\sigma(R) = \frac{\lambda_c}{2\pi} \sigma_\phi$$

where λ_c = code wavelength

The total rms range error due to phase jitter in the two loops equals

$$\sigma_n(R) \quad \frac{\lambda_c}{2\pi} \left[\sigma_{\phi 2}^2 + \left(\frac{\sigma_{\phi 1}}{N} \right)^2 + 2 \frac{\sigma_{\phi 1}}{N} \sigma_{\phi 2} \rho_{12}(0) \right] \quad (4)$$

where N = carrier-to-code frequency ratio

$\rho_{12}(0)$ = cross-correlation between Costas loop and DLL noise

The function $\rho_{12}(0)$ can be neglected since it is close to zero[3]. Therefore, Equation 4 reduces to

$$\sigma_n(R) = \frac{\lambda_c}{2\pi}\left[\sigma_{\phi 2}^2 + \left(\frac{\sigma_{\phi 1}}{N}\right)^2\right]^{\frac{1}{2}} \tag{5}$$

The rms quantization errors for the coarse and fine range measurements are as follows:

$$\sigma_{qc}(R) = \frac{\lambda_c}{\sqrt{12}} \tag{6}$$

$$\sigma_{qf}(R) = \frac{\lambda_c}{\sqrt{12}\,N} \tag{7}$$

When the delay lock loop is locked and tracking, the tracking error oscillates in a random manner about the true range with a rms error of one carrier wavelength in range. This is because the loop is always adding or deleting one count of the N divider between the carrier and code loops. This results in an instrumentation error equal to the carrier wavelength. Thus

$$\sigma_I(R) = \lambda_c/N \tag{8}$$

A range bias error occurs from receiver accelerations due to an additional phase tracking error in the receiver carrier loop. For an acceleration step, the peak phase error in the carrier loop is

$$\Delta\phi = \frac{1.845}{B_1^2}\left(\frac{A}{\lambda_s}\right) \tag{9}$$

where A = acceleration (fps^2)

λ_s = carrier wavelength = λ_c/N

The resulting error in the code loop is 1/N as large, and so the resulting bias error is

$$\Delta R = \frac{\lambda_c}{2\pi}\frac{\Delta\phi}{N} \tag{10}$$

Combining Equations 9 and 10

$$\Delta R = \frac{1.845}{2\pi B_1^2} A \tag{11}$$

6.1.2 Range Rate Errors

The total random range rate error is the rms sum of the noise and quantization errors as follows:

$$\sigma_T(\dot{R}) = \left[\sigma_n^2(\dot{R}) + \sigma_q^2(\dot{R})\right]^{\frac{1}{2}} \tag{12}$$

The range rate noise error results from the phase noise in the carrier tracking loop. The receiver range rate counter averages the carrier frequency for T sec. Thus, the range rate measurement equals

$$\dot{R} = \frac{\lambda_s}{2\pi T}\left[\phi(t+T) - \phi(t)\right] \tag{13}$$

where $\phi(t)$ = carrier loop VCO phase

T = averaging time for doppler count

The rms range rate error is then

$$\sigma_n(\dot{R}) = \frac{\lambda_s\sqrt{1-\rho_{11}(T)}}{2\pi T}\left(\sqrt{2}\,\sigma_{\phi 1}\right) \tag{14}$$

where $\rho_{11}(T)$ = autocorrelation of carrier loop noise

$\sigma_{\phi 1}$ = carrier loop phase noise [Equation (2)]

The autocorrelation function $\rho_{11}(T)$ is negligibly small[3], and can be assumed to be zero. Thus

$$\sigma_n(\dot{R}) = \frac{\lambda_s}{2\pi T}\left[\frac{2B_1}{(C/N_o)_{L1}}\left(1+\frac{W}{2(C/N_o)_{L1}}\right)\right]^{\frac{1}{2}} \tag{15}$$

The rms quantization errors for the coarse and fine range rate measurements are as follows:

$$\sigma_{qc}(\dot{R}) = \frac{\lambda_s}{T\sqrt{12}} \tag{16}$$

$$\sigma_{qf}(\dot{R}) = \frac{\lambda_s}{2NT\sqrt{12}} \tag{17}$$

The range rate bias error due to receiver accelerations is equal to

$$\Delta\dot{R} = \frac{\lambda_s}{2\pi}\left(\frac{\Delta\phi}{T}\right) \tag{18}$$

Where $\Delta\phi$ is the phase tracking error in the carrier loop as given by Equation 9.

6.1.3 Performance Summary

Table 6-3 gives a tabulation of the values of the various user receiver parameters. Table 6-4 is a tabulation of the range and range-rate bias errors due to the relative user acceleration for the parameter values of Table 6-3.

Both the P- and C/A-signal user receivers are expected to be implemented using a τ-dither configuration. The τ-dither implementation results in a reduction in signal-to-noise density into the two tracking loops (carrier and code). The input signal-to-noise density ratio into the two tracking loops is related to the effective received carrier-to-noise density ratio* as follows[3]:

$$(C/N_o)_C = (C/N_o)_e - 1.3 \text{ dB} \tag{19}$$

$$(C/N_o)_{DLL} = (C/N_o)_e - 4.8 \text{ dB} \tag{20}$$

The total random range errors, $\sigma_T(R)$ as a function of effective received carrier-to-noise density are presented in Figures 6-1 and 6-2 for the P-signal and C/A-signal, respectively. The total random range errors plus bias errors due to user accelerations are also presented. Figure 6-3 presents total random range rate error as a function of effective received carrier-to-noise density. The total random range rate error plus bias errors due to user accelerations are also presented.

