

PRECISION GEODESY AND ASTROMETRY VIA
VERY-LONG-BASELINE INTERFEROMETRY

by

Alan Robert Whitney

B.S., Massachusetts Institute of Technology

M.S., Massachusetts Institute of Technology

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January, 1974

PART I

Signature of Author _____
Department of Electrical Engineering, January 16, 1974

Certified by _____ Thesis Supervisor

Accepted by _____
Chairman, Departmental Committee on Graduate Students



PRECISION GEODESY AND ASTROMETRY VIA
VERY-LONG-BASELINE INTERFEROMETRY

by

ALAN ROBERT WHITNEY

Submitted to the Department of Electrical Engineering
on January 16, 1974 in partial fulfillment of the re-
quirements for the degree of Doctor of Philosophy.

ABSTRACT

Very-long-baseline interferometry (VLBI) has the potential to perform geodetic measurements accurate to the centimeter level and astrometric measurements accurate to a very small fraction of a second of arc. Until recently, only phase-delay-rate (i.e. fringe-rate) measurements had been used for such purposes. The accuracies achievable with such measurements are on the order of a few meters in baseline components and about 0.1 arc-second in source positions. Also, only the equatorial components of the baseline are measurable with this method. Accurate group delay measurements, on the other hand, which are relatively insensitive to fringe-phase variations with periods of minutes, were not made because of the limited bandwidth of the recording systems. Group-delay measurements, if accurate to the nanosecond or better level, would overcome most of the limitations of the phase-delay-rate measurements and, as well, allow the determination of clock-synchronizations and all three components of the baseline.

In this thesis, a technique is described, known as "frequency-switched" VLBI, which allows a wide recording bandwidth to be synthesized by sequential sampling of several frequency "windows" distributed widely compared to the instantaneous recording bandwidth. In this fashion, the effective recorded bandwidth may easily exceed the instantaneous recorded bandwidth by a factor of a hundred or more, allowing group-delay measurements to be made to nanosecond or sub-nanosecond accuracy. When implemented with highly-stable frequency standards (such as the hydrogen maser) and digital recording, the

"frequency-switched" technique is significantly more powerful, for geodetic and astrometric measurements, than measurements of phase-delay rate alone.

The thesis begins with a brief review of the development of VLBI. The "frequency-switched" technique is then presented, along with analyses of maximum-likelihood estimates of observables and analysis of signal-to-noise relations. The effect of one-bit clipping of the data is discussed and compared to the analog case. Practical data-processing algorithms are developed and discussed in detail and the limitations of the algorithms are examined.

The application of this "frequency-switched" technique in a transcontinental 3-station VLBI experiment in October 1969, using antennas in Massachusetts, West Virginia, and California is then described. Observations were made at both L-band (~1660 MHz) and X-band (~7800 MHz), with a "synthesized" bandwidth spanning 36 MHz around each of these frequencies. From these data, the components of all three baselines were estimated to accuracies of a few meters and the positions of eight sources estimated to within a few tenths of an arc-second. More recent experiments, the results of which are briefly discussed, have considerably improved upon these accuracies. The future of the technique is discussed and suggestions are made for continuing work.

Thesis Supervisor: Irwin I. Shapiro

Title: Professor of Geophysics and Physics

ACKNOWLEDGEMENTS

The work reported in this thesis has been made possible only through the cooperation and support of many people and institutions.

I am greatly indebted to my thesis supervisor, Prof. I. I. Shapiro, who provided constant support, expertise, and enthusiasm throughout my thesis research and most of my graduate school career at M.I.T. Many people have been sources of encouragement and helpful technical discussions. Among them are Dr. A. E. E. Rogers, Dr. T. A. Clark, Dr. H. F. Hinteregger, Dr. J. M. Moran, Jr., Prof. C. C. Counselman III, C. A. Knight, and D. S. Robertson. I am especially grateful to the staff of the Haystack Observatory, whose knowledge, helpfulness, and support have immeasurably aided in the many aspects of this theses.

The experiment reported here would not have been possible without the support and cooperation of the National Radio Astronomy Observatory in Green Bank, West Virginia, the Owens Valley Radio Observatory in Big Pine, California, and the Space Sciences Computing Center of the NASA Goddard Space Flight Center. To all of these I extend my thanks and appreciation.

A special thanks are due to Dr. and Mr. T. A. Clark, who so generously received me during my long stay at the NASA Goddard Space Flight Center while the data-reduction software was being developed.

Finally, I am grateful to my parents, Mr. and Mrs. Robert H. Whitney, who wholeheartedly supported my early interest in science. And special thanks to my father, who volunteered to do the drawings for this thesis.

The final preparation of this thesis was much aided by the skill and constant good humor of my typist, Mrs. Judith Ungermann.

This work was supported in part by the National Science Foundation (Grant GA-36283X1) and in part by the Advanced Research Projects Agency.

TABLE OF CONTENTS

	<u>Page</u>
CHAPTER 1. INTRODUCTION	15
1.1 Brief History of Very-Long-Baseline Interferometry	15
1.2 Analog to Aperture-Synthesis Antenna Array	21
1.3 Early Frequency-Switched VLBI Experiments	26
1.4 Synopsis of Contents of Thesis	32
CHAPTER 2. THE GEOMETRY OF VLBI MEASUREMENTS	35
2.1 Basic VLBI Geometry	35
2.2 Interferometer Resolution	43
2.3 Measurement Sensitivities	46
2.3-1 Sensitivity of Delay and Fringe Rate to Small Changes in Model Parameters	46
2.3-2 Some Comments on the Character of $\Delta\tau$ and $\Delta\nu$	52
CHAPTER 3. SIGNAL AND NOISE ANALYSIS	57
3.1 Introduction	57
3.2 Optimum Estimates of Observables	59
3.2-1 Introduction	59
3.2-2 Maximum-Likelihood Estimates of Observables: Frequency Domain Analysis	59
3.2-3 Maximum-Likelihood Estimate of Observables: Time Domain Analysis	70
3.3 Signal-to-Noise Analysis	76

	<u>Page</u>
3.4 Calculation of the Probability Density Functions of Correlation Amplitude and Phase	89
3.4-1 Probability Density Functions for $T_a \ll T_s$	89
3.4-2 Alternative Derivation of Probability Density for the "No-Signal" Case	95
3.5 Effects of Infinite Clipping on Cor- relation Coefficient and SNR	99
3.6 Group-Delay and Phase-Delay Rate Estimation	108
3.6-1 Single-Band Delay Estimation	108
3.6-2 Switched-Frequency Delay Estimation	110
3.6-3 Single-Band Delay Rate Estimation	112
3.6-4 Switched-Frequency Delay Rate Estimation	113
3.7 Signal Detection Characteristics	117
3.7-1 Signal Detection Probabilities	117
3.7-2 Distribution of Maximum Cor- relation Amplitude in No- Signal Case	121
3.8 Summary of Important Results	124

	<u>Page</u>
CHAPTER 4. DATA PROCESSING - ALGORITHMS AND SOFTWARE	129
4.1 The Basic Data-Processing Procedure	129
4.2 Historical Setting	132
4.3 The Basic Delay and Delay Rate Estimation Algorithm	134
4.4 The VLBI-1 Algorithms	139
4.5 The VLBI-2 Algorithms	145
4.5-1 Cross-Spectrum Computation	147
4.5-2 "Coarse-Search" in Delay and Delay Rate	147
4.5-3 "Fine-Search" in Delay and Delay Rate	151
4.6 Analysis of Possible Systematic Errors Due to Data-Processing Algorithms	159
4.7 Implementation and Testing of VLBI-1 and VLBI-2	178
4.7-1 Implementation of VLBI-1 and VLBI-2	178
4.7-2 Testing of VLBI-1 and VLBI-2	180
4.8 Brief Description of VLBI-3	183
CHAPTER 5. DESIGN OF A 3-STATION FREQUENCY-SWITCHED VLBI EXPERIMENT	186
5.1 General Outlines and Goals of the Experiment	186
5.2 Choice of RF Observing Frequencies	188
5.3 Choice of Switched-Frequency Windows	191
5.4 Expected System Performance and Measurement Accuracies	194

CHAPTER 6. INSTRUMENTATION FOR EXPERIMENT	198
6.1 Introduction	198
6.2 Basic System Description	199
6.3 Feed Horn System	202
6.4 RF Amplification System	203
6.5 Local-Oscillator System	209
6.5-1 General Comments	209
6.5-2 High-Frequency Local Oscil- lators	211
6.5-3 Low-Frequency Local Oscil- lators	216
6.6 Frequency Converters and IF System	219
6.7 Recording Systems	221
6.7-1 General Comments	221
6.7-2 The NRAO Mark I Recording System	223
6.8 Frequency Standards	226
6.8-1 Characteristics of Existing Standards	226
6.8-2 Frequency-Stability Require- ments	229
6.8-3 Frequency-Standards Used for October 1969 Experiment	233
6.9 The OVRO Antenna	234
6.10 Construction, Testing, and Instal- lation of the Receiver at OVRO	235

	<u>Page</u>
6.10-1 Receiver Construction	235
6.10-2 Receiver Testing and In- stallation	240
CHAPTER 7. CONDUCT, ANALYSIS AND RESULTS OF THE EX- PERIMENT	242
7.1 Preliminary Setup and Testing	242
7.2 Conduct of the Experiment	245
7.2-1 Scheduling of Observations and Sources Observed	245
7.2-2 Observing Frequencies	247
7.2-3 Details of a Single Observation	249
7.2-4 Operational Difficulties Encountered	250
7.3 Data Processing through VLBI-1 and VLBI-2	253
7.3-1 Determination of the Coarse Clock Synchronizations	253
7.3-2 Determination of Phase-Calibration Correction Phases	254
7.3-3 Mass Processing of Data through VLBI-1 and VLBI-2	256
7.3-4 Examples of Results of Individual Observations	256
7.3-5 Delay and Delay Rate Closure Tests	276

	Page
7.4 Baseline and Source Position	280
Results	
7.4-1 Baseline Solutions	283
7.4-2 Baseline Closure Test	286
7.4-3 Character of the Post-Fit	291
Residuals	
7.4-4 Comparison of VLBI and	295
Survey Baseline Results	
7.4-5 Source Position Results	298
7.5 Conclusions and Suggestions for	303
Further Work	

	<u>Page</u>
NOTATION FOR APPENDICES	306
APPENDIX A Covariance Analysis of the Analog Cross-Correlation Estimate	308
APPENDIX B Covariance Analysis of the Normalized Analog Cross-Correla- tion Estimate	313
APPENDIX C Cross-Covariance Analysis of the Analog Cross-Correlation and Auto-Correlation Estimates	319
APPENDIX D Mean and Variance of the Norma- lized Cross-Correlation Estimate from an Infinitely-Clipped Signal	326
APPENDIX E Review of Least-Squares Fit of Data to a Straight Line	336
APPENDIX F Derivation of the Algorithms Used by VLBI-1 and VLBI-2	344
REFERENCES	351
BIOGRAPHICAL NOTE	355

LIST OF ILLUSTRATIONS

Figure	Page
1-1 The Delay Resolution Function	23
1-2 Simplified Block Diagram of a Frequency-Switched VLBI System	28
2-1 Basic VLBI Coordinate System	36
3-1 Receiver Models for Maximum-Likelihood Analysis	60
3-2 Graphic Representation of the Signal and Noise Relations	84
3-3 Vector Diagram of Signal and Noise Vectors for Derivation of Amplitude and Phase Distribution of \hat{R}	90
3-4 The Probability Density of Correlation Amplitude	94
3-5 Probability of Successful Signal Detection	119
3-6 Probability Density Function for the Maximum Correlation Amplitude in the No-Signal Case	123
4-1 Vector Representation of Delay Resolution Function	157
4-2 Consequences of Computing Truncated Cross-Correlation Function	163
6-1 Block Diagram of OVRO Receiving and Recording System	200
6-2 Block Diagram of High-Frequency Oscillator System	214
6-3 RMS Fractional Frequency Stability for a Variety of Commonly-Used Frequency Standards	227

Figure	Page
6-4 Completed Receiver Box	238
6-5 Interior Equipment Rack	238
6-6 Receiver Box with Top Cover Removed	239
7-1 X-Band Switched-Frequency Observation on NRAO-Haystack Baseline	262
7-2 X-Band Switched-Frequency Observation on NRAO-OVRO Baseline	264
7-3 X-Band Switched-Frequency Observation on Haystack-OVRO Baseline	266
7-4 L-Band Single-Frequency Observation on NRAO-Haystack Baseline	269
7-5 L-Band Single-Frequency Observation on NRAO-OVRO Baseline	271
7-6 L-Band Single-Frequency Observation on Haystack-OVRO Baseline	273
7-7 Delay and Delay-Rate Closure Errors	279
7-8 Baseline Solution Results	284
7-9 Typical Delay and Delay-Rate Residuals from Haystack-NRAO Baseline	293
7-10 Typical Delay and Delay-Rate Residuals from Haystack-OVRO Baseline	294
D-1 Bias of Infinitely-Clipped Cross-Correlation Function	332

LIST OF TABLES

Table		Page
5-1	Approximate Expected Parameters for the Individual Antenna Sites	196
5-2	Expected Signal Parameters	197
7-1	Primary Sources Observed and their <u>A Priori</u> Coordinates	246
7-2	Frequency Sequences Used During the Experiment	248
7-3	Results of X-Band Switched-Frequency Observation	268
7-4	Results of L-Band Single-Frequency Observation	275
7-5	Number of Successful Observations on each Source and Baseline	281
7-6	Processing Parameters for Baseline Solutions	285
7-7	Baseline-Length Solutions	287
7-8	Results of Baseline Closure Test	290
7-9	Comparison of VLBI and Survey Baseline Results	297
7-10	Comparison of Source Position Results	300
D-1	Bias and Variance of Infinitely-Clipped Cross-Correlation Function	333

CHAPTER 1

INTRODUCTION

1.1 BRIEF HISTORY OF VERY-LONG-BASELINE INTERFEROMETRY

Very-long-baseline interferometry (VLBI) is still a relatively young science. The very first experiments were conducted by a group at the University of Florida (References 1, 2) in 1965 to investigate radio bursts from the planet Jupiter. Data were recorded independently by analog recorders at each end of the baseline, and the recorded bandwidth of 2.4 kHz was small enough that timing control pulses could be sent over telephone lines. The observing frequency of 18 MHz was low enough that frequency signals from WWV could be used to control the local oscillators.

It was not until 1967 that VLBI measurements at microwave frequencies with truly independent time and frequency standards was demonstrated by researchers in Canada (Reference 3), closely followed by NRAO-Cornell (Reference 4) and independently by an MIT-NRAO group (Reference 5 ¹). The system developed by the Canadian researchers was a sophisticated extension of the analog recording and processing techniques used earlier at Florida. The recorded bandwidth was extended to 4 MHz through the use of video tape machines and special techniques were developed to synchronize recordings during playback for

¹Both the NRAO-Cornell and MIT-NRAO groups, however, used the same digital recording system (so-called "Mark I" system) developed by NRAO.

processing.² The system developed by NRAO, on the other hand, made use of a fully digitized data recording system which had the major advantage that the recordings were fully compatible with tape systems on large general-purpose computer systems, leading ultimately to greatly increased flexibility and reliability of VLBI measurements. The one major drawback of the digital system, however, was a limitation of the recorded bandwidth to ~ 360 kHz, resulting in lower signal-to-noise ratios on continuum observations than were obtainable with the Canadian analog system.

The early VLBI measurements made by NRAO, Cornell, and MIT were primarily aimed at developing high-resolution maps of radio sources and/or placing upper bounds on their angular sizes. Spectral-line sources, such as OH, were prime targets for source-structure studies because individual features may be identified by Doppler differences and their relative positions may be determined from fringe-rate differences between the features (Reference 6). No particular attempts were made in these early measurements to determine highly-accurate absolute source positions or baseline components, due primarily to the limitations imposed by the then-available frequency standards. The rubidium atomic clocks then generally in use as frequency standards possessed fractional stabilities on the order of a few parts in 10^{12} over periods from a few seconds

²Although the recording bandwidth capability of the Canadian system was 4 MHz, the actual video bandwidth used in early experiments was limited to 1 MHz in order to reduce playback-synchronization requirements.

to a few hours, which translates into uncertainties of a large fraction of a microsecond per day. This limits absolute source position measurements to accuracies of the order of several seconds of arc or more, and baseline measurements to several tens of meters, making any such measurements relatively uncompetitive with short-baseline-interferometer position measurements or conventional high-precision ground surveys (References 7, 8, 9).

In late 1967 the first hydrogen-maser frequency standards became available for limited use in VLBI experiments. The hydrogen maser offered at least a ten-fold improvement in frequency stability over rubidium standards, reportedly as high as a few parts in 10^{14} over periods of hours. Immediate impact was felt on VLBI measurements, for suddenly it became possible, at least in principle, to model the reference clocks to a few nanoseconds per day, dramatically increasing the potential use of VLBI to determine high-accuracy source positions and baseline components. Potentially, both delay and delay-rate measurements could be made with sufficient accuracy to determine source positions to small

fractions of a second of arc and baseline components to the several-centimeter level. The prospect, though tantalizing, was still a long way from reality as limitations unrelated to the frequency standards became dominant. The existing VLBI recording systems were primarily of use to measure accurate phase-delay rates, sometimes called "fringe rates". Measurements of total phase-delay, however, were impossible due to insufficient knowledge of initial conditions. Precise measurements of group-delay, on the other hand, required the examination of the relative phases of the received signals over a bandwidth wider than was possible to obtain with the existing VLBI recording systems. The accuracy of the group-delay measurement is inversely proportional to the recorded bandwidth, the group-delay measurement accuracy of the 360 kHz bandwidth digital recording system being on the order of a few tenths of a microsecond. Similarly, the Canadian 4 MHz bandwidth analog recording system has the potential of group-delay measurement accuracies of a few tens of nanoseconds. These accuracies, although perhaps impressive, are still far too small to take full advantage of the few-nanoseconds-per-day stability offered by hydrogen masers.

High-accuracy group-delay measurements offer some advantages over phase-delay-rate measurements although they are in some ways complementary. As we shall see in Chapter 2, phase-delay rate is independent of the polar component of the baseline, so that only the equatorial components may be estimated. In addition, the declination of at least one source must be known a priori in order to solve for source positions. Group-delay measurements, on the other hand, suffer from neither of these disadvantages. In addition, clock synchronization errors may be determined (a posteriori, of course) to approximately the same level of accuracy as the accuracy of the group-delay measurements. No other technique, so far as is known to the author, has the realistic potential of synchronizing widely-separated clocks to sub-nanosecond accuracy.

Owing to the seeming advantages of group-delay measurements in conjunction with phase-delay-rate measurements, much thought was given as to how such measurements might be made within the basic framework of the existing VLBI recording hardware, in particular the 360 kHz bandwidth digital recording system. The problem was solved in a rather ingenious manner after a suggestion by A. E. E. Rogers (Reference 10) and has become commonly known as "frequency-switched" VLBI. Rogers suggested sequential sampling of several "frequency windows", each of bandwidth 360 kHz but spanning overall a much larger bandwidth. The sampling of these windows would be cyclical so that each frequency window would be sampled many times during the course of a

few-minute observation. When the tapes are brought together for processing each frequency window may be processed as a separate observation, yielding a "fringe-phase" for each window. Then, provided the frequency windows are not too widely spaced and that some means is available to calibrate the instrumental phases for each window, it should be possible to extrapolate the fringe phase from one window to the next and hence synthesize an apparent recorded bandwidth equal to the largest separation of the frequency windows. Such a procedure is independent of the instantaneous recorded bandwidth, provided that the signal-to-noise ratio in each frequency window remains acceptably high. An appreciation of the possible improvement in group-delay measurement accuracy may be had by noting that a group-delay uncertainty of 0.1 microsecond, typical for a relatively high signal-to-noise measurement over a single 360 kHz bandwidth, may be reduced, by frequency-switching over a total bandwidth 100 times as large (36 MHz), to an uncertainty of about one nanosecond.

1.2 ANALOG TO APERTURE-SYNTHESIS ANTENNA ARRAY

The "frequency-switched" VLBI technique for determining group delay has a nearly exact analog in the linear multiple-antenna aperture-synthesis array. It is useful to pursue this analogy to understand the principle and technique of frequency-switched VLBI. Consider a phased linear-array of antennas. The Fourier transform of the spatial array structure defines a "beam pattern" on the sky (see, for example, Reference 11) which, depending on the details of the array spacings, generally has many repeating "fringe groups", of which each group consists of a "central fringe" plus several sidelobes. One of these fringe groups is the "central fringe group". The "central fringe" of the "central fringe group" is defined as that single fringe which would be present if the entire linear array were filled. The approximate angular width of this fringe is λ/D where λ is the wavelength and D is the maximum spacing of the array. Now suppose a point source is moving through the beam of the antenna array and the problem is to determine when the source is at the peak of the "central fringe" of the "central fringe group". This determination may be made in three logical steps: 1) Determine when the source is within the "central fringe group", 2) Determine when the source is on the "central fringe" of that fringe group, and 3) Ascertain when the peak of the central fringe has been reached. Step 1 is

usually accomplished by choosing the "fringe group" spacings, controlled primarily by the minimum array spacing, to be large enough that the proper fringe group can be chosen from a priori knowledge of the source position. Picking the proper fringe within the fringe group, Step 2, is usually assisted by choosing the relative array spacings so that the sidelobes immediately surrounding the central fringe of the group are substantially lower in amplitude. Step 3 is then simply a matter of "peaking up" the amplitude, assuming that the proper fringe has been located by Steps 1 and 2.