* Effective carrier-to-noise density ratio includes jammer noise and is related to the unjammed carrier-to-noise density ratio as follows (for the P-signal):

$$(C/N_o)_e = \frac{(C/N_o)_{signal}}{1 + \left(\frac{10^{-7}}{1.023}\right)(J/S)(C/N_o)_{signal}}$$

TABLE 6-3
USER RECEIVER PARAMETERS

Parameter	Value	
	P-Code	C/A-Code
B_1	10 Hz	10 Hz
B_2	0.5 Hz	0.5 Hz
W	50 Hz	200 Hz
B_{IF}	100 Hz	400 Hz
λ_c	96.25 ft	962.5 ft
N	154	1540
$\lambda_s = \lambda_c / N$	.625 ft	.625 ft
T	0.2	0.2

TABLE 6-4
RANGE AND RANGE RATE ACCELERATION BIAS ERRORS

P- and C/A-Code Receivers

Quantity	Equation	Accelerations (g's)			
		1	2	4	7
ΔR (ft)	(11)	0.095	0.19	0.38	0.66
ΔR (fps)	(18)	0.47	0.94	1.89	3.3

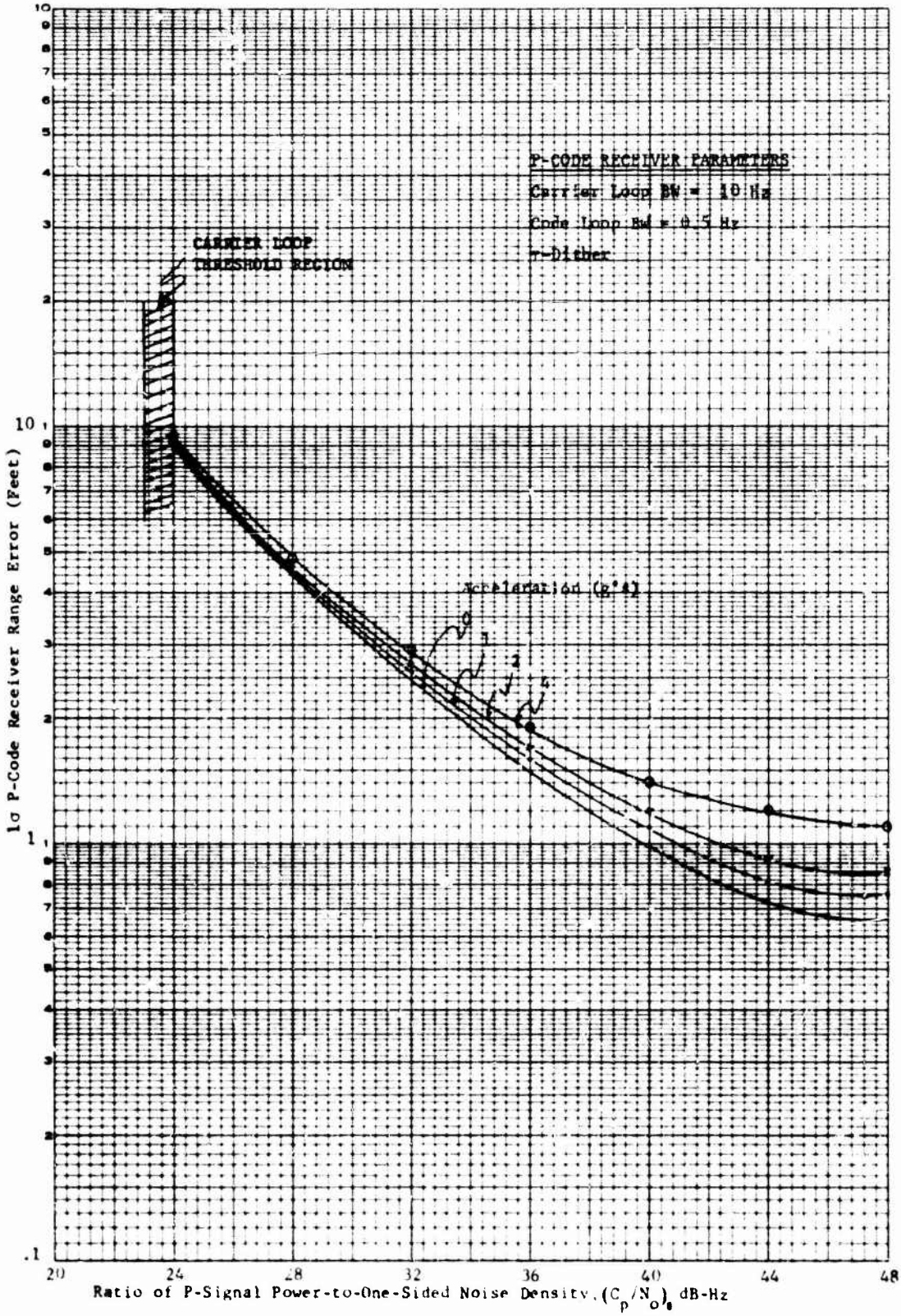


Figure 6-1 P-Signal Receiver Range Errors

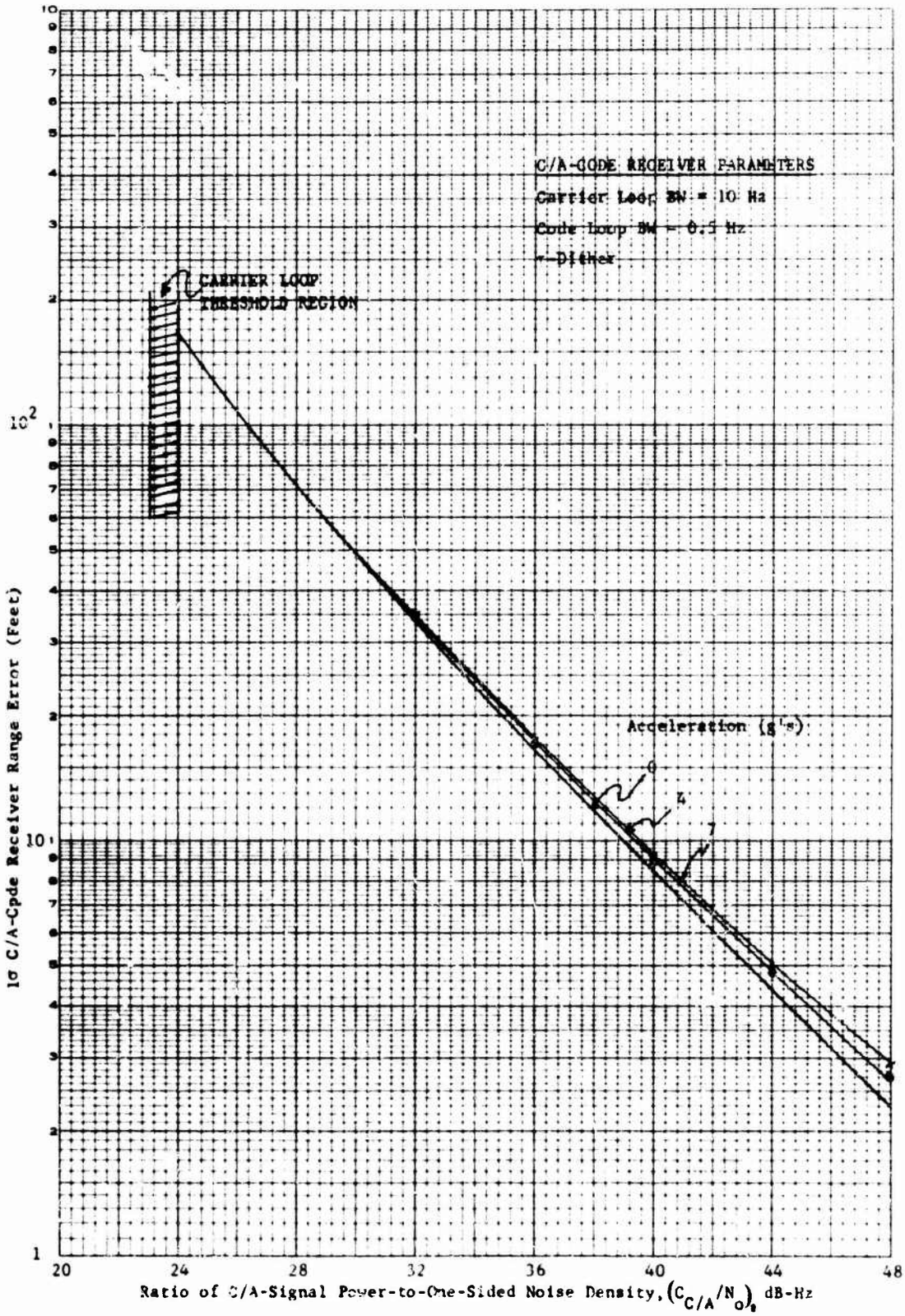


Figure 6-2 C/A Signal Receiver Range Errors