If, instead of antenna spacings, we think of frequency-window spacings, the Fourier transform of these frequency-window spacings forms a function known as the "delay resolution function", which is entirely analogous to the "beam pattern" of the antenna array. Figure 1.1a shows schematically an array of six frequency windows placed at the relative frequencies of 0, 1, 4, 6, 24, and 36 MHz, respectively. Figures 1-1b and 1-1c show the corresponding delay resolution function. A sharply defined peak occurs every 1 microsec (inverse of the minimum window spacing), each peak being surrounded by complicated lower sidelobes. Each such peak and its immediate sidelobes are exactly analogous to a "fringe group" of the antenna array, while

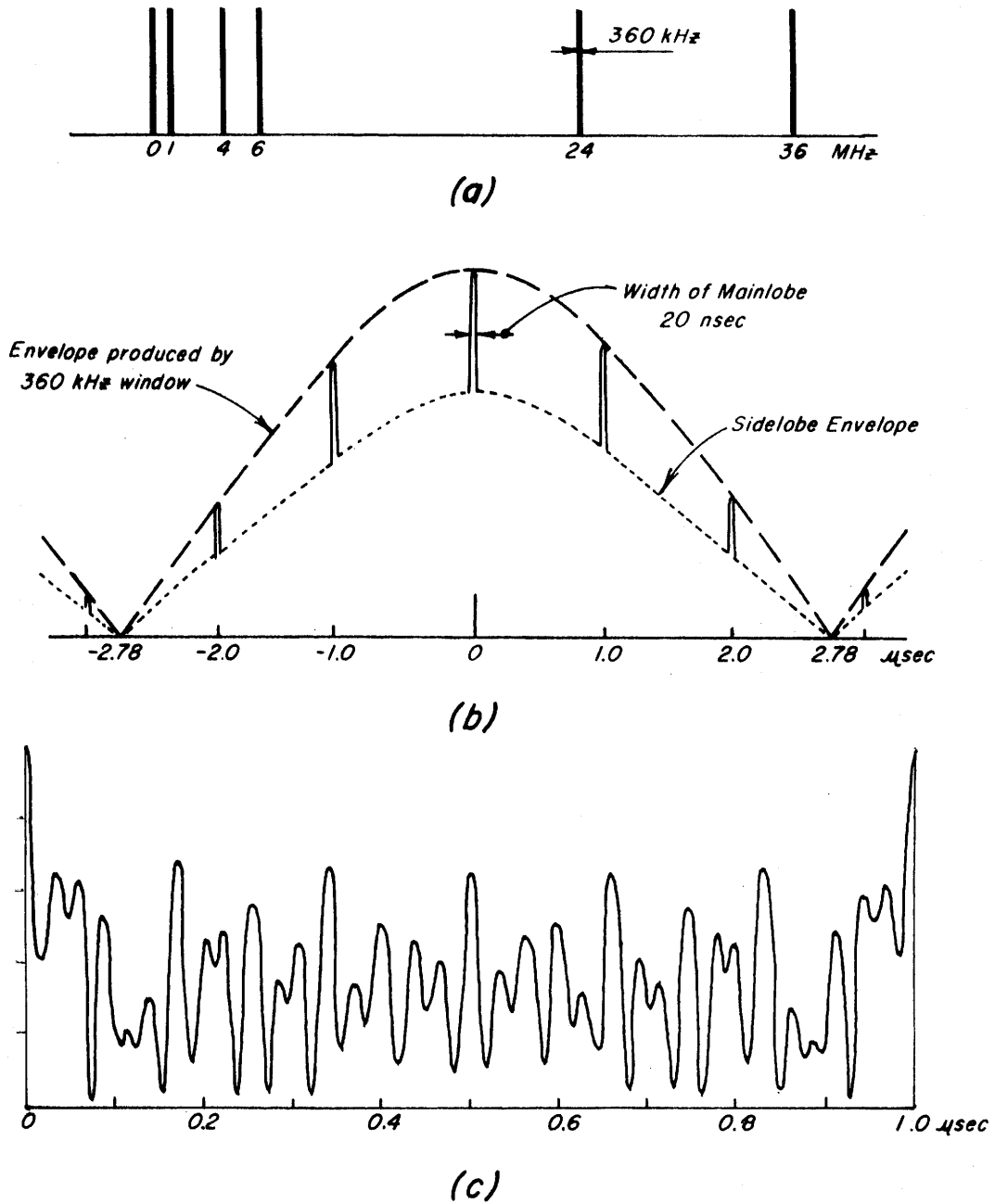


Figure 1-1

a) Schematic representation of frequency windows placed at relative frequencies of 0, 1, 4, 6, 24 and 36 MHz.

b) Delay resolution function corresponding to this set of frequency windows. Detailed sidelobe structure omitted.

c) Details of sidelobe structure between 0 and 1.0 μsec , ignoring envelope produced by 360 kHz window.

each peak corresponds to the "central fringe" of that group. The width of each peak is approximately 15 nanoseconds, approximately the inverse of the maximum window spacing, and the detailed sidelobe structure between peaks is determined by the spacing of the interior windows. A broad envelope, produced by the 360 kHz bandwidth of the individual windows, surrounds the entire function, in the same way that the size of the individual antennas in an array imposes a broad envelope over the beam pattern.

Our analogy to the antenna array is now nearly complete, and the three-step procedure outlined to "determine when a source is on the peak of the central fringe of the central fringe group" can be translated directly, by analogy, to "find the maximum of the main peak of the central delay ambiguity". The procedure, then, is to 1) use a priori knowledge to determine the proper ambiguity of the delay resolution function; 2) find the main peak within this ambiguity, and 3) maximize on the main peak.

The reader may be prompted to ask why a priori knowledge is generally used to find the proper delay ambiguity because, as indicated in Figure 1.1b, the peak of the proper ambiguity is higher than any of the others and should require no a priori knowledge to properly choose it. The answer is twofold: 1) noise on a given observation may, in fact, cause the wrong ambiguity to be higher than the proper one, and 2) data processing in a

computer is much speeded if each frequency window is processed individually and then only the amplitude and phase from each window is used to construct the delay resolution function; this effectively removes the broad envelope, due to the 360 kHz individual-window bandwidth, from the delay resolution function and makes all ambiguities of equal amplitude.

A selection of the exact spacing of the frequency windows is of course very important. The minimum spacing must be chosen small enough that a priori information is sufficient to choose between the major ambiguities of the delay resolution function. The maximum spacing in large part determines the width of the main peak of the delay resolution function, but must not be chosen so large that sidelobes become too high. The intermediate frequencies are chosen to lower the sidelobe levels of the delay resolution function as much as possible. Generally a set of frequency spacings must be chosen which reflects a compromise between high delay resolution and tolerable sidelobe levels. Higher sidelobe levels may be tolerated as a priori knowledge of the group delay increases. Some semi-empirical optimization of window-spacings has been attempted (Reference 10) but no general rules of optimization have yet been formulated (see also Section 5.3).

1.3 EARLY FREQUENCY-SWITCHED VLBI EXPERIMENTS

The first frequency-switched VLBI experiment was conducted in April 1968 between the 120' antenna of the Haystack Observatory in Tyngsboro, Mass. and the 84' radio facility of Chalmers University at Onsala, Sweden. The observing frequency was chosen to be ~1660 MHz, near the frequency of earlier OH observations. In order to implement this experiment a considerable amount of new equipment was built and tested. The results from this first experiment, however, were very few and poor due mainly to the problems involved in "shaking down" all of the new equipment.

In October of 1968 a second switched-frequency experiment was attempted between Haystack and the NRAO 140' antenna at Green Bank, W. Va. The primary purpose of this experiment was to monitor the apparent position difference between the quasars 3C273B and 3C279 as the sun travelled through their vicinity. The results, it was hoped, would enable a precise determination of the relativistic bending of electromagnetic waves in a gravitational field. The experiment was conducted at two widely separated frequency bands, the first near 7500 MHz (X-band) spanning a synthesized bandwidth of 44 MHz with 16 frequency windows and the second near 1610 MHz

(L-band) spanning a synthesized bandwidth of ~5 MHz with 8 frequency windows. Using two widely-separated bands of frequencies it was hoped to be possible to measure the effect of the solar corona on the delay measurements, and hence remove its effect in the gravitational-bending determination.

Figure 1-2 shows a basic block diagram of the equipment used for the October 1968 experiment and is typical of frequency-switched VLBI experiments. At each station the incoming signal is amplified by a low-noise amplifier before being mixed with a radio-frequency local-oscillator signal. (Separate amplifiers and local oscillators were used for L-band and X-band, but for simplicity only one is shown in Figure 1-2.)

A second stage of mixing translates the signal to a band spanning 0-360 kHz, where it is clipped, sampled and recorded on magnetic tape. The independent hydrogen masers at each site control all local oscillator signals as well as all timing and control signals. Both local oscillators at each site are controlled in a programmed manner so that different "frequency windows" are sequentially selected. During the October 1968 experiment both the first and second local oscillators at each site were programmed to "switch frequencies"; it was necessary to switch first local oscillators because of the relatively narrow bandwidth of the IF between the first and second mixing stages. Switching of the first local-oscillator

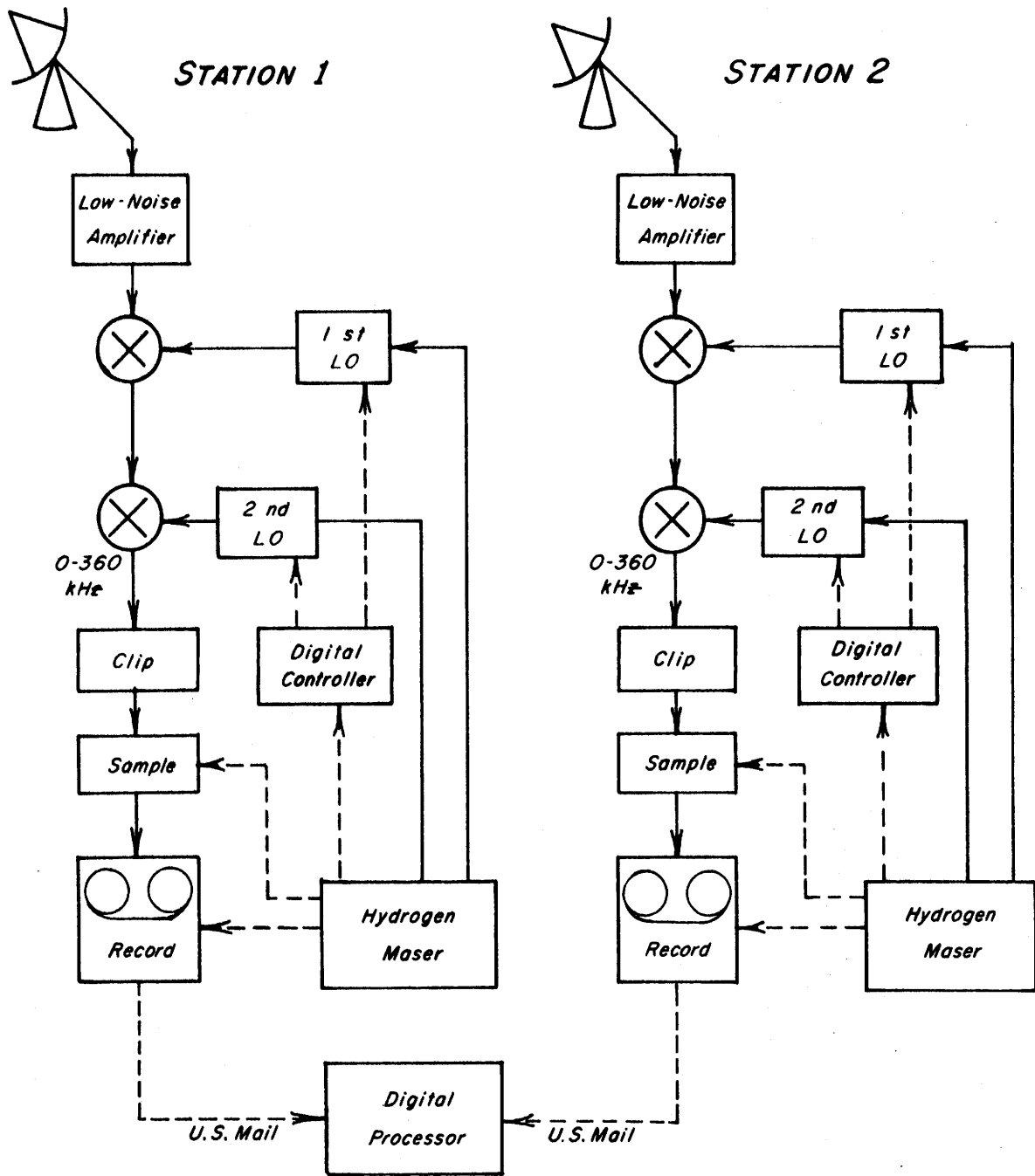


Figure 1-2

Simplified block diagram of frequency-switched VLBI system

signal proved to be particularly troublesome because it was generated with the aid of a phase-locked loop which had to be retuned and relocked at each change of frequency, at a switching rate of five times per second.

Despite the equipment troubles encountered, many observations from the October 1968 experiment showed good fringes but the delay results were disappointing. They showed scatter of many tens of nanoseconds, far more than could be accounted for if all the equipment were performing properly. Subsequent investigations revealed that the most probable primary cause of the observed problems were switching transients in the first local oscillator signal. These transients were apparently not so large or fast as to prevent fringes from being observed but caused the observed fringe phases to drift randomly over an unacceptably large range. The net result was that the wide bandwidth could not be adequately "synthesized", and the measured group delays were unreliable (Reference 12).

After some improvements in equipment and procedure, another switched-frequency VLBI experiment was conducted in January 1969 between Haystack and the NRAO 140' antenna (Reference 13). All observations were made at L-band near 1660 MHz, spanning a synthesized bandwidth

of 110 MHz with six frequency windows. Both the first and second local oscillators were switched and, in addition, the L-band parametric amplifiers were synchronously retuned by means of programmed bias voltages. Reduction of the data from this experiment yielded a 3-component vector baseline between a well-defined point on each antenna as well as high-accuracy source positions. The agreement of the VLBI-determined baseline with that of a high-precision survey was to within 2 meters in length and to about 5 meters in orientation (References 13, 14). The position of six radio sources were simultaneously determined to accuracies varying from ~ 0.2 to 1.0 arc-seconds. These results still were not up to initial expectations, but represented a significant improvement in the state-of-the-art.

Following the January 1969 experiment, a reassessment of the switched-frequency technique was made. A persistent and nagging problem still remained in switching the first local oscillator. Even though improvements had been made in the area, switching transients was still a major concern and no really suitable solutions seemed in sight. Fortunately, at approximately the same time, very wide-band solid-state IF amplifiers were becoming available. With the availability of these amplifiers, IF bandwidths of 100 MHz were not uncommon and the need to switch the first local oscillator suddenly disappeared. In ad-

dition, broadband low-noise microwave amplifiers were also becoming available so that the need to synchronously retune the rf amplifiers also vanished. The result of these developments was that the switched-frequency VLBI hardware could be significantly simplified, and only the second local oscillator needed to be switched. This simplified procedure has been used in all succeeding switched-frequency VLBI experiments.

In the spring of 1969 planning began for another switched-frequency experiment to be conducted in October 1969, the goals being to make high-accuracy geodetic and astrometric measurements as well as another attempt to measure the gravitational bending of radio waves. The experiment was to be conducted at two frequencies, L-band and X-band, and involved the participation of three major radio antennas: the 120' Haystack antenna in Tyngsboro, Mass., the 140' NRAO antenna at Green Bank, W. Va., and one of the 90' antennas at the Owens Valley Radio Observatory near Big Pine, Calif. This was the first switched-frequency VLBI experiment to be attempted using three antennas and offered a unique opportunity to collect simultaneous data on three baselines, allowing, for the first time, internal consistency checks to be made around

three baselines. Among other checks, each baseline vector may be determined individually and a realistic estimate of the measurement accuracy determined by examining the sum of the three independently-determined baseline vectors around the three baselines. Similarly, source positions may be determined from data taken on each individual baseline, and the resulting positions compared for consistency. Goals for measurement accuracies were set at about 1 meter in baseline vector components and a tenth of a second of arc for source positions.

Although the initial goals for the October 1969 experiment proved to be somewhat optimistic, the result of the experiment yielded results significantly improved over previous results and a still greater appreciation of some of the remaining problems. The author has been heavily involved with the development of the switched-frequency VLBI technique starting with the first experiment in April 1968 and has contributed to the effort in almost all of its phases, including theory, hardware, and software.

1.4 SYNOPSIS OF CONTENTS OF THESIS

By its very nature, VLBI involves many people working simultaneously on its many aspects. The work presented in this thesis is part of a continuing program

of VLBI development being carried out at MIT and elsewhere. Earlier work in this area has been published in theses by Moran (Reference 6) and Hinteregger (References 13, 14). Preston (References 15, 16) has applied the technique of VLBI to satellite tracking. Rogers (References 10, 17) has made extensive contributions to both the engineering of VLBI systems and the analysis of VLBI data. The reader is referred to these sources for extensive background material. In addition, a number of good general-interest articles have been published on the subject (References 18, 19, 20, 21, 22).

Chapters 2 and 3 of this thesis are intended to outline the basic principles underlying VLBI. Chapter 2 details the basic geometry of VLBI measurements and shows how VLBI may be applied to precision geodesy and astrometry. The sensitivity of interferometer measurements to geodetic and astrometric parameters is discussed in some detail. Chapter 3 discusses the signal and noise analysis problems of VLBI from the viewpoint of a relatively simple model. Signal detection problems are studied and the effects of infinite-clipping of the signals are discussed and compared to the analog signal case. Estimates of group-delay and phase-delay-rate measurement accuracies are made, resulting in simple, useful formulas.

Chapter 4 outlines in detail the data processing

algorithms and procedures that have actually been implemented and used for the analysis of VLBI data. Also, an analysis is made of possible systematic errors introduced into the data processing due to the many approximations that had to be made in the VLBI data-processing system.

Chapters 5, 6, and 7 deal with the design and execution of the October 1969 VLBI experiment which was briefly described in the previous section. The design and goals of the experiment are presented, the instrumentation is discussed in some detail, and the actual execution of the experiment is outlined. Chapter 7 gives the results and post mortems of the experiment and discusses possible areas for improvement and future work.

CHAPTER 2

THE GEOMETRY OF VLBI MEASUREMENTS

2.1 BASIC VLBI GEOMETRY

The basic geometry of VLBI measurements has been discussed by several authors [among them Moran (Reference 6) and Williams (Reference 23)] and we shall briefly review it here in the light of the experiments conducted for this thesis. In this simple analysis, we shall ignore any effects of polar motion, aberration, and relativity. For more details, the reader is referred to Ash (Reference 24) and Williams (Reference 23).

Consider the two right-handed frames of reference shown in Figure 2-1, designated as S and S', with the origin of both frames being at the center-of-mass of the earth. Frame S is defined by the non-rotating axes x, y, z. The z-axis is aligned with the instantaneous spin axis of the earth. The x-axis is defined to lie in the direction of the true equinox of date and the y-axis completes a right-handed coordinate set. The second reference frame, designated S' and with axes x', y', z', is fixed in the rotating earth. The axes z and z' are coincident. The axis x' is in the direction of the Greenwich meridian and the y' axis completes the right-handed set. The right-ascension of date of the Greenwich meridian, designated α_G , relates S and S'. The transformation from S to S' is

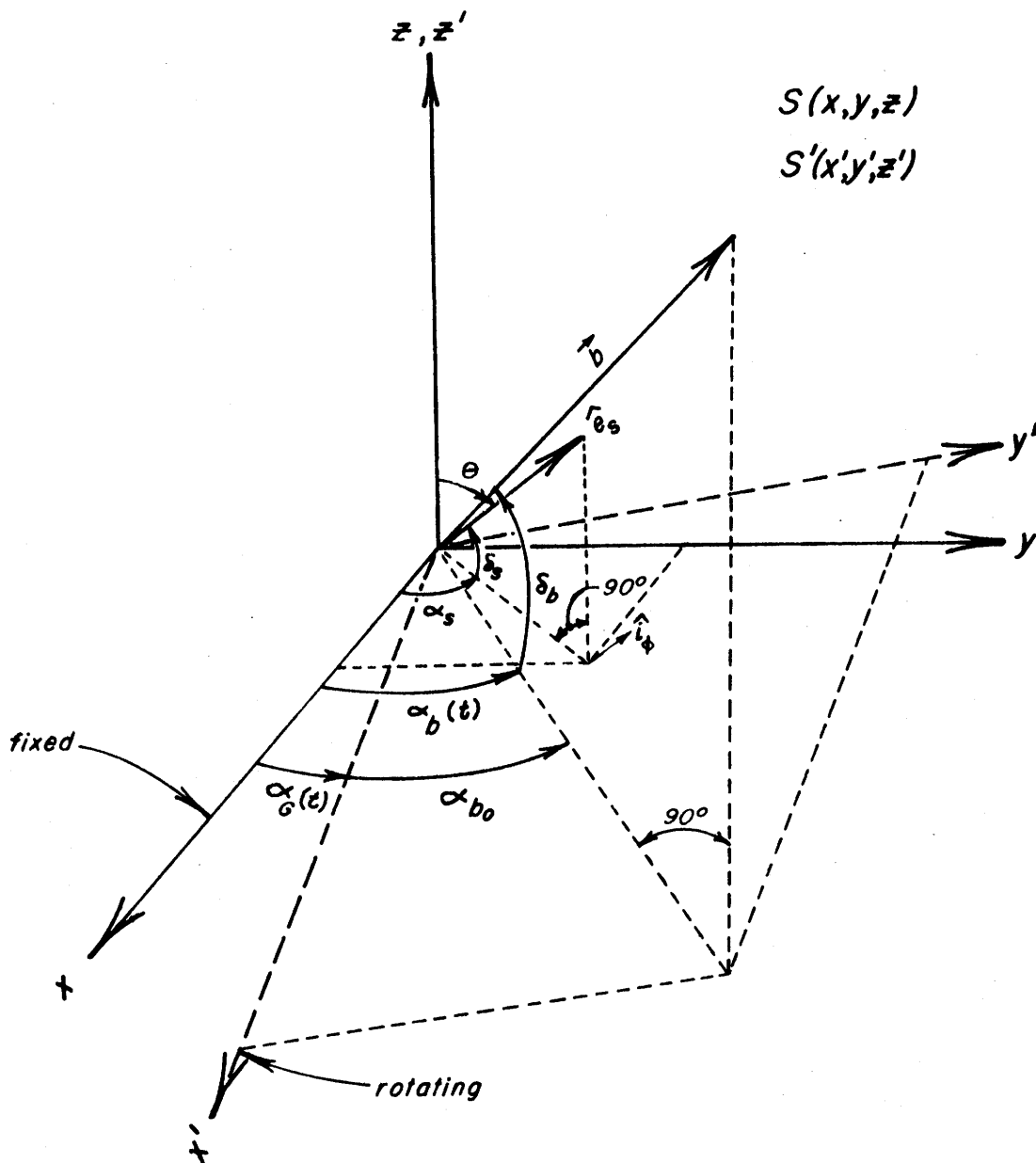


Figure 2-1

Basic VLBI Coordinate System

$$\begin{aligned}
 \hat{i}'_x &= \hat{i}_x \cos \alpha_G + \hat{i}_y \sin \alpha_G \\
 \hat{i}'_y &= -\hat{i}_x \sin \alpha_G + \hat{i}_y \cos \alpha_G \\
 \hat{i}'_z &= \hat{i}_z
 \end{aligned}
 \tag{2-1}$$

and, similarly, from S' to S by

$$\begin{aligned}
 \hat{i}_x &= \hat{i}'_x \cos \alpha_G - \hat{i}'_y \sin \alpha_G \\
 \hat{i}_y &= \hat{i}'_x \sin \alpha_G + \hat{i}'_y \cos \alpha_G \\
 \hat{i}_z &= \hat{i}'_z
 \end{aligned}
 \tag{2-2}$$

Define the baseline vector \vec{b} fixed in S'

$$\vec{b} = b'_x \hat{i}'_x + b'_y \hat{i}'_y + b'_z \hat{i}'_z
 \tag{2-3}$$

where b'_x , b'_y , b'_z are constant components of the baseline and not a function of time. (In Figure 2-1 we show \vec{b} translated in the S' frame so that one end of it is fixed at the origin of S'. This is done simply to clarify the geometry of Figure 2-1 and will have no effect on the results of our computations as long as we assume that the observed sources are infinitely distant.) If we define the baseline right-ascension and declination, α_b and δ_b , respectively, as shown in Figure 2-1, then \vec{b} may be written in the S frame as

$$\vec{b} = b [\cos \delta_b (\cos \alpha_b \hat{i}_x + \sin \alpha_b \hat{i}_y) + \sin \delta_b \hat{i}_z]
 \tag{2-4}$$

where $b \equiv |\vec{b}|$ is the length of the baseline and α_b , δ_b are the instantaneous right-ascension and declination, respectively, of \vec{b} within the S coordinate frame. α_b is a function of time

$$\alpha_b(t) = \alpha_{b_0} + \alpha_G(t) \quad (2-5)$$

where α_{b_0} is the baseline right ascension at the instant S and S' are aligned (i.e. when the sidereal time of the Greenwich meridian is zero).

In a similar manner we may define the "source vector" \hat{e}_s to be a unit vector in the direction of the radio source fixed in the S coordinate system

$$\hat{e}_s = \cos\delta_s (\cos\alpha_s \hat{i}_x + \sin\alpha_s \hat{i}_y) + \sin\delta_s \hat{i}_z \quad (2-6)$$

where α_s, δ_s are the right-ascension and declination of date, respectively, of the source. We may use the transformation of Equation (2-2) to write \hat{e}_s in the S' frame as

$$\hat{e}_s = \cos\delta_s [\cos(\alpha_G - \alpha_s) \hat{i}'_x - \sin(\alpha_G - \alpha_s) \hat{i}'_y] + \sin\delta_s \hat{i}'_z \quad (2-7)$$

If we define τ_g to be the travel time of a wavefront, propagating from the source in the direction $-\hat{e}_s$, from the "reference station" at the foot of \vec{b} to the "remote station" at the head of \vec{b} , then

$\tau_g = -\frac{1}{c}(\vec{b} \cdot \hat{e}_s)$. From Equations (2-3) and (2-7),

$$\tau_g = -\frac{1}{c} \{ \cos\delta_s [b'_x \cos(\alpha_G - \alpha_s) - b'_y \sin(\alpha_G - \alpha_s)] + b'_z \sin\delta_s \} \quad (2-8)$$

where c is the speed of light. Using Equations (2-4) and (2-6) we may rewrite Equation (2-8):

$$\tau_g = -\frac{b}{c}[\cos\delta_s \cos\delta_b \cos(\alpha_b - \alpha_s) + \sin\delta_b \sin\delta_s] \quad (2-9)$$

We should note two points regarding the results of Equations (2-8) and (2-9): 1) these results do not include any effects of polar motion, aberration, relativity, or transmission media, and 2) since this analysis is based on the propagation of a single wavefront from one station to the other, these are, strictly speaking, phase delay (as opposed to group delay) results. For propagation in a vacuum, as we have assumed, phase and group delay are, of course, the same.