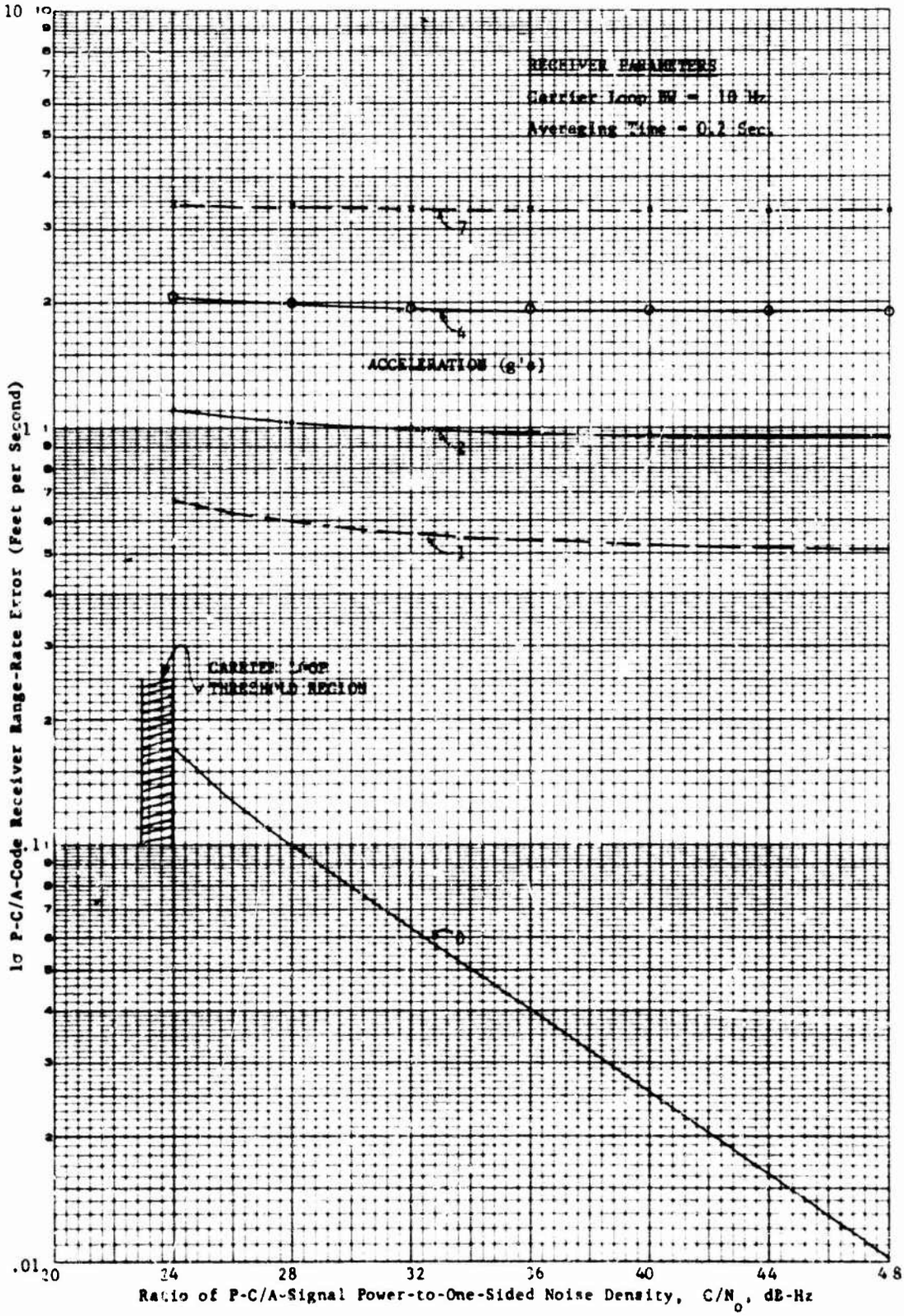


Figure 6-3 P or C/A Signal Receiver Range Rate Errors

REFERENCES

(1) TRW Memo DNSDP-JKH-154, "Performance and Stability Considerations of the DNSS Receiver I-Q Loop", By J. K. Holmes, 3 December 1973

(2) Gill, W. J., "A Comparison of Binary Delay-Lock Tracking Loop Implementations", IEEE Transactions AES Vol. AES2, No. 4, July 1966

(3) TRW Systems Group, "System 621B Timeshared User Equipment Study", Vol II, Appendix 11, SAMSO TR 72-202, Vol II, January 1973

6.2 Reference Link C/N_o

This section establishes a reference C/N_o at a 5 degree elevation angle for a C.P. isotropic antenna and shows how C/N_o varies with elevation angle. Probability distributions of satellite elevation angles are given for Phases I, II, and III for the baseline orbits for user and monitor receivers.