We note that the bracketed term of Equation (2-8) and the first term of Equation (2-9) vary sinusoidally with a period of one sidereal day; the remaining term, the so-called "polar-component" of τ_g , is constant and, in a real experiment, is indistinguishable from a clock synchronization error if observations are made only on a single source.

In a realizable VLBI experiment, τ_g cannot be measured directly since it is not readily separable from the effects of media propagation and instrumental delays. More specifically, if we define τ_t to be the total propagation time of a particular wavefront as measured by comparing the clock readings made by the crew at each station upon the arrival of that wavefront at their station, we would find τ_t to be made up of several independent components

$$\tau_t = \tau_g + \tau_c + \tau_i + \tau_p \quad (2-10)$$

where τ_c = synchronization error of the clocks at the two sites

τ_i = instrumental delay differences (due to amplifiers, waveguides, cables, mixers, etc.) between the two sites

τ_p = propagation media delay differences due to atmosphere, ionosphere, interplanetary and interstellar plasmas (although scale sizes of interstellar matter are usually much larger than the earth-based baseline and hence negligible)

Complicating the measurements is the fact that τ_c , τ_i , and τ_p may be time-dependent and behave in an unpredictable and/or unmeasurable manner. We shall examine each of these in somewhat more detail later.

We may take the time derivative of Equation (2-10) to get the total delay rate

$$\dot{\tau}_t \equiv \frac{d\tau_t}{dt} = \frac{d\tau_g}{dt} + \frac{d\tau_c}{dt} + \frac{d\tau_i}{dt} + \frac{d\tau_p}{dt} \quad (2-11)$$

where we have assumed that τ_g , τ_c , τ_i and τ_p may all be described independently. The quantity known as total "fringe rate"¹ is then

¹The term "fringe rate" has grown naturally into the VLBI repertoire of nomenclature, and has an intuitive meaning if one visualizes the "fringe pattern" to be a set of parallel lines in the sky, fixed in the frame of the rotating earth, and sweeping past the source at a rate of $2\pi v_t$ lines per second, where v_t is defined in Equation (2-12).

$$v_t = \omega \dot{i}_t + \Delta\omega_{\ell_0} \quad (2-12)$$

where ω is the rf observing frequency that is translated to zero-frequency (i.e. DC) in the video spectrum of the "reference" station and where $\Delta\omega_{\ell_0}$ is the local oscillator "offset" at the "remote" station.

The main component of v_t is the component due to \dot{i}_g . From Equations (2-5), (2-8), and (2-9) we have the geometric component of the fringe rate, v_g ,

$$v_g = \omega \dot{i}_g \quad (\text{rad/sec}) \quad (2-13a)$$

$$= \frac{\omega}{c} \cos\delta_s [b'_x \sin(\alpha_G - \alpha_s) + b'_y \cos(\alpha_G - \alpha_s)] \frac{\partial\alpha_G}{\partial t} \quad (2-13b)$$

$$= \frac{\omega b}{c} \cos\delta_s \cos\delta_b \sin(\alpha_b - \alpha_s) \frac{\partial\alpha_G}{\partial t} \quad (2-13c)$$

where $\frac{\partial\alpha_G}{\partial t} \approx 0.73 \times 10^{-4}$ rad/sec is the rotation rate of the earth. For $\omega = 10$ GHz and $b = 6000$ km, we see from Equation (2-13) that v_g may be as large as $\sim 10^4$ Hz. The maximum v_g occurs, as we would intuitively expect, when $\alpha_s - \alpha_b = \frac{\pi}{2}$ (i.e. when $\vec{b} \cdot \hat{e}_s = 0$). We note that only the "equatorial component" of the baseline, $b \cos\delta_b$, appears in the expression for v_g , so that fringe rate measurements alone can measure only b'_x and b'_y of Equation (2-3). No information is gained about the z-component, b'_z . It is also interesting to note from Equation (2-13c) that the

source declination is inseparable from the equatorial baseline component, except by observing several sources. And even then, if the equatorial baseline component is not known, at least one additional constraint must be applied in order to solve for the source declination. For example, the declination of one source could be fixed, or a priori source declination information could be used to restrain the "average" declination corrections.

2.2 INTERFEROMETER RESOLUTION

The minimum fringe spacing of the interferometer occurs along a direction in the sky parallel to the projection of the baseline vector, \vec{b} , onto a plane perpendicular to \hat{e}_s . If, as before, \hat{e}_s and \vec{b} are the source and baseline vectors, respectively, then this projection is, from simple geometry, just the vector

$$\vec{b}_p \equiv (\hat{e}_s \times \vec{b}) \times \hat{e}_s \quad (2-14)$$

The minimum fringe spacing occurs along this direction and may be visualized as a set of parallel lines drawn perpendicular to the plane which contains \vec{b}_p and \hat{e}_s . The angular spacing of the lines is $\lambda/|\vec{b}_p|$, where λ is the wavelength at which the observation is being made. We recognize $\lambda/|\vec{b}_p|$ as just the angle through which \vec{b}_p must be rotated in order to change the signal delay (i.e. the difference in time of arrival at the two ends of the baseline of a signal from the source) by λ/c .

In most VLBI work, the "resolution" is expressed in terms of the fringe spacing in the "East-West" and "North-South" directions. If we define a right-handed spherical coordinate system (r, θ, ϕ) coincident with the already defined Cartesian coordinate system S of Figure 2-1, then the East-West component of \vec{b}_p , as viewed from the source, is $\vec{b}_p \cdot \hat{i}_\theta$, where \hat{i}_θ is perpendicular to \hat{e}_s (see Figure 2-1).

The resolution in the East-West direction (i.e. \hat{i}_θ direction), conventionally designated as u and expressed as inverse fringe spacing, is

$$\begin{aligned} u &= \frac{\vec{b}_p \cdot \hat{i}_\theta}{\lambda} \\ &= \frac{b}{\lambda} \cos\delta_b \sin(\alpha_b - \alpha_s) \quad (\text{fringes/radian}) \quad (2-15) \end{aligned}$$

where we have used Equations (2-5), (2-7), and (2-14).

Similarly, the "North-South" resolution (more correctly, the resolution in the \hat{i}_ϕ direction) is designated by v and is given by

$$\begin{aligned} v &= \frac{\vec{b}_p \cdot \hat{i}_\phi}{\lambda} \\ &= -\frac{b}{\lambda} [\sin\delta_s \cos\delta_b \cos(\alpha_b - \alpha_s) - \sin\delta_b \cos\delta_s] \quad (2-16) \end{aligned}$$

The quantities u and v are, of course, functions of time as the baseline rotates. From Equations (2-15) and (2-16) we note that

$$\frac{u^2}{a^2} + \frac{(v-v_0)^2}{b^2} = 1 \quad (2-17)$$

where

$$a = \frac{b}{\lambda} \cos\delta_b$$

$$b = \frac{b}{\lambda} \cos\delta_b \sin\delta_s$$

$$v_0 = \frac{b}{\lambda} \sin\delta_b \cos\delta_s$$

Equation (2-17) indicates that u and v trace out an ellipse as a result of the earth's rotation. It should also be noted that any earth-based VLBI experiment may make measurements only over the part of the u - v "track" for which the source is visible to both stations. In the special case that both stations are in the same hemisphere and the magnitude of the source declination is larger than the magnitude of the site's latitudes (North Pole in the Northern Hemisphere, South Pole in the Southern Hemisphere), the source will be visible to both stations at all times and the entire u - v track may be observed.

2.3 MEASUREMENT SENSITIVITIES

2.3-1 Sensitivity of Delay and Fringe Rate to Small Changes in Model Parameters

In this section we shall examine the effects on τ_g and v_g of small errors in the baseline vector, source position vectors, polar motion and UT1.² In doing so we may gain an idea of the accuracy with which these quantities may be determined with VLBI. In this discussion we shall ignore propagation media and instrumentation effects except for clock synchronization and rate errors. Analyses of this type have been done by others [see, for example, Moran (Reference 6), Rogers (Reference 10), Williams (Reference 23)] and here we shall give a brief discussion relevant to the experiments carried out for this thesis. We shall also examine the relative sensitivities of delay and fringe-rate data and show how delay and fringe-rate data are in large part complementary.

Consider an observation made at time t which yields a measured delay τ_m and measured fringe rate v_m . In order to carry out the processing of the data, a set of a priori baseline, source position, and clock synchronization parameters $b'_x, b'_y, b'_z, \alpha_s, \delta_s, \tau_c$ [as defined by Equations (2-3), (2-6), and (2-10)] are assumed. Parameters associated with precession and nutation appear only when the coordinate system of date is transformed into a reference coordinate

²UT1 is a measure of time defined by the earth's rotation angle. Small irregularities in the rotation rate of the earth may be measured by determining UT1 as a function of atomic time.

system (such as the mean equinox and equator of 1950.0), and need not immediately concern us here. The observed delay τ_m may be written

$$\tau_m = \tau_{g_o} + \tau_{c_o} + \Delta\tau \quad (2-18)$$

where τ_{g_o} is the a priori geometric delay as given by Equation (2-8) or (2-9), τ_{c_o} is the a priori clock synchronization error, and $\Delta\tau$ is the residual delay unaccounted for by the a priori model. If the measurement errors are negligible and if the errors in the model are small, we may do a first-order expansion of $\Delta\tau$ around the a priori

$$\begin{aligned} \Delta\tau = & \frac{\partial\tau_g}{\partial b'_x} \Delta b'_x + \frac{\partial\tau_g}{\partial b'_y} \Delta b'_y + \frac{\partial\tau_g}{\partial b'_z} \Delta b'_z + \frac{\partial\tau_g}{\partial \alpha_s} \Delta \alpha_s \\ & + \frac{\partial\tau_g}{\partial \delta_s} \Delta \delta_s + \Delta\tau_c \end{aligned} \quad (2-19)$$

where $\Delta b'_x$, $\Delta b'_y$, $\Delta b'_z$, $\Delta \alpha_s$, $\Delta \delta_s$ and $\Delta\tau_c$ are the errors in the model parameters and all the partial derivatives are understood to be evaluated at the a priori values.

Similarly, the observed fringe rate ν_m may be written as

$$\nu_m = \nu_{g_o} + \omega \dot{\tau}_{c_o} + \Delta\nu \quad (2-20)$$

where ν_{g_o} is the a priori fringe rate, ω is the observing frequency, and $\dot{\tau}_{c_o}$ is the assumed clock-rate offset. For small errors in the model we may write

$$\Delta v = \frac{\partial v}{\partial b'_x} \Delta b'_x + \frac{\partial v}{\partial b'_y} \Delta b'_y + \frac{\partial v}{\partial b'_z} \Delta b'_z + \frac{\partial v}{\partial \alpha_s} \Delta \alpha_s + \frac{\partial v}{\partial \delta_s} \Delta \delta_s + \omega \Delta \tau_c \quad (2-21)$$

where again it is understood that all partial derivatives are to be evaluated at the a priori values. We note that the radio frequency ω is defined with respect to the frequency standard of each station and is not an unknown or variable parameter of the model.

We now wish to evaluate each of the partial derivatives in Equations (2-19) and (2-21) and make order-of-magnitude estimates of their values.³ To meaningfully examine these estimates we need to keep in mind that the approximate accuracies of real measurements using the switched-frequency technique with the available equipment in 1969 was on the order of a few nanoseconds (10^{-9} sec) in delay and a millihertz (10^{-3} Hz) in fringe rate.

First consider the dependence of τ_g upon the baseline components. From Equation (2-8)

$$\frac{\partial \tau_g}{\partial b'_x} = - \frac{1}{c} \cos \delta_s \cos (\alpha_G - \alpha_s) \quad (2-22a)$$

$$\frac{\partial \tau_g}{\partial b'_y} = \frac{1}{c} \cos \delta_s \sin (\alpha_G - \alpha_s) \quad (2-22b)$$

$$\frac{\partial \tau_g}{\partial b'_z} = - \frac{1}{c} \sin \delta_s \quad (2-22c)$$

These all have an order-of-magnitude value of $1/c$ or

³This type of sensitivity analysis is very simplified. A more meaningful discussion would require a complete correlation or covariance analysis.

about 3 ns/meter, independent of the length of the baseline. Similarly, the fringe-rate partials are, from Equation (2-13b)

$$\frac{\partial v_g}{\partial b'_x} = \frac{\omega}{c} \cos \delta_s \sin(\alpha_G - \alpha_s) \frac{\partial \alpha_G}{\partial t} \quad (\text{rad/sec-m}) \quad (2-23a)$$

$$\frac{\partial v_g}{\partial b'_y} = \frac{\omega}{c} \cos \delta_s \cos(\alpha_G - \alpha_s) \frac{\partial \alpha_G}{\partial t} \quad (2-23b)$$

$$\frac{\partial v_g}{\partial b'_z} = 0 \quad (2-23c)$$

Recalling that $\frac{\partial \alpha_G}{\partial t} \approx 0.73 \times 10^{-4}$ rad/sec is just the rotation rate of the earth, and choosing the observing frequency to be ~ 7.8 GHz, we find the sensitivity of v_g to the baseline components b'_x and b'_y to be about 2×10^{-3} Hz/m. Note that, as with the delay partials, the sensitivity is independent of the baseline. Comparing the sensitivity to the corresponding delay sensitivity of ~ 3 ns/m we see that, for the assumed observation frequency and observational accuracies, delay and fringe-rate measurements will lead to nearly the same accuracies in the equatorial baseline components b'_x and b'_y . Fringe-rate information alone, of course, gives us no information about the polar baseline component, b'_z .

Next let us examine the sensitivities of τ_g and ν_g to small changes in source position. From Equation (2-8)

$$\frac{\partial \tau_g}{\partial \alpha_s} = - \frac{1}{c} \cos \delta_s [b'_x \sin(\alpha_G - \alpha_s) + b'_y \cos(\alpha_G - \alpha_s)] \quad (2-24a)$$

$$\frac{\partial \tau_g}{\partial \delta_s} = \frac{1}{c} \{ \sin \delta_s [b'_x \cos(\alpha_G - \alpha_s) - b'_y \sin(\alpha_G - \alpha_s)] - b'_z \cos \delta_s \} \quad (2-24b)$$

We recognize the quantity in brackets in Equation (2-24a) to be dependent on the equatorial components of the baseline. Order-of-magnitude calculations with a baseline length of ~3000 km yield a sensitivity of about 50 ns/arc-sec.

Similarly, the partials of fringe rate with respect to source position may be found from Equation (2-13b)

$$\frac{\partial \nu_g}{\partial \alpha_s} = - \frac{\omega}{c} \cos \delta_s [b'_x \cos(\alpha_G - \alpha_s) - b'_y \sin(\alpha_G - \alpha_s)] \frac{\partial \alpha_G}{\partial t} \quad (2-25a)$$

$$\frac{\partial \nu_g}{\partial \delta_s} = - \frac{\omega}{c} \sin \delta_s [b'_x \sin(\alpha_G - \alpha_s) + b'_y \cos(\alpha_G - \alpha_s)] \frac{\partial \alpha_G}{\partial t} \quad (2-25b)$$

Order-of-magnitude values for these partials for a baseline length of ~3000 km and an observing frequency of ~7800 MHz are about 25×10^{-3} Hz/arc-sec.

Finally, the sensitivity of τ_g and ν_g to UT1 may be found by differentiating Equations (2-8) and (2-13b) with respect to the Greenwich hour angle α_G

$$\frac{\partial \tau_g}{\partial \alpha_G} = \frac{1}{c} \cos \delta_s [b'_x \sin(\alpha_G - \alpha_s) + b'_y \cos(\alpha_G - \alpha_s)] \quad (2-26a)$$

$$\frac{\partial \nu_g}{\partial \alpha_G} = \frac{\omega}{c} \cos \delta_s [b'_x \cos(\alpha_G - \alpha_s) - b'_y \sin(\alpha_G - \alpha_s)] \frac{\partial \alpha_G}{\partial t} \quad (2-26b)$$

We note from Equations (2-24a) and (2-26a) that

$$\frac{\partial \tau_g}{\partial \alpha_s} = - \frac{\partial \tau_g}{\partial \alpha_G} \quad (2-27)$$

and from Equations (2-25a) and (2-26b) that

$$\frac{\partial \nu_g}{\partial \alpha_s} = - \frac{\partial \nu_g}{\partial \alpha_G} \quad (2-28)$$

so that α_s and α_G are, in fact, directly related and hence cannot be determined independently. What Equations (2-27) and (2-28) really tell us is that the origin of right-ascension cannot be determined by measurements of τ_g and ν_g . In general, a right-ascension origin must be established by other means; for example, a priori source right-ascension information may be used to constrain the "average" right-ascension corrections to a minimum.

2.3-2 Some Comments on the Character of $\Delta\tau$ and $\Delta\nu$

Suppose a series of observations are made on one source so that $\Delta\tau(t)$ and $\Delta\nu(t)$ of Equations (2-19) and (2-21) are measured at several times during a single pass of the source. Suppose further that the independent clocks used at the two sites have an initial synchronization error τ_c at $t=0$ and a constant rate offset $\dot{\tau}_c$. Then Equations (2-22) and (2-24) may be used in Equation (2-19) to write

$$\Delta\tau(t) = A_\tau \cos(\alpha_G - \alpha_S) + B_\tau \sin(\alpha_G - \alpha_S) + C_\tau + D_\tau t \quad (2-29)$$

where

$$A_\tau = \frac{1}{c} [-\Delta b'_x \cos \delta_s - b'_y \Delta \alpha_s \cos \delta_s + b'_x \Delta \delta_s \sin \delta_s] \quad (2-30)$$

$$B_\tau = \frac{1}{c} [\Delta b'_y \cos \delta_s - b'_x \Delta \alpha_s \cos \delta_s - b'_y \Delta \delta_s \sin \delta_s] \quad (2-31)$$

$$C_\tau = -\frac{1}{c} [\Delta b'_z \sin \delta_s + b'_z \Delta \delta_s \cos \delta_s] + \tau_c \quad (2-32)$$

$$D_\tau = \dot{\tau}_c \quad (2-33)$$

Thus we see that a series of observations on one source will yield 4 knowns ($A_\tau, B_\tau, C_\tau, D_\tau$) compared to the 6 unknowns ($\Delta b'_x, \Delta b'_y, \Delta b'_z, \Delta \delta_s, \tau_c, \dot{\tau}_c$. ($\Delta \alpha_s$ must be fixed since VLBI gives no information about a right-ascension origin.) Clearly this is insufficient for a complete solution. Observations taken on each additional source will yield 3 more knowns (since D_τ is common for all sources) along with 2 more unknowns ($\Delta \alpha_s, \Delta \delta_s$ for each additional source), so that observations

on three sources (10 unknowns) are the minimum necessary to determine all the unknown parameters. An absolute minimum set of observations to determine all parameters (except one right-ascension) would consist of 4 observations on one source and 3 observations each on the remaining two sources.

In a similar fashion Equations (2-23) and (2-25) may be used in Equation (2-21) to write the fringe-rate error Δv as

$$\Delta v(t) = A_v \cos(\alpha_G - \alpha_s) + B_v \sin(\alpha_G - \alpha_s) + C_v \quad (2-34)$$

where

$$A_v = \frac{\omega}{c} \frac{\partial \alpha_G}{\partial t} [\Delta b'_y \cos \delta_s - b'_x \Delta \alpha_s \cos \delta_s - b'_y \Delta \delta_s \sin \delta_s] \quad (2-35)$$

$$B_v = \frac{\omega}{c} \frac{\partial \alpha_G}{\partial t} [\Delta b'_x \cos \delta_s + b'_y \Delta \alpha_s \cos \delta_s - b'_x \Delta \delta_s \sin \delta_s] \quad (2-36)$$

$$C_v = \omega \tau_c \quad (2-37)$$

But comparison of Equations (2-29) and (2-30) shows immediately that

$$A_v = \omega \frac{\partial \alpha_G}{\partial t} B_\tau \quad (2-38)$$

$$B_v = -\omega \frac{\partial \alpha_G}{\partial t} A_\tau \quad (2-39)$$

$$C_v = \omega D_\tau \quad (2-40)$$

These last equations imply that the constants to be solved for in Equations (2-29) and (2-30) are equivalent, except that b'_z and τ_c do not appear in the expression for $\Delta v(t)$. Thus, the addition of fringe-rate data to an

existing set of delay data does not reduce the minimum number of sources on which observations are necessary in order to solve for all the unknown parameters.

For fringe rate data alone, observations on a single source yield 3 knowns (A_v, B_v, C_v). Each additional source yields 2 more knowns (since C_v is common to all sources).

As we implied earlier, the declination of one source and the right-ascension of one source must be fixed, or some equivalent constraints applied.

Consider observations on n different sources. Then, 3 observations on one source and 2 observations on each of the $(n-1)$ other sources allows complete solution for $\Delta b'_x, \Delta b'_y, \tau_c, (n-1)$ right-ascension corrections and $(n-1)$ declination corrections.

It is interesting to note how delay and fringe-rate observations complement one another. In order to show this we shall define the "sensitivity vectors" \vec{S}_τ and \vec{S}_v for delay and delay-rate measurements, respectively. Consider a source with assumed a priori coordinates α_s, δ_s , and with coordinate errors $\Delta\alpha_s$ and $\Delta\delta_s$. The actual arc-lengths of the position errors are $\Delta\delta_s$ in the direction of increasing declination and $\Delta\alpha_s \cos\delta_s$ in the direction of increasing right-ascension. We are therefore led to define

$$\dot{S}_\tau = \frac{\partial \tau_g}{\partial \alpha_s} \frac{1}{\cos \delta_s} \hat{i}_x + \frac{\partial \tau_g}{\partial \delta_s} \hat{i}_y \quad (2-41a)$$

$$\dot{S}_\nu = \frac{\partial \nu_g}{\partial \alpha_s} \frac{1}{\cos \delta_s} \hat{i}_x + \frac{\partial \nu_g}{\partial \delta_s} \hat{i}_y \quad (2-41b)$$

where the length of the \hat{i}_x component is a measure of the sensitivity in the direction of right-ascension and the length of the \hat{i}_y component is a measure of the declination sensitivity. The partial derivatives may be evaluated directly from Equations (2-9) and (2-13c) to yield

$$\begin{aligned} \dot{S}_\tau = -\frac{b}{c} \{ \cos \delta_b \sin(\alpha_b - \alpha_s) \hat{i}_x - [\sin \delta_s \cos \delta_b \cos(\alpha_b - \alpha_s) \\ + \sin \delta_b \cos \delta_s] \hat{i}_y \} \end{aligned} \quad (2-42a)$$

$$\begin{aligned} \dot{S}_\nu = -\frac{\omega b}{c} \frac{\partial \alpha_G}{\partial t} [\cos \delta_b \cos(\alpha_b - \alpha_s) \hat{i}_x \\ + \sin \delta_s \cos \delta_b \sin(\alpha_b - \alpha_s) \hat{i}_y] \end{aligned} \quad (2-42b)$$

Comparing these with the expressions for u and v given in Equations (2-15) and (2-16) we see that the above may be rewritten

$$\dot{S}_\tau = -\frac{1}{c} [(\lambda u) \hat{i}_x + (\lambda v) \hat{i}_y] \quad (2-43a)$$

$$\dot{S}_\nu = -\frac{\omega}{c} \frac{\partial}{\partial t} [(\lambda u) \hat{i}_x + (\lambda v) \hat{i}_y] \quad (2-43b)$$

The quantities λu and λv are just the projections of the baseline on a plane perpendicular to the source direction \hat{e}_s , so we see from these equations that \vec{S}_τ is proportional to these projections, whereas \vec{S}_ν is proportional to their rate of change.