The reference values are based upon the GPS System Specification, SS-GPS-101A, Appendix II, paragraph 20.8, Receive Signal Level which states:

> 20.8 Receive signal level. The receive signal levels are specified in Table 20.8.1. The received signal levels shall exceed the levels given in the table whenever the spacecraft is 5° above the horizon with respect to a ground-based user at sea level. The signals shall be Right-Hand Circularly Polarized (RHCP). The receive signal levels in the table shall be achieved at the output of a serially connected 0 db RHCP antenna. The user equipment shall be designed for the receive signal levels shown in Table 20.8.1.

"TABLE 20.8.1"

Receive Signal Level

Channel \ Signal	P	C/A
L_1	-163 dBw	-160 dBw
L_2	-166 dBw	N/A

The effective receiver system noise temperature/noise power density referred to the receiver preamplifier input is derived as follows.[1]

$$T_S = LT_A + (1 - L)T_L + (F - 1)T_o$$

where

T_S = receiving system noise temperature

T_A = 130^oK - a conservative estimate of antenna temperature of hemispheric beam looking upward with some radiation from earth coming in via sidelobes

T_L = $290^{\circ}K$ - temperature of antenna coupling circuit and coax line connecting antenna feed and receiver input

L = -1.0 dB (0.794) nominal

Based on 25 ft of RG-type coax cable between antenna feed and receiver input or 100 ft of rigid copper coax. This value is highly user installation dependent. If cable runs become too long, receiver preamp can be located at antenna feed to minimize cable losses.

F = 4.5 dB (2.82) nominal

Estimated overall receiver noise figure based on available L-band transistor devices and including preselector filter and mismatch losses.

T_o = $290^{\circ}K$ - reference temperature.

Using the above values, the receiver system noise temperature is

T_S = $690.7^{\circ}K$ nominal

Table 6-5 summarizes the signal and noise level values into the receiver at the reference 5 degree elevation angle when a circularly polarized isotropic antenna is used.

Reference (1) also calculates the effect of elevation angle on signal strength. The increased path length loss is shown on Figure 6-4 (a) and the relative satellite antenna gain (off-beam-center loss) is shown on Figure 6-4 (b). The latter includes a pointing error of 1 degree.

Combining the data from Table 6-5 with Figure 6-4 yields Figure 6.5. This shows the C/N_o at the receiver input as a function of elevation angle when the user has a circularly polarized isotropic antenna with no VSWR or polarization mismatch losses.

TABLE 6-5

REFERENCE C/No FOR ISOTROPIC ANTENNA AT 5° ELEV. ANGLE

RANGING CODE		P-CODE		C-CODE
FREQUENCY		L_1	L_2	L_1
Signal Level (GPS Spec.) 5° Elev. Angle, CP Isotropic Ant.	dBw	-163.0	-166.0	-160.0
Line Loss	dB	- 1.0	- 1.0	- 1.0
Net Signal Into Rcvr	dBw	-164.0	-167.0	-161.0
Boltzmann's Constant	dBw /Hz/K	-228.6	-228.6	-228.6
System Noise Temp (591°K)	dB-K	28.4	28.4	28.4
Net Noise Density	dBw/Hz	-200.2	-200.2	-200.2
Net C/No at 5° Elev Angle and CP Isotrople Ant.	dB·Hz	36.2	33.2	39.2

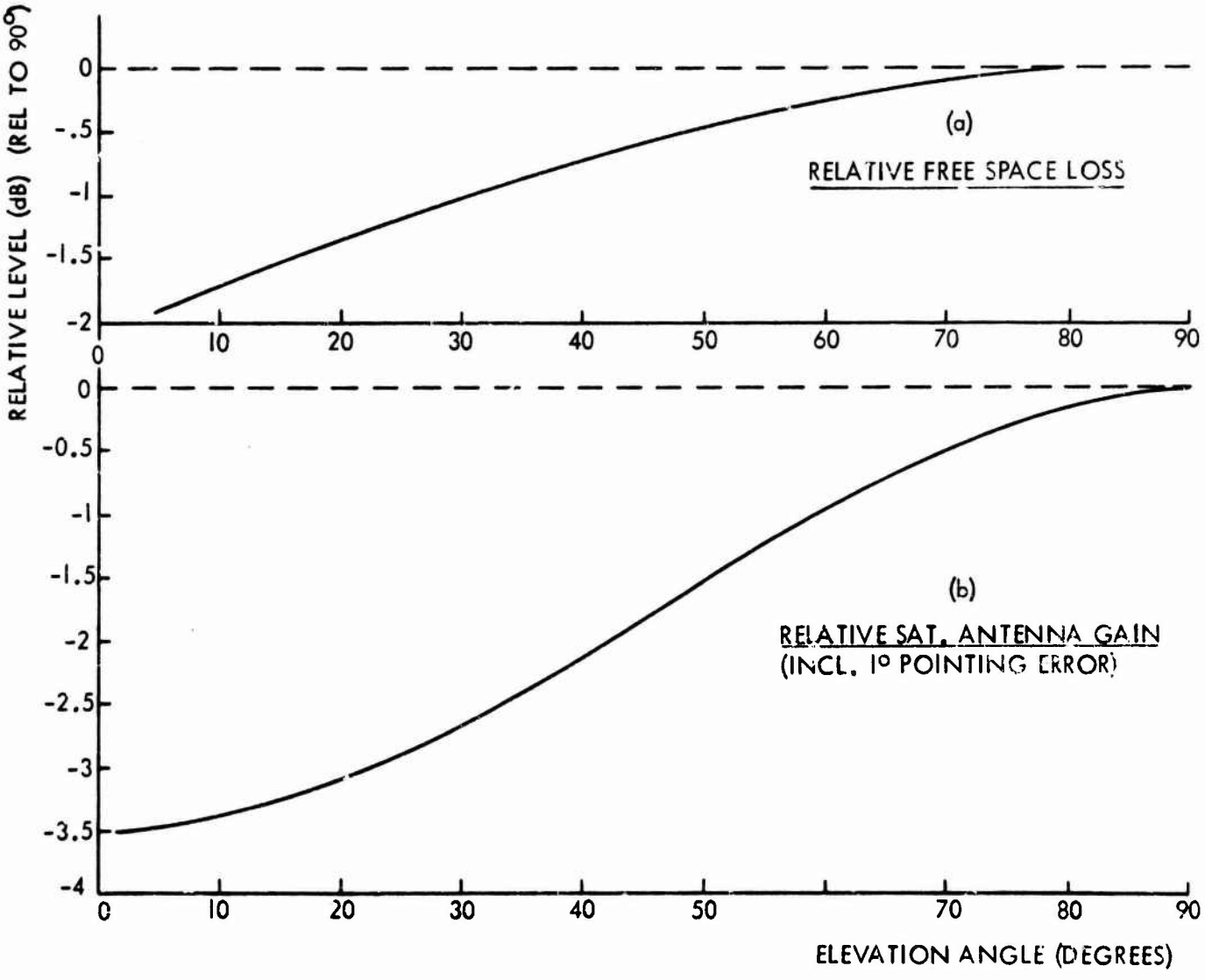
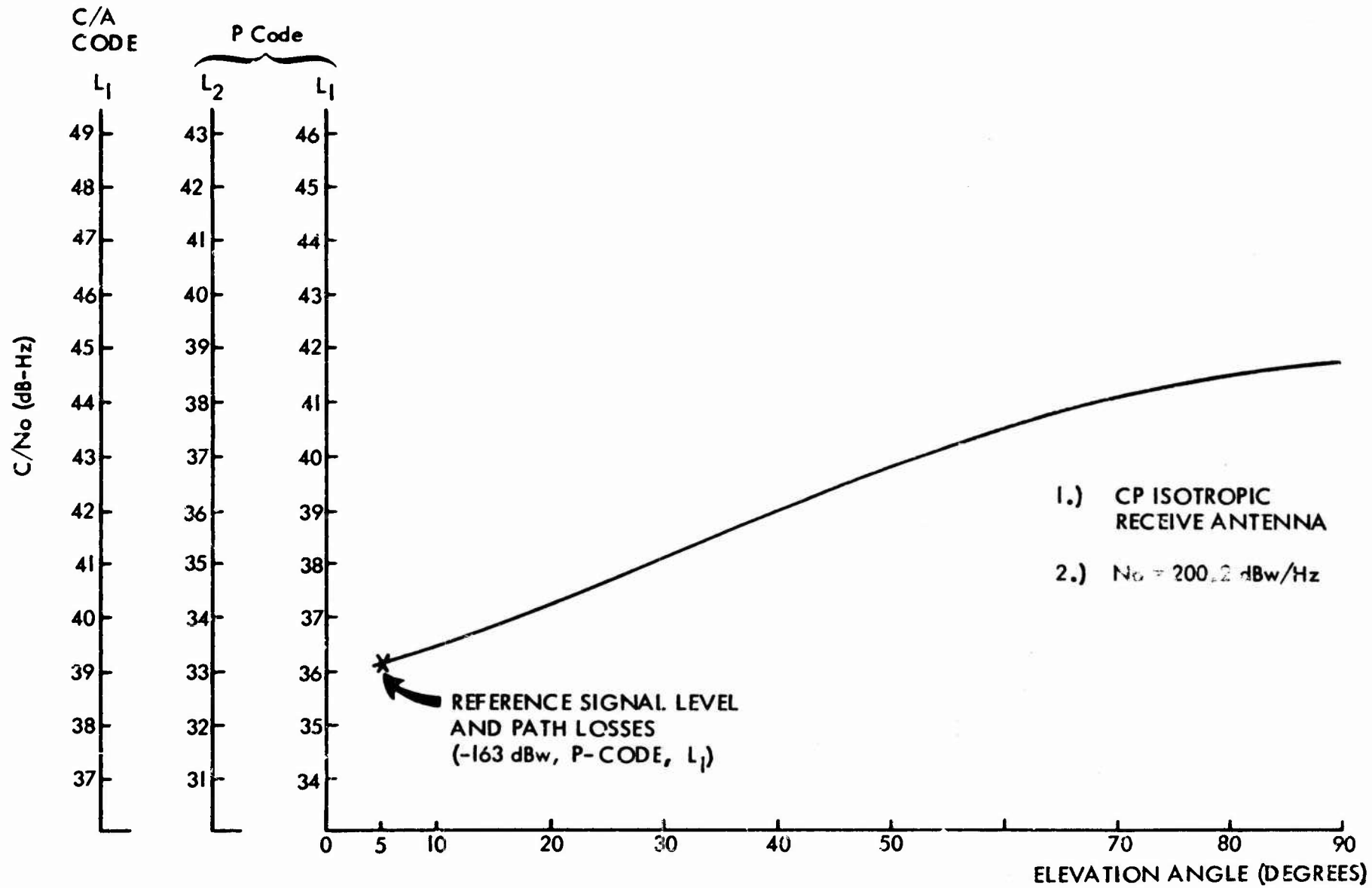


Figure 6-4 ELEVATION ANGLE EFFECTS

Figure 6-5 C/No vs ELEVATION ANGLES (ISOTROPIC ANTENNA)



6.3 User Accuracy

The probability distribution of the satellite elevation angle for users during Phases I, II, and II are shown on Figure 6-6. This assumes the baseline orbits as given in Reference (2). The Phase I distribution assumes the user is located at White Sands Missile Range (WSMR) and covers the 2.4 hour test period that all four satellites are in view simultaneously. The Phase IIA distribution is also based on a user at WSMR and is for the 7.75 hour test period. The Phase IIB and III user distributions are based on 24 hour per day operation for users at 33° latitude. It can be seen that during Phase I, the average elevation angle is 27°. During Phase IIA this increases to 34° and during Phases IIB and III it is 29°. Discussions with manufacturers of the user equipment and testing of L-band antennas developed for aircraft communications indicates that the power transfer (including polarization and impedance mismatch) of realizable antennas is as good as a circularly polarized isotropic antenna at these elevation angles. [3, 4, 5, 6] From Figure 6-5 the C/N_o provided by a 0 dB antenna 25° is 37.5 dB-Hz for the L_1 P-Code signal. Table 6-6 lists the one sigma range and range rate errors associated with this configuration derived from Figures 6-1, 6-2 and 6-3 of section 2.1.