From Equation (2-42) it is evident that $\vec{S}_\tau(t)$ and $\vec{S}_\nu(t)$ both trace out ellipses in the x-y plane with the time-variable components exactly 90° (i.e. 6 hours) out of phase. The ellipses are identical except for a scale factor and for the fact that \vec{S}_τ is offset by a constant vector in the \hat{i}_y direction. In fact, we can easily relate \vec{S}_τ and \vec{S}_ν directly

$$\vec{S}_\tau(t+6 \text{ hrs}) = \left(\omega \frac{\partial \alpha_G}{\partial t}\right)^{-1} \vec{S}_\nu(t) - \frac{b}{c} \sin \delta_b \cos \delta_s \hat{i}_y \quad (2-44)$$

Over large areas of the sky, in particular where $\sin \delta_b \cos \delta_s < \sin \delta_s \cos \delta_b$, $\vec{S}_\tau(t)$ and $\vec{S}_\nu(t)$ are approximately orthogonal. Over these areas of the sky a single measurement of delay and fringe rate is sufficient to establish the position of a source, provided all baseline, clock, atmosphere, etc., parameters are assumed known.

CHAPTER 3

SIGNAL AND NOISE ANALYSIS

3.1 INTRODUCTION

In this chapter we shall examine the signal, noise, and signal-detection problems that arise in VLBI work. First, in Section 3.2, we shall examine a simple statistical model of VLBI observations to determine the optimum procedure, based on the "maximum-likelihood" and "least-squares" criteria, for estimating the VLBI observables. These results will be applied in Section 3.3 to a detailed signal-to-noise analysis; simple expressions for the signal-to-noise ratio and the probability densities of correlation amplitude and phase are developed.

In Section 3.5 we shall examine the effect of infinite-clipping of the recorded signals, which is a common technique used for many VLBI recording systems, and which results in a simple modification of the signal-to-noise expressions derived in Section 3.3. Expressions for the mean and variance of the correlation amplitude estimates for both analog and infinitely-clipped signals are also derived and compared.

In Section 3.6 we examine the statistics of the group delay and phase-delay rate estimates based on the signal-to-noise results of Section 3.3 and on the actual estimation

procedures used in actual VLBI data processing. Simple useful expressions are developed for both switched-frequency and non-switched-frequency observations.

The signal detection problem as applied to VLBI is examined in Section 3.7. Computations of signal detection probabilities are made as a function of the expected signal-to-noise ratio (SNR) and as a function of the supposed accuracy of the a priori model.

Finally, in Section 3.8 we shall give a brief summary of the important results of the chapter that are most often used in day-to-day VLBI work.

3.2 OPTIMUM ESTIMATES OF OBSERVABLES

3.2-1 Introduction

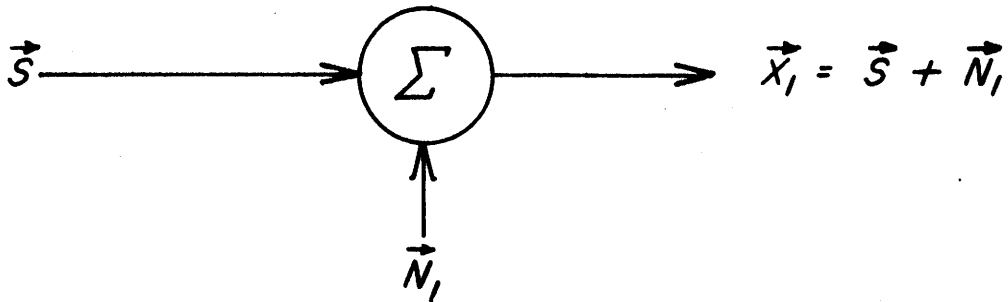
The usual method of VLBI data reduction is to "cross-correlate" the data from the two ends of the baseline. In this section we wish to show that the cross-correlation procedure is in fact the optimum processing procedure for obtaining so-called "maximum-likelihood" estimates of delay, delay rate, and fringe phase. Two slightly different approaches, one in the frequency domain and the other in the time domain, will be taken in Sections 3.2-2 and 3.2-3 to show that both approaches arrive at the same result. Unless otherwise explicitly stated, "delay" will mean group delay and "delay rate" will mean "phase delay rate".

3.2-2 Maximum-Likelihood Estimates of Observables: Frequency Domain Analysis

The maximum likelihood (ML) estimates of delay and fringe phase may be found after the method outlined by Wozencraft and Jacobs (Reference 25), and as applied by Rogers (Reference 10). Here we review and somewhat expand those calculations. For purposes of this analysis, we consider the two receiver systems as diagrammed in Figure 3-1.¹ The signal

¹For simplification of the analysis and "clarity" of presentation of the main idea, we shall not consider the actual heterodyning operations that take place in an actual receiver. The results of such an analysis are equivalent (see Appendix F).

Receiver 1



Receiver 2

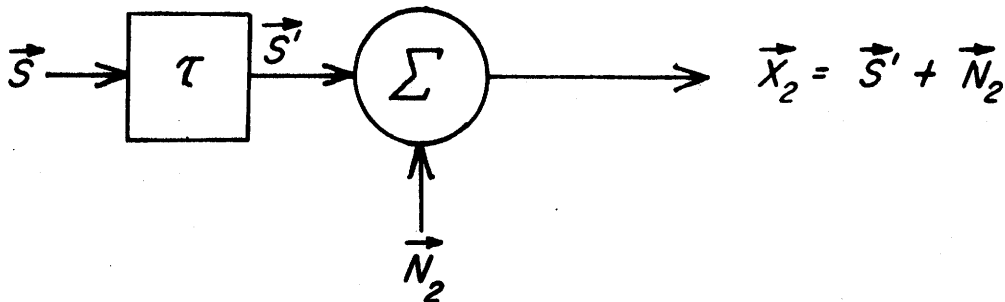


Figure 3-1

Receiver models for maximum-likelihood estimate analysis

\vec{S} is Gaussian noise and is represented by N complex components

$$\vec{S} = (S_1, \dots, S_j, \dots, S_N); \quad 1 \leq j \leq N \quad (3-1)$$

for which each complex component has independent Gaussian real and imaginary parts, and for which all components are independent, i.e. $\langle S_i S_j \rangle = 0$ for $i \neq j$. In practice, each component of \vec{S} corresponds to an independent point in the received frequency spectrum. In the second receiver the signal S suffers a delay so that S' is related to S by

$$\vec{S}' \equiv \vec{S} e^{-i\omega\tau} \quad (3-2)$$

where ω is a frequency in the observed spectrum.

At each station the signals are corrupted by additive independent Gaussian noise signals \vec{N}_1 and \vec{N}_2 ($\vec{N}_1, \vec{N}_2, \vec{S}$ and all their components are assumed independent Gaussian). The observed quantities are then the quantities $\vec{X}_1 = \vec{S} + \vec{N}_1$ and $\vec{X}_2 = \vec{S}' + \vec{N}_2$.

The ML estimate of τ , which we shall denote as $\hat{\tau}$, is defined (see Van Trees, Reference 26) as the value of τ which maximizes the conditional

probability $p(\vec{X}_1, \vec{X}_2 | \tilde{\tau})$, where $\tilde{\tau}$ is the parameter whose value is varied to maximize $p(\vec{X}_1, \vec{X}_2 | \tilde{\tau})$.²

As the first step towards finding the ML estimate of τ we need to compute $p(\vec{S}, \vec{N}_1, \vec{N}_2)$, which is simply

$$p(\vec{S}, \vec{N}_1, \vec{N}_2) = \prod_{j=1}^N \frac{1}{2\pi\sigma_{s_j}^2} \exp\left\{-\frac{|s_j|^2}{2\sigma_{s_j}^2}\right\} \frac{1}{2\pi\sigma_{n_{1j}}^2} \exp\left\{-\frac{|N_{1j}|^2}{2\sigma_{n_{1j}}^2}\right\} \frac{1}{2\pi\sigma_{n_{2j}}^2} \exp\left\{-\frac{|N_{2j}|^2}{2\sigma_{n_{2j}}^2}\right\} \quad (3-3)$$

where we have assumed identical variances for the real and imaginary parts of all signals. Now recalling

² The so-called "maximum a posteriori" (MAP) estimate, on the other hand, is chosen to maximize $p(\tilde{\tau} | \vec{X}_1, \vec{X}_2)$; see, for example, Van Trees (Reference 26). From Bayes' rule

$$\ln p(\tilde{\tau} | \vec{X}_1, \vec{X}_2) = \ln p(\vec{X}_1, \vec{X}_2 | \tilde{\tau}) + \ln p(\tilde{\tau}) - \ln p(\vec{X}_1, \vec{X}_2)$$

Now $p(\vec{X}_1, \vec{X}_2)$ is not a function of $\tilde{\tau}$ and $p(\tilde{\tau})$ approaches a constant for the case where a priori knowledge of τ approaches zero. Therefore, we see that the ML estimate corresponds to the limiting case of a MAP estimate for which a priori knowledge approaches zero.

that $\vec{X}_1 = \vec{S} + \vec{N}_1$, $\vec{X}_2 = \vec{S}' + \vec{N}_2$, that \vec{S}' is related to \vec{S} by Equation (3-2), and that the Jacobian is unity, we have

$$\begin{aligned}
 p(\vec{X}_1, \vec{X}_2, \vec{S} | \tilde{\tau}) &= \prod_{j=1}^N \frac{1}{(2\pi)^3 \sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \cdot e^{-\frac{|s_j|^2}{2\sigma_{s_j}^2}} \\
 &\cdot e^{-\frac{|x_{1j} - s_j|^2}{2\sigma_{n_{1j}}^2}} \cdot e^{-\frac{|x_{2j} - s_j e^{-i\omega_j \tilde{\tau}}|^2}{2\sigma_{n_{2j}}^2}} \\
 &= \prod_{j=1}^N \frac{1}{(2\pi)^3 \sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \\
 &\cdot \exp\left\{ -\frac{|x_{1j}|^2}{2\sigma_{n_{1j}}^2} - \frac{|x_{2j}|^2}{2\sigma_{n_{2j}}^2} \right\} \\
 &\cdot \exp\left\{ -\frac{1}{2} [|s_j|^2 \frac{\sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2 + \sigma_{s_j}^2 \sigma_{n_{2j}}^2 + \sigma_{s_j}^2 \sigma_{n_{1j}}^2}{\sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2}] \right. \\
 &\left. + \operatorname{Re} \left\{ s_j \cdot \left[\frac{x_{1j}^*}{\sigma_{n_{1j}}^2} + \frac{x_{2j}^* e^{-i\omega_j \tilde{\tau}}}{\sigma_{n_{2j}}^2} \right] \right\} \right\} \quad (3-4)
 \end{aligned}$$

Now define

$$\sigma_j^2 = \frac{\sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2}{\sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2 + \sigma_{s_j}^2 \sigma_{n_{2j}}^2 + \sigma_{s_j}^2 \sigma_{n_{1j}}^2} \quad (3-5)$$

$$Y_j = \frac{x_{1j}^* \sigma_j^2}{\sigma_{n_{1j}}^2} + \frac{x_{2j}^* e^{-i\omega_j \tilde{\tau}} \sigma_j^2}{\sigma_{n_{2j}}^2} \quad (3-6)$$

and integrate Equation (3-4) over \vec{S} to get the desired probability density

$$\begin{aligned} p(\vec{x}_1, \vec{x}_2 | \tilde{\tau}) &= \int_{-\infty}^{\infty} p(\vec{x}_1, \vec{x}_2, \vec{S} | \tilde{\tau}) d\vec{S} \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^N \frac{1}{(2\pi)^3 \sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \exp\left\{-\frac{|x_{1j}|^2}{2\sigma_{n_{1j}}^2} \right. \\ &\quad \left. - \frac{|x_{2j}|^2}{2\sigma_{n_{2j}}^2} \right\} \exp\left\{-\frac{|S_j|^2}{2\sigma_j^2} + \frac{\text{Re}[S_j \cdot Y_j]}{\sigma_j^2}\right\} d\vec{S} \quad (3-7) \end{aligned}$$

Recalling that we must integrate over both the real and imaginary parts of S_j , we can complete the square of the last factor of Equation (3-7) to do the integration

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{|S_j|^2}{2\sigma_j^2} + \frac{\text{Re}[S_j \cdot Y_j]}{\sigma_j^2}\right\} dS_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma_j^2} [|S_j - Y_j|^2 - |Y_j|^2]\right\} d[\text{Re}(S_j)] d[\text{Im}(S_j)]$$

$$= 2\pi \sigma_j^2 \exp\left\{\frac{|Y_j|^2}{2\sigma_j^2}\right\} \quad (3-8)$$

so that Equation (3-7) becomes

$$p(\vec{X}_1, \vec{X}_2 | \tilde{\tau}) = \prod_{j=1}^N \frac{\sigma_j^2}{(2\pi)^2 \sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \exp\left\{-\frac{|X_{1j}|^2}{2\sigma_{n_{1j}}^2} - \frac{|X_{2j}|^2}{2\sigma_{n_{2j}}^2} + \frac{|Y_j|^2}{2\sigma_j^2}\right\} \quad (3-9)$$

The ML estimate of τ is now just the value of $\tilde{\tau}$ which maximizes this quantity. Of the terms in the exponent of Equation (3-9) only the last is a function of $\tilde{\tau}$.

From Equation (3-6)

$$|Y_j|^2 = \frac{\sigma_j^4}{\sigma_{n_{1j}}^4} |X_{1j}^*|^2 + \frac{\sigma_j^4}{\sigma_{n_{2j}}^4} |X_{2j}^* e^{-i\omega_j \tilde{\tau}}|^2 + \frac{2\sigma_j^4}{\sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \text{Re}[X_{1j} \cdot X_{2j}^* e^{-i\omega_j \tilde{\tau}}] \quad (3-10)$$

But $|x_{2j}^* e^{-i\omega_j \tilde{\tau}}| = |x_{2j}|$ so that only the last term of Equation (3-10) is a function of $\tilde{\tau}$. To find the ML estimate of τ , then, it is sufficient to maximize

$$\prod_{j=1}^N \frac{\sigma_j^2}{\sigma_{s_j}^2 \sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \exp\left\{ \frac{\sigma_j^2}{\sigma_{n_{1j}}^2 \sigma_{n_{2j}}^2} \operatorname{Re}[X_{1j} \cdot X_{2j}^* e^{-i\omega_j \tilde{\tau}}] \right\} \quad (3-11)$$

If the variances are independent of j (as is the usual case for continuum interferometry) this is equivalent to maximizing over

$$\sum_{j=1}^N \operatorname{Re}[X_{1j} \cdot X_{2j}^* e^{-i\omega_j \tilde{\tau}}] \quad (3-12)$$

But this is just cross-correlation (with fringe rotation) followed by summation over independent points in the frequency spectrum, which is the long-used VLBI estimation procedure (see, for example, Moran, Reference 6, or Hinteregger, Reference 13).

This analysis can easily be extended to find the ML estimate of ϕ , fringe phase, ϕ , simply by substituting

$$s_j = \tilde{s}_j e^{-i[\omega_j \tau + \phi]} \quad ; \quad 0 < t < T \quad (3-13)$$

for Equation (3-2) and carrying through the analysis in a similar manner. Usually, there is no a priori knowledge of the value of ϕ , so that

the ML estimates of τ and ϕ are the same as the MAP estimates and are found by

maximizing $p(\vec{X}_1, \vec{X}_2 | \tilde{\tau}, \tilde{\phi})$ simultaneously over the search parameters $\tilde{\tau}$ and $\tilde{\phi}$. We shall denote the resulting estimate as $\hat{\tau}$ and $\hat{\phi}$, respectively. In this case, it is easy to see that the quantity which is to be maximized is (for the case of variances independent of j)

$$\sum_{j=1}^N \operatorname{Re} \{ X_{1j} \cdot X_{2j}^* e^{-i[\omega_j \tilde{\tau} + \tilde{\phi}]} \} \quad (3-14)$$

Since the sum of the real parts is the real part of the sum, maximizing expression (3-14) is the same as

$$\begin{aligned} & \max_{\tilde{\tau}, \tilde{\phi}} \operatorname{Re} \left[\sum_{j=1}^N X_{1j} X_{2j}^* e^{-i[\omega_j \tilde{\tau} + \tilde{\phi}]} \right] \\ & = \max_{\tilde{\tau}, \tilde{\phi}} \operatorname{Re} \left\{ e^{-i\tilde{\phi}} \sum_{j=1}^N X_{1j} \cdot X_{2j}^* e^{-i\omega_j \tilde{\tau}} \right\} \end{aligned} \quad (3-15)$$

But, clearly, this quantity can always be maximized by choosing $\hat{\phi}$ to make the sum of (3-15) entirely real. Therefore, expression (3-14) may be maximized by first performing

$$\max_{\tilde{\tau}} \left| \sum_{j=1}^N X_{1j} \cdot X_{2j}^* e^{-i\omega_j \tilde{\tau}} \right| \quad (3-16)$$

to find $\hat{\tau}$ and then choosing $\hat{\phi}$ to make the bracketed {} part of expression (3-16) lie entirely along the real axis, i.e. choosing $\hat{\phi}$ such that

$$\tan \hat{\phi} = \frac{\text{Im} \left\{ \sum_{j=1}^N X_{1j} \cdot X_{2j}^* e^{-i\omega_j \hat{\tau}} \right\}}{\text{Re} \left\{ \sum_{j=1}^N X_{1j} \cdot X_{2j}^* e^{-i\omega_j \hat{\tau}} \right\}} \quad (3-17)$$

where $\hat{\tau}$ is the value of delay which maximizes expression (3-16).

We may extend this analysis further to estimate the delay rate, $\dot{\tau}$. For the case where $|\dot{\tau}| \ll 1$, which is applicable to all VLBI work, an observation may be broken into many time-segments such that the change of delay during each segment is small. Expression (3-12) may then be evaluated for each of these segments and then summed over the observation, i.e. if t_k is the time at the middle of the k^{th} data segment the estimates of τ , $\dot{\tau}$ and ϕ may be found by evaluating

$$\max_{\tilde{\tau}, \dot{\tilde{\tau}}, \tilde{\phi}} \left[\sum_{k=1}^K \sum_{j=1}^N \text{Re} \{ X_{1j} \cdot X_{2j}^* e^{-i [\omega_j (\tilde{\tau} + \dot{\tilde{\tau}} t_k) + \tilde{\phi}]} \} \right] \quad (3-18)$$

where $\tilde{\tau}$, $\dot{\tilde{\tau}}$, and $\tilde{\phi}$ are the "trial" parameters, K is the total number of data segments in the observation, and X_{1j} , X_{2j} are, by implication, functions of k . Following the same argument that led to Expressions (3-16) and (3-17), we find that the delay and delay rate estimates,

designated $\hat{\tau}$ and $\hat{\tau}$ respectively, may first be found by evaluating

$$\max_{\tilde{\tau}, \tilde{\tau}} \left| \left\{ \sum_{k=1}^K \sum_{j=1}^N X_{1j} \cdot X_{2j} e^{-i\omega_j (\tilde{\tau} + \tilde{\tau} t_k)} \right\} \right|$$

The fringe phase estimate, $\hat{\phi}$, may then be found in a similar manner as given by expression (3-17):

$$\tan \hat{\phi} = \frac{\operatorname{Im} \left\{ \sum_{k=1}^K \sum_{j=1}^N X_{1j} \cdot X_{2j}^* e^{-i\omega_j (\hat{\tau} + \hat{\tau} t_k)} \right\}}{\operatorname{Re} \left\{ \sum_{k=1}^K \sum_{j=1}^N X_{1j} \cdot X_{2j}^* e^{-i\omega_j (\hat{\tau} + \hat{\tau} t_k)} \right\}} \quad (3-20)$$

where τ and $\hat{\tau}$ have been found from Expression (3-19).

3.2-3 Maximum Likelihood Estimate of Observables: Time Domain Analysis²

Consider a simple model similar to that of Figure 3-1, except generalized to K stations, each with associated delay τ_k , where τ_k may be a slowly varying function of time (i.e. $\dot{\tau}_k \ll 1$). Let $x_k(t)$ be the signal observed at site k at time t, where $x_k(t)$ is composed of a "signal" component $s(t)$ and a "noise" component $n_k(t)$ such that

$$x_k(t) = s(t - \tau_k) + n_k(t)$$

The noise signals $n_k(t)$ are assumed to be non-zero Gaussian and statistically independent of $s(t)$ and of $n_l(t)$ for $l \neq k$.