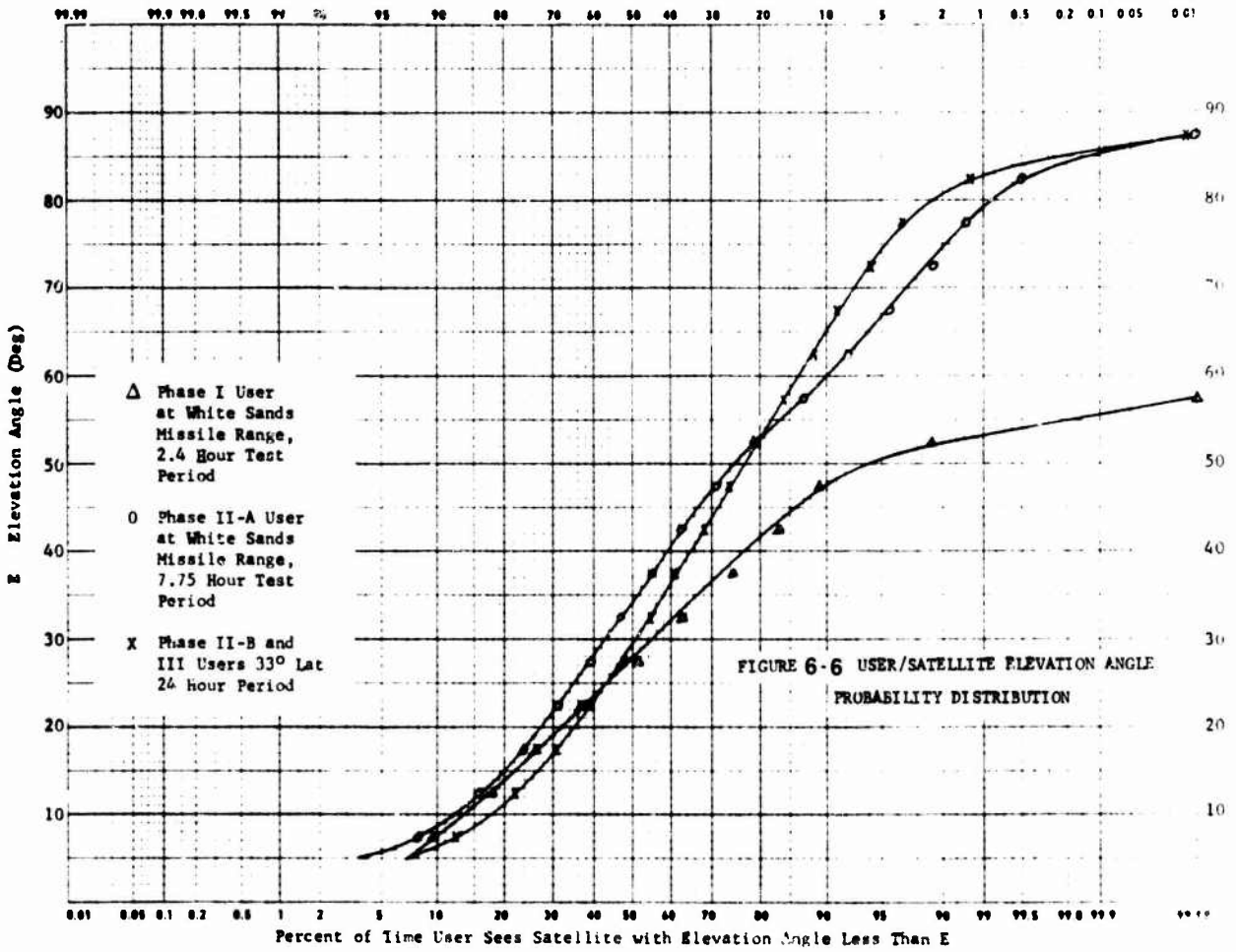


FIGURE 6-6 User/Satellite Elevation Angle Probability Distribution

TABLE 6-6
1σ RECEIVER NOISE AND QUANTIZATION ERRORS
(0 dB ANTENNA, 25° ELEVATION ANGLE)

PARAMETER MEASURED	RANGE			RANGE RATE		
CODE	P		C/A	P		C/A
FREQUENCY	L_1	L_2	L_1	L_1	L_2	L_1
C/No (dB-Hz)	37.5	34.5	40.5	37.5	34.5	40.5
USER EQUIP. ACCELERATION (G's)						
0	1.25	1.8	8.0	.036	.05	.026
1	1.35	1.9	8.2	.51	.52	.50
2	1.45	2.0	8.4	1.0	1.0	1.0
4	1.60	2.1	8.7	1.9	1.9	1.9
7	1.9	2.4	9.0	3.3	3.3	3.3

REFERENCES

1. Section 2, this report

2. Philco-Ford "RF Link Analysis for Global Positioning System," GPS-TM003A, Baseline Orbits," GPS-TM-005, 28 January 1974.

3. TRW, "Navigation/Traffic Control Satellite Mission Study," TRW No. 09778-6009-RO-00, for NASA ERC, June 1969.

4. Boeing, "Experimental L-Band SST Satellite Communication/Surveillance Terminal Study, "D6-60105-3, for NASA ERC, November 1968.

5. C. A. Lindberg, "A Shallow Cavity UHF Crossed-Slot Antenna," MIT Lincoln Labs, Technical Report 446, March 1968.

6. Olsson, et al, "L-Band Orthogonal-Mode Crossed-Slot Antenna," The Boeing Company Report No. DOT-TSC-NASA-72-2, August 1972.