On the basis of $x_k(t)$ from all of the stations we wish to make a maximum-likelihood (ML) estimate of some parameter set $\vec{\alpha}$ which is deterministically related to a set of model

²This analysis was originally outlined by I. I. Shapiro, private communication.

delays $\tilde{\tau}_k(\vec{\alpha})$. For the case of additive Gaussian noise it is easy to show that the ML estimate is equivalent to the "least mean-square error" estimate.³ We are therefore led to define the "mean-square error" function $Q(\vec{\alpha})$

$$Q(\vec{\alpha}) \equiv \sum_{k=1}^K \int_0^T [x_k(t) - s(t - \tilde{\tau}_k(\vec{\alpha}))]^2 dt$$

where K is the number of observing stations and $0 < t < T$ is the period over which the observation is conducted. In order to find the least-mean-square (or ML) estimate of $\vec{\alpha}$, we simply minimize $Q(\vec{\alpha})$ over the parameter set $\vec{\alpha}$. For convenience, let us define $x_k(t) = s(t - \tau_k) \equiv 0$ outside of the interval $0 < t < T$ so that we may rewrite $Q(\vec{\alpha})$

³Consider a k -component signal vector $\underline{x}(\underline{a}, t)$ with zero-mean additive Gaussian noise \underline{n} such that the noisy received signal is

$$\underline{y} = \underline{x}(\underline{a}, t) + \underline{n}$$

The probability density of \underline{y} is simply

$$p[\underline{y}|\underline{x}(\underline{a}, t)] = \frac{1}{\sqrt{(2\pi)^k |\underline{M}|}} \exp \left\{ -\frac{1}{2} (\underline{y}^T - \underline{x}^T) \underline{M}^{-1} (\underline{y} - \underline{x}) \right\}$$

where $\underline{M} = \underline{n} \underline{n}^T$ is the covariance matrix of the noise. The ML estimator maximizes $p[\underline{y}|\underline{x}(\underline{a}, t)]$ over $\underline{x}(\underline{a}, t)$, which is equivalent to minimizing the exponent $(\underline{y}^T - \underline{x}^T) \underline{M}^{-1} (\underline{y} - \underline{x})$. But this is equivalent to minimizing the mean-square error.

as

$$Q(\vec{\alpha}) = \sum_{k=1}^K \int_{-\infty}^{\infty} [x_k(t) - s(t - \tilde{\tau}_k)] dt \quad (3-21)$$

where, for brevity, we have dropped the explicit dependence of $\tilde{\tau}_k$ on $\vec{\alpha}$. Assuming that $\tilde{\tau}_k$ is a slowly varying function of time (i.e., $|\dot{\tilde{\tau}}_k| \ll 1$), we may make a simple change of variables to rewrite Equation (3-21)

as

$$Q(\vec{\alpha}) \approx \sum_{k=1}^K \int_{-\infty}^{\infty} [x_k(t + \tilde{\tau}_k) - s(t)]^2 dt$$

$$\approx \sum_{k=1}^K \int_{-\infty}^{\infty} x_k^2(t + \tilde{\tau}_k) dt \quad (3-22)$$

$$+ K \left[\int_{-\infty}^{\infty} s^2(t) dt - \frac{2}{K} \int_{-\infty}^{\infty} \sum_{k=1}^K x_k(t + \tilde{\tau}_k) s(t) dt \right]$$

With another change of variables the first term of expression (3-22) may be written

$$\sum_{k=1}^K \int_{-\infty}^{\infty} x_k(t + \tilde{\tau}_k) dt = \sum_{k=1}^K \int_{-\infty}^{\infty} x_k(t) dt \quad (3-23)$$

But the right-hand side of Equation (3-23) is independent of $\vec{\alpha}$ and so is not involved in the minimization of $Q(\vec{\alpha})$. We are therefore left with minimizing only over the second term of expression (3-22) which, upon completion of the

square, reduces the least mean-square criterion to

$$\min_{\vec{\alpha}} \left\{ K \int_{-\infty}^{\infty} \left[s(t) - \frac{1}{K} \sum_{k=1}^K x_k(t + \tilde{\tau}_k) \right]^2 dt - \frac{1}{K} \int_{-\infty}^{\infty} \left[\sum_{k=1}^K x_k(t + \tilde{\tau}_k) \right]^2 dt \right\} \quad (3-24)$$

The value of the first integral of this expression is always ≥ 0 and may be minimized to zero by setting

$$s(t) = \frac{1}{K} \sum_{k=1}^K x_k(t + \tilde{\tau}_k) \quad (3-25)$$

But this is just the least-mean-square estimate of $s(t)$ which will result after the minimization of the entire expression (3-24). Clearly, then, the minimization of (3-24) reduces simply to the minimization of the second integral of (3-24) which is equivalent to

$$\max_{\vec{\alpha}} \int_{-\infty}^{\infty} \left[\sum_{k=1}^K x_k(t + \tilde{\tau}_k) \right]^2 dt \quad (3-26)$$

which may be rewritten as

$$\max_{\vec{\alpha}} \left\{ \sum_{k=1}^K \int_{-\infty}^{\infty} x_k^2(t + \tilde{\tau}_k) dt + \sum_{k \neq \ell} \int_{-\infty}^{\infty} x_k(t + \tilde{\tau}_k) x_\ell(t + \tilde{\tau}_\ell) dt \right\} \quad (3-27)$$

We may change variables to show that the first term of (3-27) is independent of $\vec{\alpha}$ so that the least-squares (or ML) minimization criterion finally reduces to

$$\max_{\vec{\alpha}} \left[\sum_{k \neq l} \int_{-\infty}^{\infty} x_k(t + \tilde{\tau}_k) x_l(t + \tilde{\tau}_l) dt \right] \quad (3-28)$$

For a simple 2-station interferometer where we define the delay to the reference station 1 to be zero (i.e. $\tau_1 \equiv 0$), expression (3-28) reduces to the familiar cross-correlation

$$\max_{\vec{\alpha}} \left[\int_{-\infty}^{\infty} x_1(t) x_2[t + \tilde{\tau}(\vec{\alpha})] dt \right] \quad (3-29)$$

where $\tilde{\tau}(\vec{\alpha})$ is the model delay to the "remote" station 2.

We now wish to show that expression (3-29) is equivalent to the expression (3-12) of Section 3.2-2. By a simple change of variables we may rewrite (3-29) as

$$\max_{\vec{\alpha}} \left[\int_{-\infty}^{\infty} x_1(t - \tilde{\tau}) x_2(t) dt \right] \quad (3-30)$$

If we define $X_1(\omega)$ and $X_2(\omega)$ to be the Fourier transforms of $x_1(t)$ and $x_2(t)$, respectively, we may rewrite (3-30) as

$$\begin{aligned}
 & \max_{\vec{\alpha}} \over \left[\int_{-\infty}^{\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(\omega) e^{i\omega(t-\tau)} d\omega \right] X_2(t) dt \\
 &= \max_{\vec{\alpha}} \over \left[\int_{-\infty}^{\infty} \frac{1}{2\pi} \left\{ \int_{-\infty}^{\infty} X_2(t) e^{i\omega t} dt \right\} e^{-i\omega\tilde{\tau}} X_1(\omega) d\omega \right] \\
 &= \max_{\vec{\alpha}} \over \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(\omega) X_2^*(\omega) e^{-i\omega\tilde{\tau}} d\omega \right] \quad (3-31)
 \end{aligned}$$

For real $x_1(t)$ and $x_2(t)$, $X_1(-\omega) = X_1^*(\omega)$ and $X_2(-\omega) = X_2^*(\omega)$, so that expression (3-31) is equivalent to

$$\begin{aligned}
 & \max_{\vec{\alpha}} \over \left[\int_0^{\infty} [X_1(\omega) X_2^*(\omega) e^{-i\omega\tilde{\tau}} + X_1^*(\omega) X_2(\omega) e^{+i\omega\tilde{\tau}}] d\omega \right] \\
 &= \max_{\vec{\alpha}} \over \left[2\text{Re} \left\{ \int_0^{\infty} X_1(\omega) X_2^*(\omega) e^{-i\omega\tilde{\tau}} d\omega \right\} \right] \quad (3-31a)
 \end{aligned}$$

which is equivalent to Equation (3-12).

We note that the analysis leading to Expression (3-29) places no restriction upon the parameter set $\vec{\alpha}$. In particular, $\vec{\alpha}$ may include time, so that in the case where $\tau(\vec{\alpha})$ may be approximated by a constant delay, τ , plus a delay rate, $\dot{\tau}$, Expression (3-29) may be simplified to

$$\max_{\vec{\tau}, \dot{\tau}} \over \int_{-\infty}^{\infty} x_1(t) x_2[t + (\tilde{\tau} + \dot{\tau}t)] dt \quad (3-31b)$$

3.3 SIGNAL-TO-NOISE ANALYSIS

The signal received at the output of each receiver consists of a "signal" component due to the source being observed, and a "noise" component due to all other causes (see Figure 3-1). Generally, the "signal" and "noise" components are independent and have Gaussian statistics. The outputs of the two receivers can be written in the frequency domain as⁴

$$X_1(\omega) = \sqrt{T_{a1}} S(\omega) + \sqrt{T_{s1}} N_1(\omega) \quad (3-32)$$

$$X_2(\omega) = \sqrt{T_{a2}} S'(\omega) + \sqrt{T_{s2}} N_2(\omega) \quad (3-33)$$

where (see Equation 3-13)

$$S'(\omega) = S(\omega) e^{-i[\omega(\tau+i t)+\phi]} \quad (3-34)$$

and where T_{a1} , T_{a2} , T_{s1} , T_{s2} are the antenna temperatures (due to the source) and the system noise temperatures (due mostly to receiver noise) at the two sites, and where the complex quantities $S_1(\omega)$, $N_1(\omega)$, and $N_2(\omega)$ are independent and have independent real and imaginary parts. The real and imaginary parts are all assumed to have a variance of 1/2 with identical probability density functions of the form

⁴In this analysis we shall consider $S(\omega)$ to be only the portion of the signals $X_1(\omega)$ and $X_2(\omega)$ which are correlated. Any actual lack of correlation due to a fringe visibility less than 1 may be modelled simply by including the uncorrelated portions of the signals in the independent noises $N_1(\omega)$ and $N_2(\omega)$. We assume that the radiation from the source is spatially incoherent.

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{z^2}{2\sigma^2} \right]; \quad \sigma^2 = \frac{1}{2} \quad (3-35)$$

It follows then that $X_1(\omega)$ and $X_2(\omega)$ are also Gaussian with independent real and imaginary components.

We now wish to examine the statistics of the estimates of "correlation amplitude" and "fringe phase" so that we may develop expressions for expected signal-to-noise ratios and fringe-phase errors. Assume that expression (3-16) of Section 3.2-2 has been maximized over $\tilde{\tau}$ to find the estimates, $\hat{\tau}$ and $\hat{\dot{\tau}}$, of delay and delay rate, respectively. As indicated by Equation (3-15), the estimate, $\hat{\phi}$, of fringe phase is made simply by computing the phase angle of the quantity

$$\sum_j X_1(\omega_j) X_2^*(\omega_j) e^{-i\omega_j(\hat{\tau} + \hat{\dot{\tau}}t)} \quad (3-36)$$

Later it will be convenient to normalize Equation (3-36) by the product of the observation bandwidth and length, BT . Accordingly, let us define the vector \vec{R}

$$\vec{R} \equiv \frac{1}{BT} \sum_j X_1(\omega_j) X_2^*(\omega_j) e^{-i\omega_j(\hat{\tau} + \hat{\dot{\tau}}t)} \quad (3-37)$$

The amplitude of \vec{R} , when properly normalized to the total received power, is just the "correlation amplitude estimate". It is of interest to examine the statistics of both the amplitude and phase of \vec{R} .

As the first step in examining the statistics of \vec{R} , we shall assume that the estimates $\hat{\tau}$ and $\hat{\dot{\tau}}$ are close enough to τ and $\dot{\tau}$, respectively, that $|\omega(\tau - \hat{\tau})| \ll 1$ and

$|\omega(\hat{\tau}-\hat{\tau})t| \ll 1$ for all values of the radio frequency ω_j included in the summation of (3-37) and for all values of t included in the observation, $0 < t < T$. We are led to define a "rotated cross-spectrum" $S_{12}(\omega)$

$$S_{12}(\omega) \equiv X_1(\omega)X_2^*(\omega) e^{-i\omega(\tau+\dot{t}t)} \quad (3-38)$$

which, using Equations (3-32), (3-33), and (3-34) becomes

$$\begin{aligned} S_{12}(\omega) = & \sqrt{T_{a1}T_{a2}} S(\omega)S^*(\omega)e^{i\phi_0} + \sqrt{T_{s1}T_{a2}} N_1(\omega)S^*(\omega)e^{i\phi_0} \\ & + \sqrt{T_{a1}T_{s2}} S(\omega)N_2'^*(\omega) + \sqrt{T_{s1}T_{s2}} N_1(\omega)N_2''^*(\omega) \end{aligned} \quad (3-39)$$

where

$$N_2'(\omega) = N_2(\omega)e^{i\omega(\tau+\dot{t}t)} \quad (3-40)$$

and where we shall assume that $N_2'(\omega)$ has the same statistics as $N_2(\omega)$, a generally good assumption for broadband receiver noise.

We now wish to show that all of the terms comprising $S_{12}(\omega)$ are independent and that, furthermore, the real and imaginary parts of each term are independent. Since $S(\omega)$, $N_1(\omega)$, $N_2'(\omega)$ are Gaussian, it is sufficient to show that the terms of the right-hand side of Equation (3-39) are uncorrelated (see, for example, Davenport and Root, Reference 27). This is easily shown by using the expansion of the expectation of the products of four zero-mean Gaussian variables (see again Reference 27)

$$\langle x_1x_2x_3x_4 \rangle = \langle x_1x_2 \rangle \langle x_3x_4 \rangle + \langle x_1x_3 \rangle \langle x_2x_4 \rangle + \langle x_1x_4 \rangle \langle x_2x_3 \rangle$$

Hence, for example, the correlation of the first and last terms of Equation (3-39) is

$$\begin{aligned} \langle SS^*N_1N_2'^* \rangle &= \langle SS^* \rangle \langle N_1N_2'^* \rangle + \langle SN_1 \rangle \langle S^*N_2'^* \rangle \\ &+ \langle SN_2'^* \rangle \langle S^*N_1 \rangle = 0 \end{aligned} \quad (3-42)$$

This result shows that the first and last terms of Equation (3-39) are uncorrelated and hence statistically independent. It is easy to show, in the same manner, that all of the terms of Equation (3-39) are independent. Furthermore, since $S_{12}(\omega)$ is generally evaluated as the sum over a large number of independent short time intervals spanning an observation [see Section 3.2-2 and Appendix F], we may use the Central Limit Theorem to infer that each of the terms of Equation (3-39) is approximately Gaussianly distributed.

We can now proceed to evaluate the statistics of \vec{R} of Equation (3-37). Substituting Equations (3-38) and (3-39) into (3-37), we have

$$\begin{aligned} \vec{R} &= \frac{1}{BT} [\sqrt{T_{a1}T_{a2}} e^{i\phi} \sum_j S(\omega_j)S^*(\omega_j) \\ &+ \sqrt{T_{s1}T_{a2}} e^{i\phi} \sum_j N_1(\omega_j)S^*(\omega_j) + \sqrt{T_{a1}T_{s2}} \sum_j S(\omega_j)N_2'^*(\omega_j) \\ &+ \sqrt{T_{s1}T_{s2}} \sum_j N_1(\omega_j)N_2'^*(\omega_j)] \end{aligned} \quad (3-43)$$

where the summation over ω_j is implied to be over the independent points in the frequency spectra $S_1(\omega)$, $N_1(\omega)$, and $N_2'(\omega)$. Since we have shown that the terms of $S_{12}(\omega)$

of Equation (3-39) are independent, it follows that the terms of \vec{R} of Equation (3-43) are also independent and we may examine the statistics of \vec{R} term-by-term.

Let us first examine the statistics of the first term of \vec{R} . This is the so-called "signal" term, and we see that it is a vector with a phase angle equal to the fringe phase, ϕ_0 . For an observation of length T and spanning a bandwidth B (assumed uniform), there are BT independent Fourier components in $S_1(\omega)$. Recalling that we defined the real and imaginary parts of $S_1(\omega)$ to have a variance of $\frac{1}{2}$ for all values of ω_j , the mean of the first term of \vec{R} is just $\sqrt{T_{a1}T_{a2}} e^{i\phi_0}$. The variance is easily found by using the expansion of Equation (3-41) to be $T_{a1}T_{a2}/BT$. It is convenient to define a "mean signal vector" \vec{S} with amplitude

$$|\vec{S}| = \sqrt{T_{a1}T_{a2}} \quad (3-44)$$

and a noise vector \vec{N}_s which is aligned with \vec{S} and whose amplitude has a standard deviation of

$$\sigma_{|\vec{N}_s|} = \sqrt{\frac{T_{a1}T_{a2}}{BT}} \quad (3-45)$$

It should be noted that \vec{N}_s is not receiver noise, but is simply "self-noise" due to power variations in the received signal. Furthermore, \vec{N}_s is always aligned in the same direction as \vec{S} , namely at the angle corresponding to the fringe phase, ϕ_0 .

The last three terms of Equation (3-43) are due to receiver noise and, as we have already established, are independent of the first or "signal" term. We now wish to show that not only are these three terms independent of each other, but that their real and imaginary parts are also independent. This independence will allow us to simply sum the variances of the real and imaginary parts of these terms to determine their overall contribution to \hat{R} . Consider, for example, the second term of Equation (3-43) involving the product $N_1(\omega)S^*(\omega)$. Separate $N_1(\omega)$ and $S^*(\omega)$ into their real and imaginary parts

$$\begin{aligned} N_1(\omega) &= n_r + in_i \\ S^*(\omega) &= s_r - is_i \end{aligned} \tag{3-46}$$

and form the product

$$N_1(\omega)S^*(\omega) = (n_r s_r + n_i s_i) + i(n_i s_r - n_r s_i). \tag{3-47}$$

Then the correlation between the real and imaginary parts is

$$\begin{aligned} \langle \text{Re}[N_1(\omega)S^*(\omega)] \text{Im}[N_1(\omega)S^*(\omega)] \rangle &= \langle n_r s_r n_i s_i \rangle - \\ &\quad \langle n_r s_r n_r s_i \rangle + \langle n_i s_i n_i s_r \rangle - \langle n_i s_i n_r s_i \rangle \end{aligned} \tag{3-48}$$

It is easy to show that each one of these four terms is identically zero by the expansion of the expectation of the products of four zero-mean Gaussian variables as given in

Equation (3-41). Therefore we have

$$\left\langle \text{Re}[N_1(\omega)S^*(\omega)] \text{Im}[N_1(\omega)S^*(\omega)] \right\rangle = 0 \quad (3-49)$$

implying, because all their components are Gaussian, that $\text{Re}[N_1(\omega)S^*(\omega)]$ and $\text{Im}[N_1(\omega)S^*(\omega)]$ are independent. Similar arguments can be used to show that the real and imaginary parts of the third and fourth terms of Equation (3-43) are independent.

We have succeeded, then, in showing that not only are the last three terms of Equation (3-43) independent, but that their real and imaginary components are independent as well. This result allows us to compute the variance of the last three terms simply by summing the variances of the individual terms. For an observation of bandwidth B and duration T there are again BT independent Fourier components; recalling that we defined the variance of the real and imaginary parts of $S(\omega)$, $N_1(\omega)$, and $N_2(\omega)$ to be $\frac{1}{2}$, we can easily compute the variance of the real part and the imaginary part of the last three terms of Equation (3-43) to be

$$\frac{(T_{s1}T_{a2} + T_{a1}T_{s2} + T_{s1}T_{s2})}{2BT} \quad (3-50)$$

It is easily shown (see Section 3.4) that a complex vector with independent, zero-mean Gaussian components with a common variance σ^2 leads to a Rayleigh distribution of the vector amplitude with variance $(2 - \frac{\pi}{2})\sigma^2$ and a uniform distribution in phase. The variance of the amplitude of

the noise contributed by the last three terms of Equation (3-43) is then simply $(2 - \frac{\pi}{2})$ times expression (3-50) and we are led to define a "receiver noise vector" \vec{N}_n whose amplitude has a standard deviation given by

$$\sigma_{|\vec{N}_n|} = \sqrt{\frac{T_{s1}T_{a2} + T_{a1}T_{s2} + T_{s1}T_{s2}}{2BT}} \left(2 - \frac{\pi}{2}\right) \quad (3-51)$$

and whose phase is uniformly distributed.

It is useful to construct the vector diagram of the real and imaginary parts of \vec{R} as shown in Figure 3-2. The vector \vec{S} is the "signal" component of amplitude $\sqrt{T_{a1}T_{a2}}$; its phase is the fringe phase, ϕ_0 . The vector \vec{N}_s is the "self-noise" of the signal and is aligned with \vec{S} , has zero-mean amplitude and standard deviation

$$\sqrt{\frac{T_{a1}T_{a2}}{BT}}. \quad \text{The vector } \vec{N}_n \text{ is the noise vector arising from}$$

the last three terms of Equation (3-43). It is uniformly distributed in phase and has a Rayleigh amplitude distribution with a standard deviation given by Equation (3-51).

Now that we have the statistics of the components of \vec{R} , we can compute the statistics of \vec{R} itself. For the general case this is very difficult, but fortunately most VLBI observations are made under circumstances where we may make simplifying assumptions. For almost all practical VLBI systems $T_a \ll T_s$ so that $\sigma_{|\vec{N}_s|} \ll \sigma_{|\vec{N}_n|}$, and the product BT is chosen large enough so that $\sigma_{|\vec{N}_n|} \ll |\vec{S}|$. In these

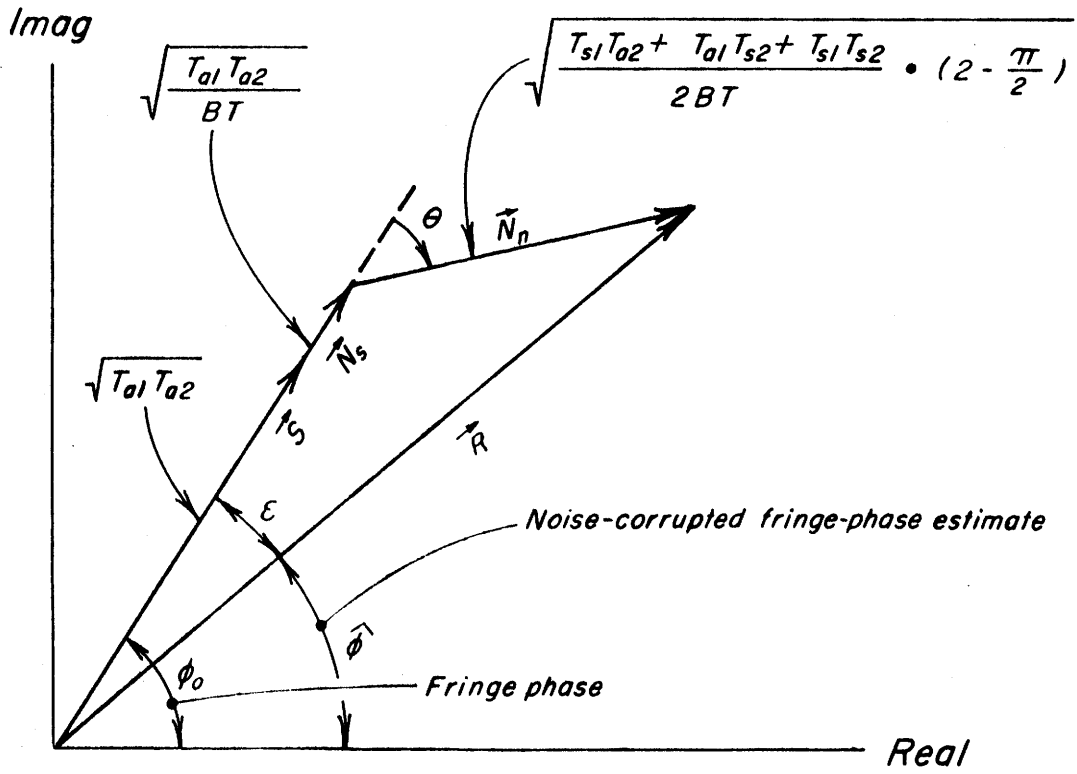


Figure 3-2

Graphic representation of
the signal and noise relations

circumstances we may neglect \vec{N}_s and consider only the effects of \vec{N}_n , which has a mean amplitude small with respect to the mean of $|\vec{S}|$. Define the phase error, ϵ , as shown in Figure 3-2. If we neglect \vec{N}_s , the phase error due to a small \vec{N}_n is approximately

$$\epsilon \approx \frac{\vec{N}_n \cdot \hat{i}_\phi}{|\vec{S}|} \quad (3-52)$$

where \hat{i}_ϕ is a unit vector perpendicular to \vec{S} and $\vec{N}_n \cdot \hat{i}_\phi$ is then the component of \vec{N}_n perpendicular to \vec{S} (see Figure 3-2). But the variance of a component of \vec{N}_n along any direction is the same as along the real or imaginary parts as given by expression (3-50). Therefore the variance of ϵ , which is identical to the variance of $\hat{\phi}$ (see Figure 3-2), is given by

(3-53)

$$\sigma_{\hat{\phi}}^2 = \sigma_{\epsilon}^2 \approx \frac{\langle \vec{N}_n \cdot \hat{i}_\phi \rangle}{|\vec{S}|} \approx \frac{T_{s1}T_{a2} + T_{a1}T_{s2} + T_{s1}T_{s2}}{T_{a1}T_{a2}} \cdot \frac{1}{2BT}$$

which, for the common case where $T_{a1} \ll T_{s1}$ and $T_{a2} \ll T_{s2}$, reduces simply to

$$\sigma_{\hat{\phi}}^2 \approx \frac{T_{s1}T_{s2}}{T_{a1}T_{a2}} \cdot \frac{1}{2BT} \quad (3-54)$$

It is convenient, at this point, to define a signal-to-noise ratio as

$$\text{SNR} \equiv \rho_0 \sqrt{2BT} \quad (3-55)$$

where ρ_0 is the "normalized correlation amplitude", defined in the general case by the ratio of the geometric means of the antenna temperatures and the total temperatures⁵

$$\rho_0 \equiv \sqrt{\frac{T_{a1} T_{a2}}{(T_{a1} + T_{s1})(T_{a2} + T_{s2})}} \quad (3-56)$$

Using Equation (3-56) in Equation (3-55) and referring to Figure 3-2, we see that SNR corresponds closely to the ratio of $|\vec{S}|$ to $|\vec{N}_n|$, so that SNR does have a "physical" meaning. For the case where $T_{a1} \ll T_{s1}$ and $T_{a2} \ll T_{s2}$, ρ_0 becomes

$$\rho_0 \approx \sqrt{\frac{T_{a1} T_{a2}}{T_{s1} T_{s2}}} \quad (3-57)$$

Using Equations (3-55) and (3-57) in Equation (3-54), we have the simple result

$$\sigma_{\hat{\phi}} \approx \frac{1}{\text{SNR}} \quad (3-58)$$

Let us proceed now to compute the variance of the amplitude of \vec{R} . For the case of small ϵ , $|\vec{R}|$ may be approximated by

$$|\vec{R}| \approx |\vec{S}| + |\vec{N}_s| + |\hat{S} \cdot \vec{N}_n| \quad (3-59)$$

where \hat{S} is a unit vector in the direction of \vec{S} .

Since \vec{S} is a constant vector and $|\vec{N}_s|$ and $|\hat{S} \cdot \vec{N}_n|$ are

⁵The definition in Equation (3-56) is strictly true only for a fringe visibility, γ , equal to 1 (i.e. a completely unresolved source). For $\gamma < 1$, the right-hand side of Equation (3-56) must be multiplied by γ .

zero-mean independent random variables, the variance of $|\vec{R}|$ is simply

$$\sigma^2_{|\vec{R}|} \approx \sigma^2_{|\vec{N}_s|} + \sigma^2_{|\hat{\vec{S}} \cdot \vec{N}_n|} \quad (3-60)$$

The expression for $\sigma^2_{|\vec{N}_s|}$ is given by Equation (3-45).

The variance of $|\hat{\vec{S}} \cdot \vec{N}_n|$ is simply the variance of the component of \vec{N}_n which lies in the direction of \vec{S} . But the variance of $|\hat{\vec{S}} \cdot \vec{N}_n|$ is the same along any given direction, and in particular is the same as along the real or imaginary parts of \vec{N}_n as given by expression (3-50). Therefore

$$\sigma^2_{|\vec{R}|} \approx \frac{(T_{a1} + T_{s1})(T_{a2} + T_{s2})}{BT} \quad (3-61)$$

A useful quantity is the ratio of $\sigma_{|\vec{R}|}$ to $|\vec{S}|$, which might be called the "normalized standard deviation" of $|\vec{R}|$. From Equations (3-44) and (3-61)

$$\frac{\sigma_{|\vec{R}|}}{|\vec{S}|} \approx \sqrt{\frac{(T_{a1} + T_{s1})(T_{a2} + T_{s2})}{BT (T_{a1} T_{a2})}} \quad (3-62)$$

or, in terms of SNR as defined by Equations (3-55) and (3-56),

$$\frac{\sigma_{|\vec{R}|}}{|\vec{S}|} \approx \frac{\sqrt{2}}{\text{SNR}} \quad (3-63)$$

An approximate lower limit of signal detectability is set when this quantity becomes equal to unity, or approximately when $\text{SNR} \approx \sqrt{2}$. In practice, the minimum

useful SNR is somewhat larger because, in general, the signal is found by searching over a large number of trial delays and delay rates, thus increasing the probability that a noise "spike" may be larger than the actual signal (see Section 3.7). For typical experimental conditions where $B \approx 360$ kHz and $T \approx 180$ seconds, the minimum useful SNR is ≈ 10 , which corresponds to a correlation amplitude of $\rho_0 \approx 0.001$.

3.4 CALCULATION OF THE PROBABILITY DENSITY FUNCTION CORRELATION AMPLITUDE AND FRINGE PHASE

3.4-1 Probability Density Functions for $T_a \ll T_s$

For the usual case where $T_a \ll T_s$ (i.e. where the system temperature greatly exceeds the source temperature), it is instructive to compute the actual probability density functions for the amplitude and phase of \vec{R} . In such a case we can neglect the self-noise \vec{N}_s (see Figure 3-2) and consider only the signal \vec{S} and noise \vec{N}_n . As we have shown in Section 3.3, the real and imaginary parts of \vec{N}_n are independent and Gaussianly distributed and each has a variance given by Equation (3-50). Figure 3-3 illustrates the situation and defines the quantities $a, b, \phi, \phi_0, \rho_0, \rho, x, y,$ and σ . The probability densities of x and y about their respective means are independent and are given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (3-64)$$

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-b)^2}{2\sigma^2}} \quad (3-65)$$

and

$$p(x, y) = p(x)p(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-a)^2 + (y-b)^2}{2\sigma^2}} \quad (3-66)$$

Using the law of cosines we have

$$|\vec{N}_n|^2 = (x-a)^2 + (y-b)^2 = \rho_0^2 + \rho^2 - 2\rho_0\rho\cos(\phi_0 - \phi). \quad (3-67)$$

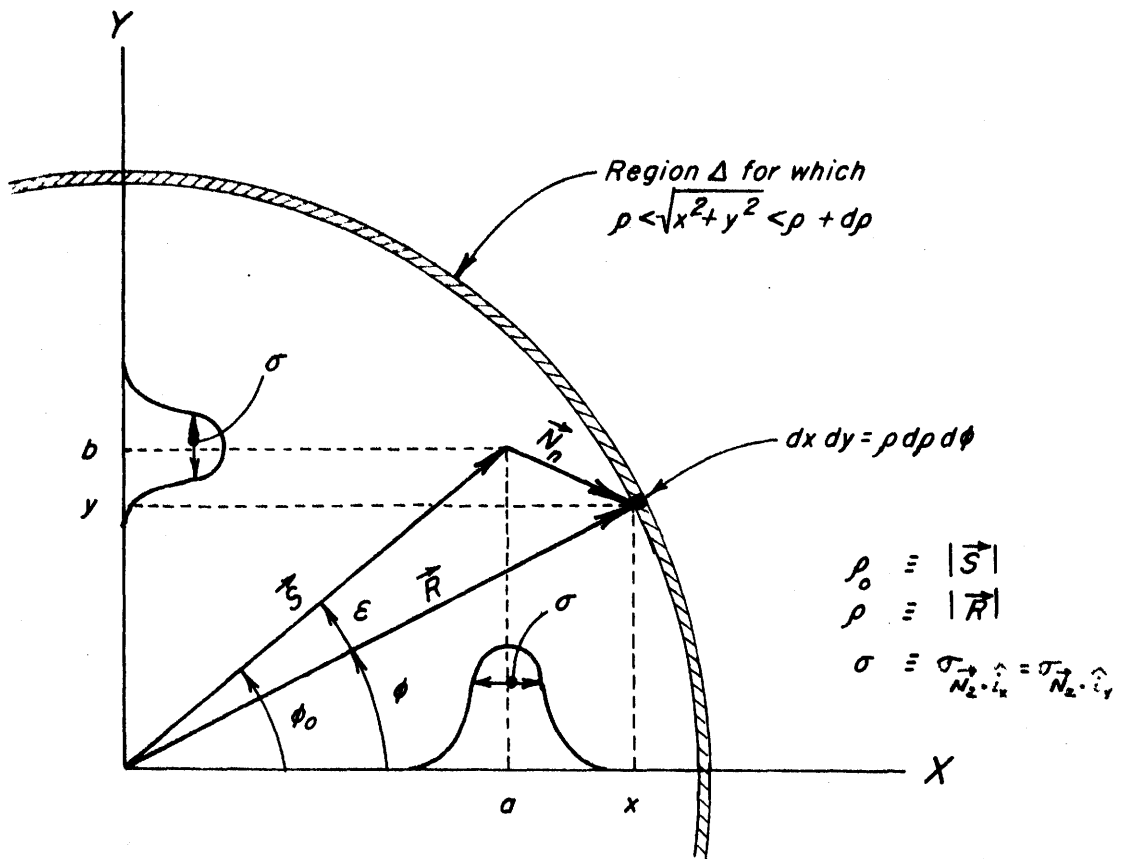


Figure 3-3

Vector diagram of signal and noise vectors for derivation of amplitude and phase distribution of \vec{R} .

The region of the x - y plane is such that

$$\rho < \sqrt{x^2+y^2} \leq \rho+d\rho \quad ; \quad \rho > 0 \quad (3-68)$$

is a circular ring with inner radius ρ and outer radius $\rho + d\rho$, shown as area Δ in Figure 3-3. With $x = \rho \cos \phi$, $y = \rho \sin \phi$, we have $dx dy = \rho d\rho d\phi$. Therefore, given ρ_0 and ϕ_0 , the probability that ρ falls between ρ and $\rho+d\rho$ is

$$\begin{aligned} p(\rho|\rho_0, \phi_0) d\rho &= \iint_{\Delta} p(x,y) dx dy \\ &= \frac{1}{2\pi\sigma^2} \int_{\phi=0}^{2\pi} e^{-\frac{1}{2\sigma^2}[\rho_0^2 + \rho^2 - 2\rho_0\rho \cos(\phi_0 - \phi)]} \rho d\rho d\phi \\ &= \frac{\rho d\rho}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[\rho_0^2 + \rho^2]} \int_0^{2\pi} e^{\frac{1}{\sigma^2} \rho_0 \rho \cos(\phi' - \phi)} d\phi' \quad (3-69) \end{aligned}$$

where we have made the change of variables $\phi' = \phi - \phi_0$.

The integral over ϕ' can be expressed in terms of the modified Bessel function of order zero (see Papoulis, p. 196, Reference 28):

$$I_0(\alpha) = \frac{1}{2\pi} \int_0^{2\pi} e^{\alpha \cos \theta} d\theta = \sum_{n=0}^{\infty} \frac{\alpha^{2n}}{2^{2n} (n!)^2} \quad (3-70)$$

We note that $p(\rho|\rho_0, \phi_0)$ is completely independent of ϕ_0 , so that the desired result is simply

$$p(\rho|\rho_0) = \frac{\rho}{\sigma^2} e^{-\frac{1}{2\sigma^2}(\rho_0^2 + \rho^2)} I_0\left(\frac{\rho_0\rho}{\sigma^2}\right); \quad \rho > 0 \quad (3-71)$$

For small values of x we see by inspection of Equation (3-70) that $I_0(x)$ may be approximated by

$$I_0(x) \approx 1 \quad (3-72)$$

so that for $\rho_0 = 0$, the distribution of Equation (3-71) becomes Rayleigh

$$p(\rho|\rho_0=0) = \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}}; \quad \rho > 0 \quad (3-73)$$

with (see Papoulis, p. 148, Reference 28)

$$\langle \rho \rangle = \sigma \sqrt{\frac{\pi}{2}}; \quad \sigma_\rho^2 = (2 - \frac{\pi}{2})\sigma^2 \quad (3-74)$$

For large values of its argument $I_0(\alpha)$ may be approximated by (Hildebrand, p. 151, Reference 29)

$$I_0(\alpha) \approx \frac{e^\alpha}{\sqrt{2\pi\alpha}} \quad (3-75)$$

so that Equation (3-71) becomes

$$\begin{aligned} p(\rho|\rho_0) &\approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\rho-\rho_0)^2} \sqrt{\frac{\rho}{\rho_0}} \\ &\approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\rho-\rho_0)^2}; \quad \rho > 0, \rho \approx \rho_0 \end{aligned}$$

The mean and variance are given by

$$\langle \rho \rangle = \rho_0 \quad ; \quad \sigma_\rho^2 = \sigma^2 \quad (3-77)$$

Figure 3-4 shows a plot of $I_0(x)$ as well as of $p(\rho|\rho_0)$ for several values of $\frac{\rho_0}{\sigma}$. We note that, for the case of large $\frac{\rho_0}{\sigma}$ (say, for $\frac{\rho_0}{\sigma} \gtrsim 5$), that

$$\frac{\rho_0}{\sigma} \approx \text{SNR} \quad (3-78)$$

where SNR is given by Equation (3-55).

The distribution of the phase angle ϕ in Figure 3-3 is a difficult function to calculate directly except in the cases $\rho_0 \approx 0$ and $\frac{\rho_0}{\sigma}$ large⁴. For the case $\rho_0 = 0$ (i.e. no signal), $\vec{R} = \vec{N}_n$ and the distribution of ϕ is uniform

$$p(\phi|0) = \frac{1}{2\pi} \quad ; \quad 0 < \phi \leq 2\pi \quad (3-79)$$

This result follows from Equation (3-66) by setting $a = 0$, $b = 0$, $x = \rho \cos \phi$, $y = \rho \sin \phi$. Thus, the ϕ dependence completely disappears, implying a uniform distribution. For the case of large signal-to-noise ratios,

⁴Papoulis (Reference 28, p. 501) gives an analytical form for the probability density of ϕ in the general case.

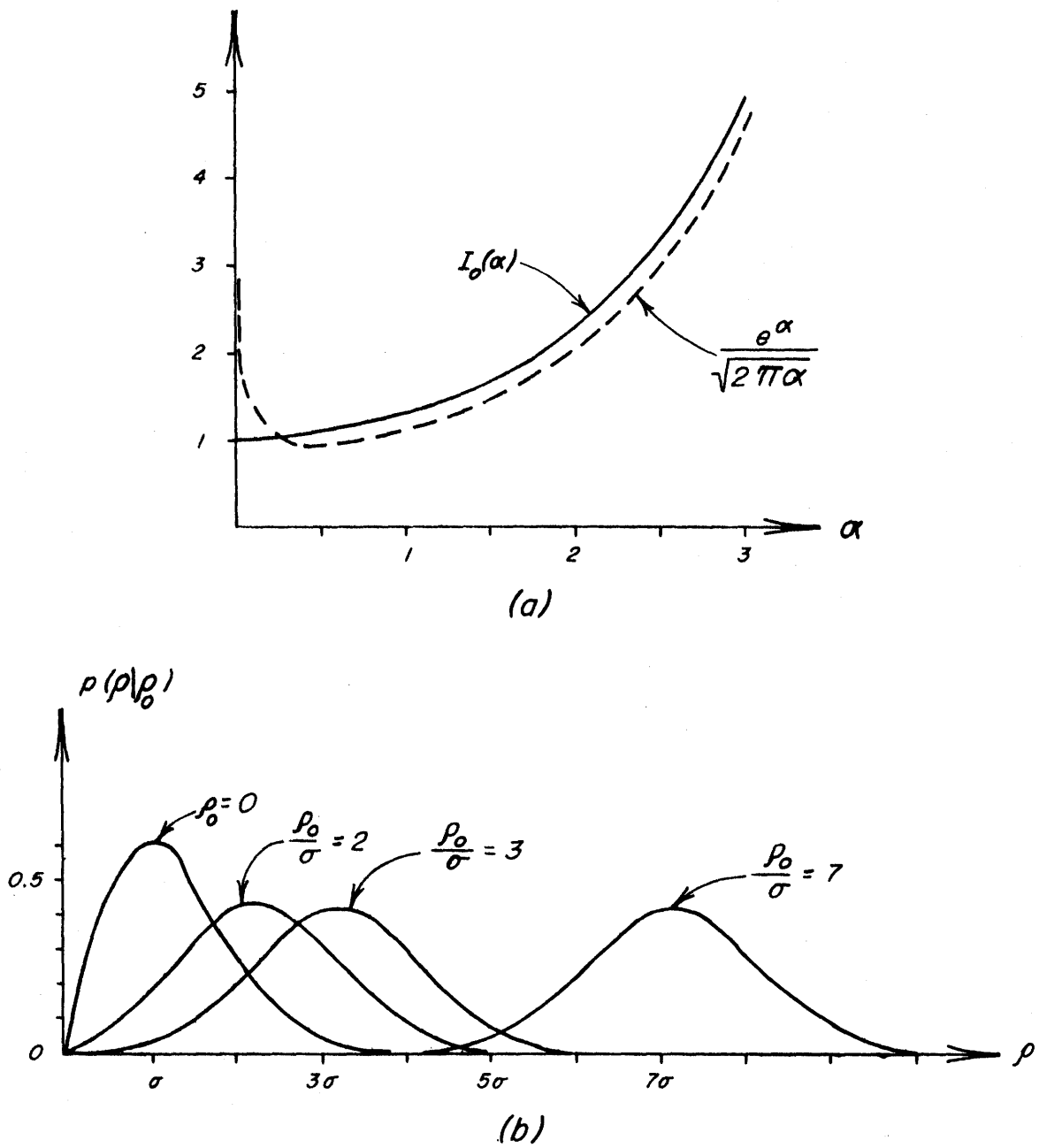


Figure 3-4

(a) $I(\alpha)$ and its approximation for large α as discussed in the text.

(b) Plot of probability density of amplitude p given a signal strength ρ_0 . See Eqn. (3-71).

(see Figure 3-3),

$$\epsilon \approx \frac{\vec{N}_n \cdot \hat{i}_\phi}{|\vec{S}|} \quad (3-80)$$

where \hat{i}_ϕ is a unit vector perpendicular to \vec{S} . Thus $\vec{N}_n \cdot \hat{i}_\phi$ is just the component of \vec{N}_n in the direction perpendicular to \vec{S} and has the same Gaussian distribution function as the x or y components of \vec{N}_n . This result also follows from Equation (3-66) by noting that it retains exactly the same form under a rotation of x, y about their mean. Therefore

$$p(\epsilon | \rho_0) \approx \frac{\rho_0}{\sqrt{2\pi}\sigma} e^{-\frac{\rho_0^2 \epsilon^2}{2\sigma^2}}; \quad \frac{\rho_0}{\sigma} \text{ large.} \quad (3-81)$$

3.4-2 Alternate Derivation of Probability Density for the "No-Signal" Case

It is illuminating to derive the no-signal amplitude distribution [see Equation (3-73)] in the absence of signal from another point of view, in particular, as a 2-dimensional random walk. Consider the case where the signal at each station is recorded as infinitely-clipped one-bit samples (this is the standard digital recording technique). The sampling period, ΔT_s , is generally chosen to satisfy the Nyquist criterion

$$\Delta T_s \approx \frac{1}{2B} \quad (3-82)$$

where B is the sampled (video) bandwidth. When this criterion is satisfied, each one-bit sample may be

considered independent of all others (see, for example, Papoulis, p. 370, Reference 28). When cross-correlation processing is done, the cross-power spectrum computation yields a vector whose two rectangular components have independent noises. Consider the no-signal case where K bit-pairs have been cross-correlated. Each component of the power spectrum may be considered to be the result of a one-dimensional random walk with K steps, where each step has equal probability of being positive or negative. Accordingly, let us define the random variable, z_i , which is the result of correlating one bit-pair which can have two equally likely outcomes

$$z_i = \begin{cases} +1/K & \text{(Matching bits)} \\ -1/K & \text{(Non-matching bits)} \end{cases} \quad (3-83)$$

$$\Pr(z_i = +1/K) = \Pr(z_i = -1/K) = 1/2$$

The outcome, x , of the 1-dimensional random walk is simply the sum of the increments z_i

$$x = \sum_{i=1}^K z_i \quad (3-84)$$

which has a range of $-1 < x < 1$, so that the value of x corresponds directly to the normalized amplitude of one component of the cross-power spectrum. Since all the z_i are independent, x has mean and variance

$$\begin{aligned} \langle x \rangle &= K \langle z_i \rangle = 0 \\ \sigma_x^2 &= K \sigma_z^2 = K \cdot \frac{1}{K^2} = \frac{1}{K} \end{aligned} \quad (3-85)$$

Since K is, in general, a very large number, we may apply the Central Limit Theorem to find the probability density function for x

$$p(x) \approx \sqrt{\frac{K}{2\pi}} e^{-\frac{Kx^2}{2}} \quad (3-86)$$

But this function is identical to the density of Equation (3-64) if we set $a = 0$ (no signal) and $\sigma = \sqrt{1/K}$. Similarly, a random-walk analysis of the other component, y , of the cross-spectrum yields an identical distribution. If, as before, we define the amplitude

$$\rho = \sqrt{x^2 + y^2} \quad (3-87)$$

then the density of ρ in the no-signal case is reduced to that of Equation (3-73) where we set $\sigma^2 = 1/K$. For an observation of length T , $K = T/\Delta T_s = 2BT$, so that the no-signal density becomes

$$p(\rho | \rho_0 = 0) \approx \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} ; \quad \rho > 0 \quad (3-88)$$

where $\sigma^2 = \frac{1}{2BT}$. The mean and variance of ρ are

$$\langle \rho \rangle = \sqrt{\frac{\pi}{2}} \sqrt{\frac{1}{2BT}} \quad (3-89)$$

$$\sigma_{\rho}^2 = \left(2 - \frac{\pi}{2}\right) \left(\frac{1}{2BT}\right)$$

and where the "dimension" of ρ is the "fraction of correlated

bits". For the no-signal case, if $BT \gg 1$, the results of Equations (3-88) and (3-89) are correct for a VLBI system using infinitely-clipped samples which satisfy the Nyquist criterion of Equation (3-82).

3.5 EFFECTS OF INFINITE CLIPPING ON CORRELATION COEFFICIENT AND SNR

Most of our discussions in the previous sections have implicitly assumed that the recorded signals at each VLBI station were faithful reproductions of the actual waveforms, i.e. analog recordings. For some VLBI recording systems this is true, but with other recording systems, including the one used in the experiment reported in this thesis, we record a substantially distorted version of the signal in a digital form. The most common technique is to clip the analog waveform (which we shall denote as $x(t)$), preserving only information about the zero-crossings, and then to sample the clipped waveform at a uniform rate. Each sample may have only one of two values: $+1$ if $x(t) \geq 0$ and -1 if $x(t) < 0$. The resulting stream of samples is then stored digitally on magnetic tape, one bit per sample. In this section we shall discuss the effect of this clipping and sampling procedure on the correlation coefficient and the signal-to-noise ratio.

The auto-correlation of clipped noise has been analyzed by Van Vleck and Middleton (Reference 30) and the same techniques have been applied by several authors [Weinreb (Reference 31), Moran (Reference 6), Staelin (Reference 32), Thomas (Reference 33)] to analyze the cross-correlation of clipped noise. For completeness we shall include a brief derivation here. Let $x(t)$ be the signal

(voltage) received by Station 1 and $y(t)$ be the signal received by Station 2, where $x(t)$ and $y(t)$ are stationary zero-mean Gaussian variables. The joint probability density of $x(t)$ and $y(t+\tau)$ is given by

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_0^2}} \exp\left\{-\frac{1}{2(1-\rho_0^2)} \left[\frac{x^2}{\sigma_1^2} - \frac{2\rho_0}{\sigma_1\sigma_2} xy + \frac{y^2}{\sigma_2^2} \right] \right\} \quad (3-90)$$

where $X \equiv x(t)$, $Y \equiv y(t+\tau)$, $\rho_0 = \rho_0(\tau)$ is the "true" correlation coefficient, and σ_1^2 , σ_2^2 are the variances (i.e. total powers) of $x(t)$ and $y(t)$ respectively. (See, for example, Wozen-craft and Jacobs, Reference 25). By direct integration of Equation (3-90) it can easily be shown that

$$\begin{aligned} \sigma_1^2 &= \langle x^2(t) \rangle \\ \sigma_2^2 &= \langle y^2(t) \rangle \\ \rho_0(\tau) &= \frac{\langle x(t)y(t+\tau) \rangle}{\sigma_1\sigma_2} \end{aligned} \quad (3-91)$$

In any real VLBI experiment it is desired to form an estimate of $\rho_0(\tau)$ from a finite-length sample of $x(t)$ and $y(t)$. If, for example, the actual signals

$x(t)$ and $y(t)$ are available for a time interval extending from $t=0$ to $t=T$, a reasonable estimate of $\rho(\tau)$, designated $\hat{\rho}(\tau)$, is

$$\hat{\rho}(\tau) \equiv \frac{\int_0^T x(t)y(t+\tau)dt}{\left[\int_0^T x^2(t)dt \int_0^T y^2(t)dt \right]^{1/2}} \quad (3-92)$$

where $\tau \ll T$. Or, in the case of uniformly-sampled analog signals,

$$\hat{\rho}(\tau) \equiv \frac{\sum_{m=1}^K x(m\Delta t)y(m\Delta t+\tau)}{\left\{ \left(\sum_{m=1}^K x^2(m\Delta t) \right) \left(\sum_{m=1}^K y^2(m\Delta t) \right) \right\}^{1/2}} \quad (3-93)$$

where the intersample time Δt is usually chosen large enough to make the samples statistically independent and K is the number of samples. [The estimator given by Equation (3-93) is extensively analyzed in Appendices A, B, and C.] On the other hand, if only infinitely-clipped replicas of $x(t)$ and $y(t)$ are available for inspection, it is not possible to use the estimation algorithm of Equation (3-93). It is this latter case that we wish to investigate further.