7.0 GPS CLOCK ERROR

GPS clock errors contribute to the User Equivalent Range Error (UERE) in three ways:

a. Satellite Clock Estimation Error. During the periods the satellites are in view, the Control Segment will monitor the satellite clocks and will make an estimate of what the clock conditions were during this past period. This estimate will contain errors which will be correlated with errors in the satellite position, since, to a certain extent, it is determined from the same tracking data used for ephemeris determination. This error is minimized by using ELM pseudorange data to measure clock condition and HAW, MAINE and VAFB range difference data to determine satellite ephemerides. This clock estimation error enters UERE as part of the satellite state vector in the navigation data and is properly treated as part of the Ephemeris Error in the Error Budget.

b. Satellite Clock Prediction Error. The satellite clocks will continue to change in phase and frequency after the last clock measurement used to make the clock estimation. This change will contribute directly to UERE and will be uncorrelated with any errors in satellite orbit predictions to the extent that the clock drift is random and unpredictable. This error is designated as GPS Clock Error in the Error Budget and is discussed in greater detail in this section.

c. User Clock Errors. Although the user employs a relatively unstable clock, his measurement of range differences over a short time span (approximately 100 milliseconds) will minimize this error contribution. The short term stability characteristics of crystal oscillators should cause this error to be less than one foot. The user's range measurements to the fourth satellite can also be used to minimize this error.

It has been found that clock models employing second order polynomials will accurately represent the deterministic satellite clock variations. The use of a higher order polynomial will not increase the accuracy of the clock as only the first three terms are categorically deterministic and non-random.[1] If the clock model is to fit only a one hour period, a linear representation, is satisfactory. Thus the clock model to be employed by the user to correct the satellite variations is:

$$T_C = T_B + T_F \cdot t$$

where

T_C = Time Correction

T_B = Clock Time Bias

T_F = Clock Frequency Error at t_o

The coefficients T_B, and T_F transmitted to the user via the navigation data message are the optimum estimate of the residuals of these parameters after the clock data has been processed.

Various recursive data filtering approaches were studied and the Carlson triangular algorithm has been selected as the baseline because of its numerical stability. A detailed analysis of Least Squares fitting to frequency was performed and the accuracy of this prediction technique was determined. The accuracy of the Carlson formulation is expected to approximate that of the Least Squares/Frequency approach.

A TRACE simulation which determined the optimum coefficients by minimizing the ranging error during the observation span was also conducted. This is

---

[1] D.W. Allan, et al, "Performance, Modeling and Simulation of Some Cesium Beam Clocks," 27th Annual Frequency Symposium, Ft. Monmouth, New Jersey

equivalent to a Least Squares fit to Phase. The detailed analysis of clock prediction is contained in Part II, Vol. C, Report A3, Clock Model Trade Study for GPS.

7.1 Clock Noise Characteristics

A major problem in determining the bias, frequency offset and drift rate is the noise in the clock. The types of predominant noise observed in clocks are (a) white noise FM, (b) flicker noise FM, and (c) random walk FM. White noise PM and flicker noise PM for averaging times > 1 second are not a significant contributors. The spectral density of the noise can be described by the following expression:

$$S(f) = h_o f^o + h_{-1} f^{-1} + h_{-2} f^{-2} \qquad f \leq 1 \text{ Hz}$$

where: h_o = white noise FM contribution

h_{-1} = flicker noise FM contribution

h_{-2} = random walk FM contribution

The Allan variance, $\sigma^2\left(\frac{\Delta f}{f}\right)$, describes the time domain instability of a clock and is related to the h_o, h_{-1} and h_{-2} parameters as follows:

$$\sigma^2\left(\frac{\Delta f}{f}\right) = \frac{h_o}{2\tau} + 2h_{-1} \ln 2 + \frac{(2\pi)^2}{6} h_{-2} \tau$$

where τ is the averaging time.

Figure 7-1 shows the relationship between the Allan Variance and averaging time for typical Rubidium, Cesium and HP super tubes.

The white noise and flicker noise values and the value of τ_1, the point at which flicker noise FM becomes predominant, will affect the selection of the optimum filter and determine the prediction accuracy of the selected algorithm.