We start by defining the infinitely-clipped replicas of $x(t)$ and $y(t+\tau)$ to be $X'(t)$ and $Y'(t)$ respectively, where

$$X'(t) = \begin{cases} +1 & \text{for } x(t) \geq 0 \\ -1 & \text{for } x(t) < 0 \end{cases} \quad (3-94)$$

$$Y'(t) = \begin{cases} +1 & \text{for } y(t+\tau) \geq 0 \\ -1 & \text{for } y(t+\tau) < 0 \end{cases}$$

Note that all total power information has disappeared from X' and Y' and that, in fact, $\langle X'^2(t) \rangle = \langle Y'^2(t) \rangle = 1$ regardless of the total powers σ_1^2, σ_2^2 of the original signals $x(t)$ and $y(t)$. We now need to examine the relationship between $\rho(\tau)$, as defined in Equation (3-91), and the correlation coefficient of $X'(t), Y'(t)$ defined by

$$\rho'(\tau) \equiv \langle X'(t)Y'(t) \rangle \quad (3-95)$$

where $Y'(t)$ is implicitly a function of $X'(t)$ through Equation (3-94). We may write $\rho'(\tau)$ explicitly in terms of the probability density of Equation (3-90) as

$$\rho'(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X'Y'p(X,Y)dXdY \quad (3-96)$$

which may be broken into integrations over each of the four quadrants of the X-Y plane

$$\rho'(\tau) = \left\{ \int_{-\infty}^0 \int_{-\infty}^0 + \int_0^{\infty} \int_0^{\infty} + \int_{-\infty}^0 \int_0^{\infty} + \int_0^{\infty} \int_{-\infty}^0 \right\} X'Y'p(X,Y) dXdY \quad (3-97)$$

which may be rewritten in a simpler form as

$$\rho'(\tau) = P_{--} + P_{++} - P_{+-} - P_{-+} \quad (3-98)$$

where, for example, P_{--} is the probability that both X' and Y' equal -1. From the symmetry of $p(X,Y)$, however, $P_{++} = P_{--}$ and $P_{+-} = P_{-+}$ so that Equation (3-98) may be rewritten

$$\rho'(\tau) = 2P_{++} - 2P_{+-} \quad (3-99)$$

Since the sum of all probabilities must be 1, we

have

$$2P_{++} + 2P_{+-} = 1 \quad (3-100)$$

which, taken with Equation (3-99), leads to

$$\rho'(\tau) = 4P_{++} - 1 = 4 \int_0^{\infty} \int_0^{\infty} p(X,Y) dXdY - 1 \quad (3-101)$$

which, substituting $p(X,Y)$ from Equation (3-90) with $\sigma_1 = \sigma_2 = 1$ (since X', Y' are completely independent of σ_1 and σ_2 , we are free to choose them for convenience) becomes

$$\rho'(\tau) = 4 \int_0^{\infty} \int_0^{\infty} \frac{1}{2\pi(1-\rho_0^2)^{1/2}} \exp\left\{-\frac{X^2 - 2\rho_0 XY + Y^2}{2(1-\rho_0^2)}\right\} dXdY - 1 \quad (3-102)$$

Making the substitution $X = r\cos\theta$, $Y = r\sin\theta$, Equation (3-102) becomes

$$\rho'(\tau) = 4 \int_0^{\pi/2} \int_0^{\infty} \frac{1}{2\pi(1-\rho_0^2)^{1/2}} \exp\left\{-\frac{r^2(1-2\rho_0 \sin\theta\cos\theta)}{2(1-\rho_0^2)}\right\} r dr d\theta - 1 \quad (3-103)$$

Performing the further substitution $\phi = 2\theta$ and

$z^2 = \frac{r^2(1-\rho_0 \sin\phi)}{2(1-\rho_0^2)}$, Equation (3-103) reduces immediately to

$$\rho'(\tau) = 4 \int_0^{\pi} \int_0^{\infty} \frac{(1-\rho_0^2)^{1/2}}{2\pi(1-\rho_0 \sin\phi)} e^{-z^2} z dz d\phi - 1 \quad (3-104)$$

which is easily evaluated with the use of tables (see, for example, Dwight, Reference 34) to yield

$$\rho'(\tau) = \frac{2}{\pi} \sin^{-1} \rho_0(\tau) \quad (3-105)$$

In general, we will only have a finite-length sample of $X'(t)$ and $Y'(t)$ from which to make an estimate of $\rho'(\tau)$. Let us denote this estimate as $\hat{\rho}'(\tau)$ and define it as

$$\hat{\rho}'(\tau) \equiv \frac{1}{K} \sum_{m=1}^K X'(m\Delta t) Y'(m\Delta t) \quad (3-106)$$

where $X'(t)$ and $Y'(t)$ have been sampled at an interval Δt sufficiently large so that the samples are statistically independent. From $\hat{\rho}'(\tau)$ we now wish to make an estimate of the true correlation coefficient $\rho_0(\tau)$, which we shall designate $\hat{\rho}_c(\tau)$ so as to distinguish it from $\hat{\rho}(\tau)$ of Equation (3-92). A logical choice for $\hat{\rho}_c(\tau)$ is based on the inverse of Equation (3-105) and we are led thereby to define the estimator

$$\hat{\rho}_c(\tau) \equiv \sin\left[\frac{\pi}{2} \hat{\rho}'(\tau)\right] \quad (3-107)$$

Since the functional relationship between $\hat{\rho}'(\tau)$ and $\hat{\rho}_c(\tau)$ is non-linear, we must carefully examine the statistics of $\hat{\rho}_c(\tau)$ to check for any possible biases or unreasonably large variance. Such an analysis has been carried out for the auto-correlation case by Weinreb (Reference 31) and is performed for the case of cross-correlation in Appendix D of this thesis. We shall summarize the results here and compare them with the statistics of the analog estimate of $\rho_0(\tau)$, [i.e. $\hat{\rho}(\tau)$ of Equation (3-92)] which is analyzed in

detail in Appendices A, B, and C.

For the case of infinitely-clipped and sampled signals, the results of Equations (D.17) and (D.18) of Appendix D show that

$$\langle \hat{\rho}_c(\tau) \rangle \approx \rho_o(\tau) \left[1 - \frac{\pi^2}{8K} (1 - \rho'^2(\tau)) \right] \quad (3-108)$$

$$\sigma_{\hat{\rho}_c}^2 \approx \frac{\pi^2}{4K} [1 - \rho_o^2(\tau)] [1 - \rho'^2(\tau)] \quad (3-109)$$

for the case where $K \gg \pi$, where K is the number of samples used to find $\hat{\rho}'(\tau)$ of Equation (3-106), and where $\rho'(\tau) = \frac{2}{\pi} \sin^{-1} \rho_o(\tau)$ as given by Equation (3-105). For large values of K , the bias in the estimate of $\rho_o(\tau)$ is small compared with the noise $\sigma_{\hat{\rho}_c}(\tau)$ and is usually neglected. We shall see later in Chapter 4, however, that the computer processing program called VLBI 1, which does the "raw" bit-by-bit correlations, has $K = 16$ so that the bias of the estimate of $\rho_o(\tau)$ is in fact non-negligible.

The results of Equations (B. 17) and (B. 21) of Appendix B show that the mean and variance of $\hat{\rho}(\tau)$, the "analog" estimate of $\rho_o(\tau)$ defined in Equation (3-93), is

$$\langle \hat{\rho}(\tau) \rangle = \rho_o(\tau) \quad (3-110)$$

$$\sigma_{\hat{\rho}(\tau)}^2 \approx \frac{1}{K} \{ 1 - \rho_o^2(\tau) [2 - \rho_o^2(0)] \} \quad (3-111)$$

From Equations (3-109) and (3-111) we can examine the increase in noise due to the infinite-clipping of

the signals. For very small correlation amplitudes (i.e. $\rho(\tau) \ll 1$), Equation (3-105) becomes

$\rho'(\tau) \approx \frac{2}{\pi} \rho(\tau)$ and Equation (3-109) becomes

$$\begin{aligned} \sigma_{\hat{\rho}_c}^2(\tau) &\approx \frac{\pi^2}{4K} [1 - \rho_o^2(\tau)] \left[1 - \frac{4}{\pi^2} \rho_o^2(\tau)\right] \\ &\approx \frac{\pi^2}{4K} [1 - 1.4 \rho_o^2(\tau)] \end{aligned} \quad (3-112)$$

Also, for $\rho(\tau) \ll 1$, Equation (3-111) becomes

$$\sigma_{\hat{\rho}}^2(\tau) \approx \frac{1}{K} [1 - 2 \rho_o^2(\tau)] \quad (3-113)$$

so that the increase in noise due to infinite-clipping is approximately

$$\frac{\sigma_{\hat{\rho}_c}^2(\tau)}{\sigma_{\hat{\rho}}^2(\tau)} \approx \frac{\pi}{2} \left[\frac{1 - 1.4 \rho_o^2(\tau)}{1 - 2 \rho_o^2(\tau)} \right]^{1/2} \approx \frac{\pi}{2} [1 + 0.3 \rho_o^2(\tau)] \quad (3-114)$$

For the case of large correlation amplitude ($\rho(\tau) \approx 1$)

Equation (3-111) becomes approximately

$$\sigma_{\hat{\rho}}^2(\tau) \approx \frac{1}{K} [1 - \rho_o^2(\tau)] \quad (3-115)$$

so that the increase in noise due to infinite-clipping in this case, using Equations (3-105), (3-109) and (3-115), is

$$\frac{\sigma_{\hat{\rho}_c(\tau)}}{\sigma_{\hat{\rho}(\tau)}} \approx \frac{\pi}{2} \left[1 - \left(\frac{2}{\pi} \sin^{-1} \rho_0(\tau) \right) \right]^{1/2} \quad (3-116)$$

For most VLBI observations $\rho_0(\tau) \ll 1$ so that the result of Equation (3-114) applies, indicating that the noise level (and hence the SNR) of the infinitely-clipped signal case is approximately a factor of $\frac{\pi}{2}$ higher than for the analog signal case. Therefore, in applying any of the earlier results of this chapter to the infinitely-clipped signal case, the SNR of Equation (3-55) should be decreased by a factor of $\frac{\pi}{2}$; or, equivalently, the true correlation amplitude ρ_0 may be replaced by an "effective true correlation amplitude", ρ_{eff} ,

$$\rho_{\text{eff}} = \frac{2}{\pi} \rho_0 \quad (3-117)$$

Usually, the small loss of SNR implied by Equation (3-114) for the infinitely-clipped signal case is more than offset by the advantages of the one-bit digital recording system, advantages which lie primarily in the ease of data-taking and data processing.

3.6 GROUP-DELAY AND PHASE-DELAY RATE ESTIMATION

3.6-1 Single-Band Delay Estimation

As we have seen in Section (3.2-2) the maximum-likelihood estimate of group delay is found simply by maximizing the expression (3-16). For a VLBI system with little or no dispersion of the signal through either the propagation medium or the electronic system, the observed fringe phase at any given frequency within the observing band is proportional to that frequency. For such a case the best estimate of group delay is simply the best-fit straight line through the measured phases as a function of frequency.

Consider an observation, of length T , covering a single frequency window of bandwidth B and having a true correlation amplitude ρ_0 . Let the observed fringe phase be computed for K uniformly-spaced independent points $\omega_1, \dots, \omega_k$ within the bandwidth B , so that the spacing between points is B/K . The signal-to-noise ratio of the point at frequency ω_k is then given by Equation (3-55) to be

$$(\text{SNR})_k = \rho_0 \sqrt{2 \left(\frac{B}{K}\right) T} \quad (3-118)$$

where we assume that ρ_0 is independent of k . The r.m.s. error of the observed fringe phase, $\hat{\phi}$, is then determined from Equation (3-58) to be

$$\sigma_{\hat{\phi}_k} \approx \frac{1}{(\text{SNR})_k} = \frac{\sqrt{K}}{\text{SNR}} \quad (\text{radians}) \quad (3-119)$$

where $\text{SNR} \equiv \rho_0 \sqrt{2BT}$ is the signal-to-noise ratio of the total bandwidth B.

The variance of the estimate of the slope of the best-fit straight line through the K points is just the variance of the estimate of the group delay, which we shall designate $\sigma_{\hat{\tau}}$. From Equation (E.35) of Appendix E it follows directly that

$$\sigma_{\hat{\tau}}^2 \approx \frac{\sigma_{\hat{\phi}_k}^2}{K \Delta\omega_{\text{rms}}^2} \quad (3-120)$$

where

$$\Delta\omega_{\text{rms}}^2 \equiv \frac{1}{K} \sum_{k=1}^K (\omega_k - \bar{\omega})^2 \quad (3-121)$$

and where

$$\bar{\omega} = \frac{1}{K} \sum_{k=1}^K \omega_k \quad (3-122)$$

For K uniformly-spaced points spanning a bandwidth B, $\Delta\omega_{\text{rms}}$ may be approximated by

$$\Delta\omega_{\text{rms}}^2 \approx \frac{1}{B} \int_0^B (\omega - \frac{B}{2})^2 d\omega = \frac{B^2}{12} \quad (3-123)$$

Using Equations (3-119) and (3-123) in Equation (3-120) we then have

$$\sigma_{\hat{t}}^2 \approx \frac{12}{B^2 (\text{SNR})^2} \quad (3-124)$$

where again $\text{SNR} = \rho \sqrt{2BT}$. Equation (3-124) is independent of K , as intuition indicates it should be. Note also that Equation (3-124) is independent of the observing frequency; the similar result for delay rate, as we shall see, is not independent of the observing frequency.

3.6-2 Switched-Frequency Delay Estimation

The case of delay estimation for switched-frequency VLBI data is a simple extension of the above. In fact, it may be considered identical except that the frequencies which are sampled are not uniformly spaced and Equation (3-120) must be evaluated explicitly. Using Equation (3-119) in Equation (3-120)

$$\sigma_{\hat{f}}^2 \approx \frac{1}{\Delta\omega_{\text{rms}}^2 \cdot (\text{SNR})^2} \quad (3-125)$$

where $\Delta\omega_{\text{rms}}$ is defined in Equations (3-121) and (3-122) and where we have assumed that the correlation amplitude is independent of frequency. Usually, the spacings of the frequency windows of a multi-frequency observation are large compared to the bandwidth of a single window so that only the spacings between the frequency windows need be used to calculate $\Delta\omega_{\text{rms}}$.

The results of Appendix E allow us to easily generalize the result of Equation (3-125) to the case of a frequency-dependent correlation amplitude. This case is encountered in practice when, for example, the gain of the front-end amplifier varies from one frequency channel to another, causing a related variation in system temperature over the frequency channels. Then, for the case of N frequency channels $\omega_1, \dots, \omega_N$, each with bandwidth B and with associated correlation amplitude ρ_n , we may use Equation (E.26) of Appendix E to find the variance of the group delay estimate for an observation of length T

$$\sigma_{\tau}^2 \approx \frac{1}{\Delta} \sum_{n=1}^N \frac{1}{\sigma_n^2} \quad (3-126)$$

where

$$\Delta \equiv \left(\sum_{n=1}^N \frac{1}{\sigma_n^2} \right) \left(\sum_{n=1}^N \frac{\omega_n^2}{\sigma_n^2} \right) - \left(\sum_{n=1}^N \frac{\omega_n}{\sigma_n^2} \right)^2 \quad (3-127)$$

$$\sigma_n^2 \approx \frac{\sqrt{N}}{\rho_{o_n} \sqrt{2BT}}$$

where ρ_{o_n} is the correction amplitude of the n^{th} frequency channel and where we have assumed that the bandwidth B is relatively small compared to the spacing of the frequency channels. For the case that ρ_{o_n} is independent of n, Equation (3-126) reduces to Equation (3-25).

3.6-3 Single-Band Delay Rate Estimation

The estimation of phase-delay rate is performed in a manner exactly analogous to the estimation of group delay. As we shall see in Section 4.4, the estimate of phase-delay rate is made by assuming that the a priori phase-delay rate computations are accurate enough so that the residual phase-delay rate is a constant over the duration of the observation. The estimation of phase-delay rate then reduces to a simple straight-line least squares fit of the residual phase-delay rate vs. time.

Consider a single-channel observation of integration time T seconds (not necessarily continuous) divided into L segments of length T/L seconds each. Assume that the correlation amplitude, ρ , is constant over the entire observation. Then from Equations (3-55) and (3-58) the rms phase error for the ℓ^{th} segment is

$$\sigma_{\hat{\phi}_\ell} = \frac{1}{\rho \sqrt{2B \left(\frac{T}{L}\right)}} = \frac{\sqrt{L}}{\text{SNR}} \quad (3-128)$$

where $\text{SNR} = \rho \sqrt{2BT}$ from Equation (3-55) and where B is the bandwidth. If the radio frequency is ω , then the corresponding rms phase-delay error is $\sigma_{\hat{\phi}_\ell} / \omega$. From Equation (E.35) of Appendix E, the variance of the estimate of the slope of the best-fit straight line through the L segments of phase delay is just

$$\sigma_{\hat{\tau}}^2 \approx \frac{(\sigma_{\hat{\phi}_\ell} / \omega)^2}{L \Delta T_{\text{rms}}^2} = \frac{1}{\omega^2 \Delta T_{\text{rms}}^2 (\text{SNR})^2} \quad (3-129)$$

where

$$\Delta T_{\text{rms}}^2 \equiv \frac{1}{L} \sum_{\ell=1}^L (t_\ell - \bar{t})^2 \quad (3-130)$$

where t_ℓ is the epoch of the center of the ℓ^{th} segment, and where

$$\bar{t} = \frac{1}{L} \sum_{\ell=1}^L t_\ell \quad (3-131)$$

In the usual case the integration period is continuous over T seconds, in which case $\Delta T_{\text{rms}}^2 = T^2/12$. Substituting this and Equation (3-128) into Equation (3-129) we then have the variance of the estimate of phase-delay rate for a continuous observation of length T

$$\sigma_{\hat{\tau}}^2 \approx \frac{12}{\omega^2 T^2 (\text{SNR})^2} \quad (3-132)$$

where $\text{SNR} = \rho\sqrt{2BT}$, ρ is the correlation coefficient and B is the bandwidth. For an observation which is non-continuous but which has a total integration time T , $\sigma_{\hat{\tau}}^2$ must be found by explicitly evaluating Equation (3-129).

3.6-4 Switched-Frequency Delay-Rate Estimation

The extension of the result of Equation (3-129) to the switched-frequency case is straight-forward. Consider the case of switching cyclically through N frequency channels with the n^{th} channel at radio frequency ω_n .

Again, let us divide the observation into L time-segments, each with phase variance σ_θ^2 . Then we may use Equation (E.26)

to calculate $\sigma_{\hat{\tau}}^2$

$$\sigma_{\hat{\tau}}^2 \approx \frac{1}{\Delta} \sum_{\ell=1}^L \frac{\omega_n^2}{\sigma_{\hat{\phi}}^2} \quad (3-133)$$

where n is implicitly a function of time-segment number ℓ , and where

$$\Delta = \left(\sum_{\ell=1}^L \frac{\omega_n^2}{\sigma_{\hat{\phi}}^2} \right) \left(\sum_{\ell=1}^L \frac{\omega_n^2 t_\ell^2}{\sigma_{\hat{\phi}}^2} \right) - \left(\sum_{\ell=1}^L \frac{\omega_n^2 t_\ell}{\sigma_{\hat{\phi}}^2} \right)^2 \quad (3-134)$$

where

$$\sigma_{\hat{\phi}} = \frac{\sqrt{L}}{\text{SNR}} \quad (3-135)$$

For convenience let us choose L to be equal to the total number of "switching times" in the observation, where a "switching time" is the time spent at one frequency-channel before switching to the next frequency channel (typically 0.2 seconds). If the N frequencies are switched through in a cyclical fashion, there will be $K = L/N$ such "switching cycles" in an observation. If the frequency-switching is very rapid compared to T, as is usually the case, we may assign an "average" epoch t_k to the k^{th} cycle and make the approximations

$$\sum_{\ell=1}^L \omega_n^2 t_{\ell}^2 \approx \left(\sum_{n=1}^N \omega_n^2 \right) \left(\sum_{k=1}^{L/N} t_k^2 \right) \quad (3-136)$$

and

$$\sum_{\ell=1}^L \omega_n^2 t_{\ell} \approx \left(\sum_{n=1}^N \omega_n^2 \right) \left(\sum_{k=1}^{L/N} t_k \right) \quad (3-137)$$

If we define

$$\overline{\omega^2} = \frac{1}{N} \sum_{n=1}^N \omega_n^2 \quad (3-138)$$

then, using Equation (3-136) and (3-137), we find that Equation (3-134) becomes

$$\begin{aligned} \Delta &\approx \frac{N^2 (\overline{\omega^2})^2}{\sigma_{\hat{\phi}}^4} \left[\sum_{k=1}^{L/N} t_k^2 - \left(\sum_{k=1}^{L/N} t_k \right)^2 \right] \\ &= \frac{N^2 (\overline{\omega^2})^2}{\sigma_{\hat{\phi}}^4} \cdot \frac{L}{N} \Delta T_{\text{rms}}^2 \end{aligned} \quad (3-139)$$

where ΔT_{rms}^2 is defined by Equation (3-130). Substituting Equations (3-135) and (3-139) into Equation (3-133) we have

$$\sigma_{\hat{\tau}}^2 \approx \frac{1}{\overline{\omega^2} \Delta T_{\text{rms}}^2 (\text{SNR})^2} \quad (3-140)$$

This is just the result we might have guessed from Equation (3-129). In the usual case of a continuous observation of length T, Equation (3-140) reduces to

$$\sigma_{\hat{t}}^2 \approx \frac{12}{\overline{\omega^2} T^2 (\text{SNR})^2} \quad (3-141)$$

where $\overline{\omega^2}$ is given by Equation (3-138) and $\text{SNR} = \rho\sqrt{2BT}$.

3.7 SIGNAL DETECTION CHARACTERISTICS

3.7-1 Signal Detection Probabilities

The usual procedure for detecting fringes in VLBI data processing is to compute the quantity $|\vec{R}|$ of Equation (3-37) for a large number of trial delays and delay rates, spanning a range within which the signal is expected, and then to simply pick the trial values which maximize $|\vec{R}|$. In the case of a weak signal, it is of some interest to compute the probability of successfully detecting the signal using this procedure. In this regard we may formalize the question to ask:

Given M independent values of $|\vec{R}|$ (such as M independent points in a trial fringe-rate spectrum), $M-1$ of which contain only Rayleigh distributed noise and one of which contains a signal of strength ρ_0 embedded in identical noise, what is the probability that the signal will be detected by the procedure of choosing the trial for which $|\vec{R}|$ is a maximum?

Let $p(\rho|\rho_0)$ be the probability density of the correlation amplitude for the channel containing the signal, where ρ_0 is the true "effective" signal correlation amplitude [see Equation (3-117)]; let $p(\rho|0)$ be the probability density for the correlation amplitudes for each of the remaining $M-1$ channels, which contain no signal. These are the same probability densities

that we have discussed earlier in Section 3.4, and are given by Equations (3-71) and (3-73), respectively. The probability that the correlation amplitude of the "signal channel" falls in the range ρ_s to $\rho_s + d\rho_s$ and that all the $M-1$ "noise channels" have correlation amplitudes less than ρ_s is

$$p(\rho_s | \rho_0) d\rho_s \left[\int_0^{\rho_s} p(\alpha | 0) d\alpha \right]^{M-1} \quad (3-142)$$

To find the probability of successful detection, given a true signal amplitude ρ_0 , we integrate over all ρ_s

$$\Pr[\rho_s > \text{all } \rho_n | \rho_0] = \int_0^{\infty} p(\rho_s | \rho_0) \left[\int_0^{\rho_s} p(\alpha | 0) d\alpha \right]^{M-1} d\rho_s \quad (3-143)$$

This function has been numerically computed and is shown in Figure 3-5 for several values of M . As we would expect, the probability of successful signal detection approaches $\frac{1}{M}$ for the small SNR case and 1 for the large SNR case. In practice M is typically on the order of a thousand or more, making the detection of very weak signals (say $\rho_0 \lesssim 5\sigma$) difficult. It is often possible, however, to substantially reduce M (often down to 10 or so) if the a priori source position, baseline and clock parameters are well enough known.