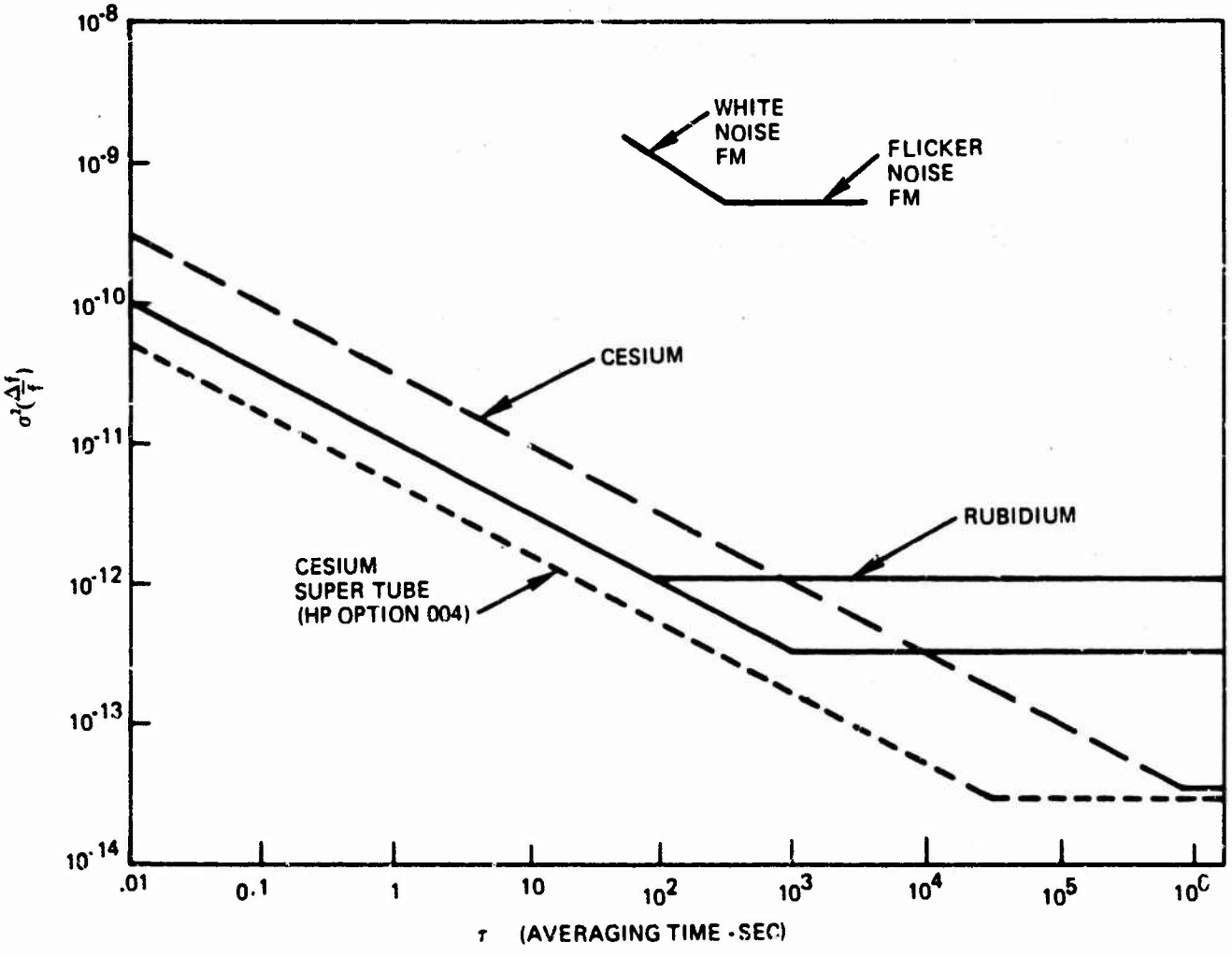


Figure 7-1 Allan Variance Vs Averaging Time

7.2 Clock Prediction Algorithms

The objective of the clock processing algorithm is to determine with the best possible accuracy, the T_B and T_F parameters of each satellite's clock for the next 24 hours and to accomplish this in the presence of significant clock noise. (The aging term T_{FD} will be supplied by the ephemeris support system)

7.2.1 Least Squares/Exponential

The least squares/exponential approach is a recursive exponential filter developed by Barnes and Allan at the National Bureau of Standards which provides a near optimum estimation of T_F in the presence of both flicker and white noise. The time or frequency offsets measured are weighted as follows:

$$\hat{T}_F\ (n\Delta t) = \frac{1}{M+1} - \left[T_F\ (nt) + m\,\hat{T}_F \left| (n-1)\ \Delta t \right| \right]$$

where M = Number of data points

m = Filter Time Constant

The mean square error of this technique considering white noise FM and flicker noise FM is:

$$\epsilon^2\ (\tau) = \frac{h_0\ (m+1)}{2\ m+1}\ \tau + \frac{h_{-1}\ \tau^2}{m^2\ (2m+1)} \sum_{K=1}^{x} \left(\frac{m}{m+1}\right)^K K^2\ \ln(K)$$

This expression has been evaluated assuming the control segment uses a cesium super tube (HP option 004) and the spacecraft uses a typical rubidium. Also assumed were m = 0.6 and a maximum prediction interval of 8.64×10^4 seconds (24 hours).

Table 7-1 summarizes the results of this analysis and lists the mean prediction error for various times since the last calibration.

TABLE 7-1

MEAN CLOCK PREDICTION ERROR

TIME SINCE LAST CALIBRATION (Seconds)	MEAN PREDICTION ERROR (ns)
1	0.014
10	0.044
10^2	0.16
10^3	0.92
10^4	8.3
10^5	81.6

ASSUMED CONDITIONS:

	$\frac{h_o}{\quad}$	$\frac{2\ln(2)\ h_{-1}}{\quad}$
Cesium	5×10^{-23}	3×10^{-14}
Rubidium	2×10^{-22}	6×10^{-13}

7.2.2 Linear

A linear filter makes an estimate of T_F considering only white noise FM on the clock frequency. The optimum estimate of clock frequency is:

$$T_F = \frac{T_i(t) - T_i(t - \tau_c)}{\tau_c}$$

Because this algorithm does not consider flicker FM noise, it will provide inferior results if this type of noise is present in the measurement data. This will be the case after 15 minutes to one hour for rubidium clocks and after 2 days to one month for cesium clocks.

7.2.3 Carlson Triangular

This technique is essentially an upper triangular square root version of the conventional covariance filter (Kalman). It carries the square root of the error covariance matrix of its state vector in upper triangular form by exploiting Cholesky decomposition. The algorithm is given explicitly in Appendix B of Reference 2. This approach which is the baseline technique, may be susceptible to flicker FM noise, however, the accuracy of this approach is expected to approximate the accuracy of Least Squares/Exponential and is less sensitive to unstable clocks. Thus, the GPS clock error contribution cited in the Summary, Section 1, is as shown in Table 7-1.

The clock errors cited above assume that the satellite clock update does not involve a physical resetting of the satellite clock. As is shown in Part II, Vol. A, Report A3, Section 4.3, a frequency change resolution of 1×10^{-14} is necessary in order to keep the clock error from exceeding 1 nanosecond after 10^5 seconds. Because the best physically realizable resolution will be significantly greater than this, the clock correction should be accomplished via the navigation data message whenever possible.

---

[2] Carlson "Fast Triangular Formulation of the Square-Root Filter," AIAA Journal, Sept. 1973.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Philco-Ford, WDL Division
Palo Alto, Calif. 94303

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
Global Positioning System (GPS).
Part II, Volume A,
System Trades and Analyses.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Final Report for GPS System Definition Study 28 Jun 73 - 28 Feb 74.

5. AUTHOR(S) (First name, middle initial, last name)

6. REPORT DATE
28 Feb 1974

7a. TOTAL NO. OF PAGES
426

7b. NO. OF REFS

8a. CONTRACT OR GRANT NO.
F04701-73-C-0295

8b. PROJECT NO.

8c.

8d.

9a. ORIGINATOR'S REPORT NUMBER(S)
WDL-TR5291-Pt-2-Vol-A

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
SAMSO TR-74-183-Pt-2-Vol-A

10. DISTRIBUTION STATEMENT

DISTRIBUTION STATEMENT B

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
HQ Space and Missile System Organization
P.O. Box 92960, Worldway Postal Center
Los Angeles, Calif. 90009

13. ABSTRACT

Provides analyses of the GPS signal structure, RF link, time references, ionospheric delay, relativistic errors, and security aspects of the navigation signal.

402 711

D FORM 1 NOV 1473

UNCLASSIFIED

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Global Positioning System						
RF Signal Structure						
RF Link Analysis						
Stable Clock Modeling						
Ionospheric Analysis						
Relativistic Clock Errors						
Secure Navigation Signal						