This technique is particularly useful for the special case in which the

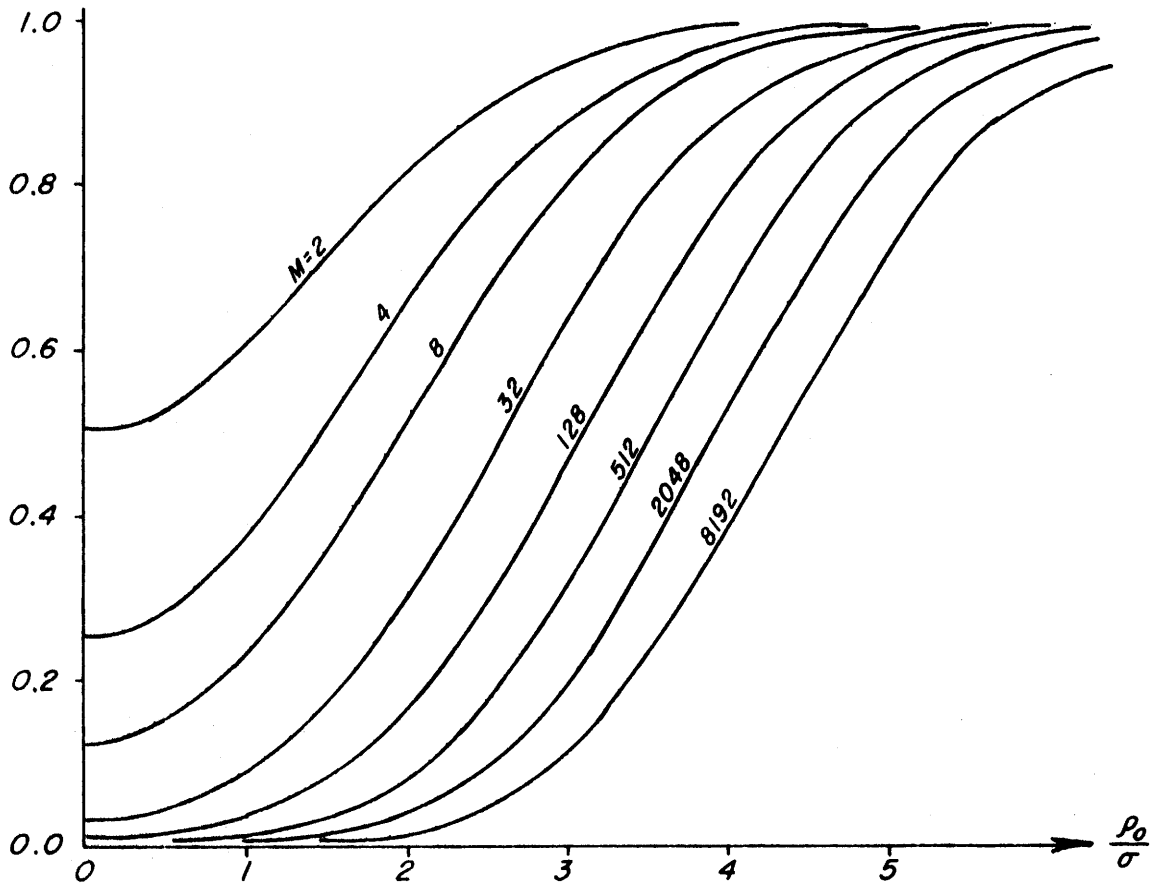


Figure 3-5

Signal detection probability for a search over M independent channels. Horizontal axis is equivalent to effective SNR.

correlation amplitude is reduced at certain baseline orientations due to small-scale source structure. Such small-scale structure is not untypical of a number of sources used for VLBI (see, for example, References). Observations made at times when the baseline orientation is such that the correlation amplitude is relatively high can be used to increase the accuracy of the a priori model and to allow the search range to be substantially reduced, thus significantly increasing the probability of detecting a weak signal.

3.7-2 Distribution of Maximum Correlation Amplitude in No-Signal Case

As we noted in the previous section the usual procedure for signal detection is simply to pick the maximum amplitude among M independent channels (or trials). In order to make a determination of whether such a maximum is indeed a signal, or is just noise, it is useful to know the expected characteristics of the no-signal case. We can find the probability distribution for the maximum amplitude in the no-signal case as a simple extension of the work of the previous section.

The probability that the maximum is found in the j^{th} channel with amplitude between ρ_{n_j} and $\rho_{n_j} + d\rho_{n_j}$ and that all other $M-1$ channels have amplitudes less than ρ_{n_j} is

$$[p(\rho_{n_j}|0)d\rho_{n_j}] \left[\int_0^{\rho_{n_j}} p(\alpha|0)d\alpha \right]^{M-1} \quad (3-144)$$

where $p(\rho|0)$ is the Rayleigh probability density given by Equation (3-88) as

$$p(\rho|0) = \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} \quad ; \quad \rho > 0 \quad (3-145)$$

where $\sigma^2 = \frac{1}{2BT}$. If all channels have the same Rayleigh statistics, then each channel is equally likely to contain the maximum. Thus the probability that the maximum

lies in the range ρ_n to $\rho_n + d\rho_n$ is

$$M[p(\rho_n|0)d\rho_n] \left[\int_0^{\rho_n} p(\alpha|0)d\alpha \right]^{M-1}$$

The probability density function for the maximum correlation amplitude is then

$$p(\rho_n) = Mp(\rho_n|0) \left[\int_0^{\rho_n} p(\alpha|0)d\alpha \right]^{M-1} \quad (3-146)$$

This function has been computed numerically and is plotted in Figure 3-6a for several values of M ; $\langle \rho_n \rangle$ and σ_{ρ_n} are plotted in Figure 3-6b as a function of M . For a typical observation recorded on the Mark I digital recording system (see Section 6.7) with $B = 360$ kHz and $T = 180$ sec, the value of σ in Equation (3-145) is $\sigma = (2BT)^{-1/2} \approx 0.88 \times 10^{-4}$. For $M \approx 4000$, as it typical, Figure 3-66 shows the expected value of the maximum noise peak to be $\langle \rho_n \rangle \approx 4.1\sigma \approx 3.6 \times 10^{-4}$. If the Van Vleck "correction" of Equation (3-107) is applied as a matter of routine to all the data (as is usually the case), this value is raised to $\frac{\pi}{2} \times 3.6 \times 10^{-4} \approx 5.7 \times 10^{-4}$, which is close to typical empirical values.

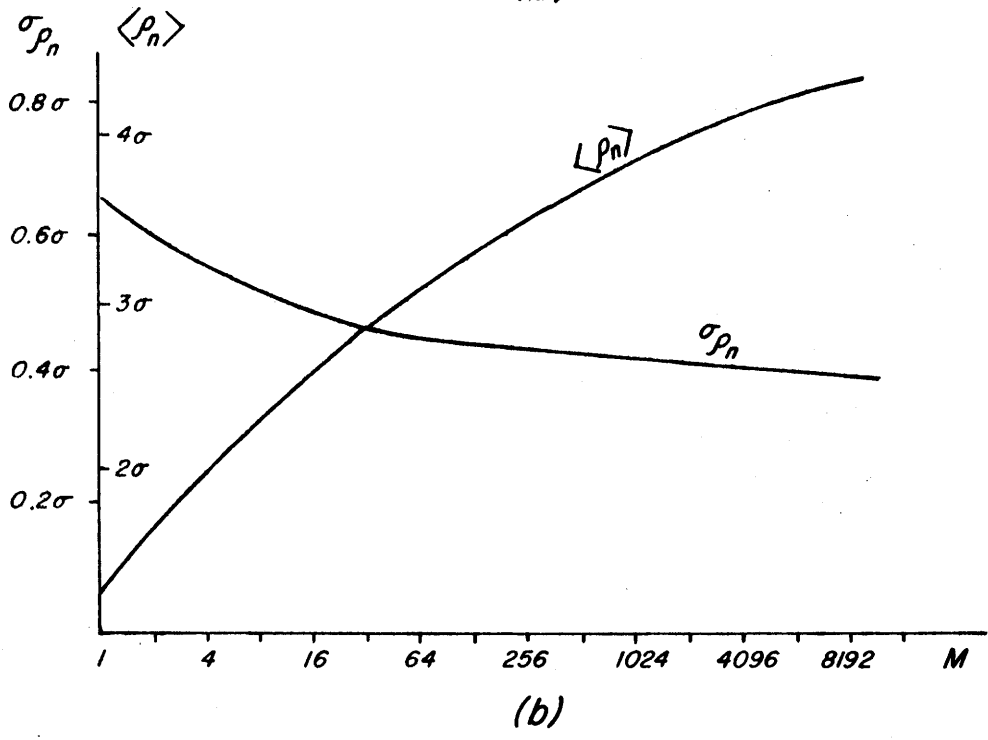
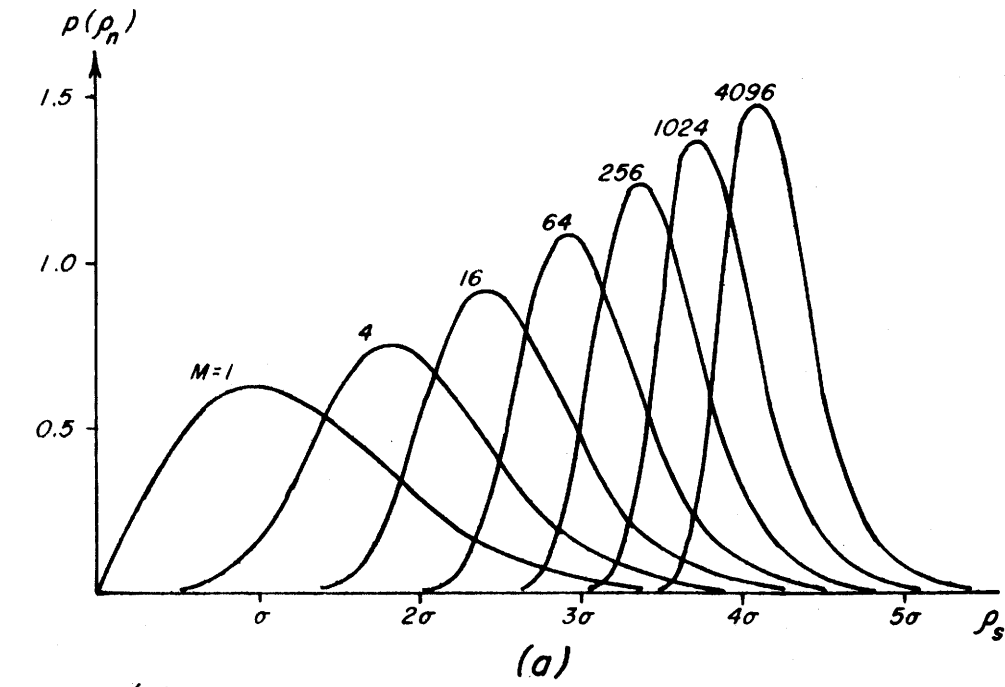


Figure 3-6

(a) Probability density for maximum correlation amplitude over M independent trials in the no-signal case.

(b) Mean and standard deviation of ρ_n as a function of M .

3.8 SUMMARY OF IMPORTANT RESULTS

In this section we shall briefly review and restate the important results of this chapter that are commonly used in "everyday" calculations. Consider a switched-frequency VLBI observation with instantaneous recording bandwidth B and duration T . Let the off-source system temperatures be T_{s_1} and T_{s_2} and let the antenna temperatures due to the source be T_{a_1} and T_{a_2} . If we assume a fringe visibility γ ($0 \leq \gamma \leq 1$), then from Equation (3-56) the "true" correlation coefficient is

$$\rho_o = \gamma \sqrt{\frac{T_{a_1} T_{a_2}}{(T_{a_1} + T_{s_1})(T_{a_2} + T_{s_2})}} \quad (3-147)$$

which, for the usual case of $T_{a_1} \ll T_{s_1}$ and $T_{a_2} \ll T_{s_2}$, reduces to

$$\rho_o \approx \gamma \sqrt{\frac{T_{a_1} T_{a_2}}{T_{s_1} T_{s_2}}} \quad (3-148)$$

In order to compute the signal-to-noise ratio (SNR), we must define an "effective" correlation coefficient, ρ_{eff} , which is dependent upon the type of recording system used. For an analog system where the exact replicas of the signals are recorded and processed,

$$\rho_{eff} = \rho_o \quad (3-149)$$

For a one-bit infinitely-clipped system, as is commonly used for many VLBI observations, the results of Section 3.5 show us that, for $\rho_0 \ll 1$,

$$\rho_{\text{eff}} \approx \frac{2}{\pi} \rho_0 \quad (3-150)$$

Then the signal-to-noise ratio is given by Equation (3-55) as

$$\text{SNR} = \rho_{\text{eff}} \sqrt{2BT} \quad (3-151)$$

We note that $2BT$ is just the number of bits recorded at each station if sampling is done at the Nyquist rate. If we let $K \equiv 2BT$, then SNR is simply

$$\text{SNR} = \rho_{\text{eff}} \sqrt{K} \quad (3-152)$$

For the usual case where $\rho \ll 1$, the rms fringe-phase error is given by Equation (3-58) as

$$\sigma_{\hat{\phi}} \approx \frac{1}{\text{SNR}} \quad (\text{radians}) \quad (3-153)$$

If the observation covers N frequency windows $\omega_1, \dots, \omega_n, \dots, \omega_N$ in a sequentially-switched manner and the correlation coefficient is independent of n , the rms error of the group delay determination is given by Equation (3-125) as

$$\sigma_{\hat{\tau}} \approx \frac{1}{\Delta\omega_{\text{rms}} \cdot \text{SNR}} \quad (3-154)$$

where

$$\Delta\omega_{\text{rms}}^2 \equiv \frac{1}{N} \sum_{n=1}^N (\omega_n - \bar{\omega})^2 \quad (3-155)$$

and

$$\bar{\omega} = \frac{1}{N} \sum_{n=1}^N \omega_n$$

The rms error in the determination of phase-delay rate, again under the assumption that the correlation coefficient is independent of frequency window, is given by Equation (3-141) as

$$\sigma_{\dot{\omega}} \approx \sqrt{\frac{12}{\bar{\omega}^2}} \frac{1}{T \cdot \text{SNR}} \quad (3-156)$$

where

$$\bar{\omega}^2 = \frac{1}{N} \sum_{n=1}^N \omega_n^2$$

For purposes of "getting the feel" of signal-to-noise ratios for real interferometer systems, it is convenient to "calibrate" oneself in terms of some set of standard parameters and then simply scale up or down, as the case may be, for any particular system of interest. Consider a typical, but hypothetical, one-bit digital VLBI system using two identical antennas and having the following easy-to-remember parameters:

Antenna diameter	$\equiv D_o = 100 \text{ ft.}$
Antenna efficiency	$\equiv \eta_o = 50\%$
Receiver system temperature	$\equiv T_{s_o} = 100^\circ\text{K}$
Recording bandwidth	$\equiv B_o = 360 \text{ kHz}$
Integration time	$\equiv T_o = 180 \text{ seconds}$
Source strength	$\equiv S_o = 1 \text{ flux unit}$
Fringe visibility	$\equiv \gamma = 1$

The antenna temperature, T_{a_o} , for this set of parameters, assuming that the antenna receives only a single polarization (either linear or circular) from a broadband, unpolarized source, is given by (see, for example, Kraus, Reference 35)

$$T_{a_o} = \frac{\eta_o A_o S_o}{2k} \approx 0.132^\circ\text{K}$$

where $A_o = \pi(D_o/2)^2$ is the antenna area,

$k = 1.38 \times 10^{-23}$ joule/ $^\circ\text{K}$ is the Boltzmann constant,

1 f.u. $\equiv 10^{-26}$ watt $\text{m}^{-2}\text{Hz}^{-1}$

Then the signal-to-noise ratio may be determined from Equations (3-147), (3-150) and (3-151) to be

$$\text{SNR} \approx 15$$

The signal-to-noise ratio for any one-bit digital VLBI system with parameters $D_1, D_2, \eta_1, \eta_2, T_{s_1}, T_{s_2}, B, T, S$, may then be easily determined from the relation

$$\text{SNR} \approx 15 \cdot \left[\frac{\eta_1}{0.5} \cdot \frac{\eta_2}{0.5} \cdot \frac{B}{360\text{kHz}} \cdot \frac{T}{180 \text{ sec}} \right. \\ \left. \cdot \frac{100^\circ\text{K}}{T_{s_1}} \cdot \frac{100^\circ\text{K}}{T_{s_2}} \right]^{\frac{1}{2}} \cdot \left[\frac{D_1}{100 \text{ ft}} \cdot \frac{D_2}{100 \text{ ft}} \cdot \frac{S}{1 \text{ f.u.}} \right]$$

(3-157)

Or, if instead, the antenna temperatures T_{a_1} and T_{a_2} are known

$$\text{SNR} \approx 15 \cdot \left[\frac{T_{a_1}}{0.13^\circ\text{K}} \cdot \frac{T_{a_2}}{0.13^\circ\text{K}} \cdot \frac{B}{360 \text{ kHz}} \cdot \frac{T}{180 \text{ sec}} \right]^{\frac{1}{2}}$$

(3-158)

CHAPTER 4

DATA PROCESSING - ALGORITHMS AND SOFTWARE

4.1 THE BASIC DATA-PROCESSING PROCEDURE

Data taken during the actual VLBI experiment, which is more fully described in Chapters 5 and 6, were collected on several thousand reels of magnetic tape. Each tape contains approximately three minutes of data from a single station with the data being a one-bit clipped-and-sampled image of a 360 kHz bandwidth video signal (see Section 6.7). The sampling rate is controlled to be precisely 720 kbits/sec relative to the station clock. While being recorded on tape, these digital data are turned off for a few milliseconds near the end of every 0.2-second period to provide a periodic timing marker. Each of these (slightly less than) 0.2 second segments of data is called a "record" and the blank portion is called a "record gap". One full tape holds approximately 900 records of data. The first bit of each record is related in a precisely known way to the reading of the clock at the observing site, so that the time associated with any given bit within a record is easily determined simply by counting bits from the beginning of the record. If, during tape playback, the bit count becomes confused within a single record, due to a bad spot on the tape, for example, timing "sync" is easily re-

established at the beginning of the next record.

The clocks at all of the stations are synchronized a priori to within a few microseconds and the observation starting times are carefully coordinated in advance of the experiment so that data records on the tapes of the various stations are recorded nearly simultaneously. The maximum signal delay from one antenna to another due to earth geometry is limited to ≈ 20 msec, so that the cross-correlation of the data from different stations may be carried out on a record-to-record basis simply by shifting one bit stream relative to the other by the amount of the expected signal delay and ignoring the relatively small amount of data lost due to non-overlapping bits at the beginning and end of each record.

Each 3-minute observation generates a set of tapes, one from each station, from which it is desired to find the best possible estimates of group delay, phase-delay rate, fringe phase, and correlation amplitude for each baseline. Measurements from many such observations are then combined to determine baseline components and source positions. The processing of data from a single 3-minute observation on a single baseline is broken, for convenience, into two steps. The first step is a basic cross-correlation and data-compression processing, dubbed "VLBI 1" which computes delays and fringe rates based on an a priori

model of baseline vectors and source coordinates, and then performs the necessary bit-by-bit cross-correlation. Cross-correlation is performed on one record at a time and covers a delay search range of about 10 microseconds, centered around the a priori computed delay. This range is generally large enough to insure that the actual signal delay is covered. Typically, after an initial search of several observations has been made to determine approximate clock synchronization errors, calculations of delay for further observations are within one or two microseconds of the actual delay. The resulting cross-correlation coefficients for each record are written on an intermediate tape, along with the relevant processing parameters. A data-compression ratio of about 200-to-1 is achieved by VLBI 1, reducing the mass of data from thousands of tapes to a much more manageable volume.

In the processing of an observation, the second step, dubbed "VLBI 2", combines the record-by-record data from the output of VLBI 1 and conducts an iterative trial search in delay and delay rate (i.e. fringe rate), around the a priori values calculated by VLBI 1, to maximize the correlation amplitude. The resultant best estimates of correlation amplitude, group delay, phase-delay rate, and fringe phase are punched onto cards as the final results of a particular observation. In the case where three stations are participating simultaneously, the data from each of the resulting three baselines are processed independently as a separate observation.

4.2 HISTORICAL SETTING

Up to the time of the October 1969 experiment described in this thesis, VLBI data processing for all experiments performed by the MIT group had been done almost entirely using the CDC 3300 computer at Haystack, with programs written initially by Moran (Reference 6) and modified for special purposes by this author and others. However, the minimum processing time for a single 3-minute continuum observation was about one hour, making the reduction of the 3000 pairs of tapes from the October 1969 experiment completely impractical. It was decided, therefore, to write an entirely new set of programs to run on an IBM 360 Model 91 computer at the NASA Goddard Space Flight Center, the use of which was made possible through the cooperation of Dr. T. A. Clark at Goddard. This machine is basically about 30 to 50 times faster than the CDC 3300, reducing the projected processing time from ~ 3000 hours to about 100 hours.

In order to prepare for the expected onslaught of data immediately following the October 1969 experiment, a large programming effort was mounted by H. F. Hinteregger starting in the summer of 1969. This effort was joined by the author (who had been busy building equipment during the summer) immediately following the experiment, and later was joined by Dr. A. E. E. Rogers. The culmination

of this effort, which was concluded some six months and more than 10,000 Fortran statements later, was a set of reasonably well-proven and efficient data-reduction and data-management programs. As always seems the case when data-reduction readiness falls far behind data-collection enthusiasm, strenuous attempts are made to get at least some data-reduction underway as soon as possible, even though programs are still evolving and being proven. This procedure inevitably causes at least some problems and incompatibilities with later-reduced data. Fortunately, such evolution problems were minimal with the October 1969 data. This was due, at least partly, to the adoption of the two-stage VLBI 1-VLBI 2 data-reduction technique which allowed the conceptually simpler VLBI 1 program to be developed, proven, and put into service long before the more complicated and subtle VLBI 2 program was finally tested and proven. Confident processing of the data could proceed through the time-consuming VLBI 1 program, with the easy reprocessing through VLBI 2 being accomplished as upgraded versions of VLBI 2 became available.

4.3 THE BASIC DELAY AND DELAY RATE ESTIMATION ALGORITHMS

In Section 3.2 we analyzed two simple statistical models of the VLBI problem, one in the frequency domain and the other in the time domain, to determine the maximum-likelihood (ML) estimates of delay and delay rate. In the frequency domain we found [see Equation (3-19)] that such estimates were given by

$$\max_{\tilde{\tau}, \dot{\tilde{\tau}}} \left| \sum_{kj} X_1(\omega_j) X_2^*(\omega_j) e^{-i\omega_j(\tilde{\tau} + \dot{\tilde{\tau}}t_k)} \right| \quad (4-1)$$

where $X_1(\omega)$, $X_2(\omega)$ are the spectra of the signals received at the two stations and $\tilde{\tau}, \dot{\tilde{\tau}}$ are the "trial" values of delay and delay rate. Implicitly assumed in Expression (4-1) are 1) the points, ω_j , in the received spectrum of a given station are statistically independent, 2) the signal power is independent of ω_j at each station, and 3) the delay rate, τ , is $\ll 1$. In the time-domain, the equivalent ML expression is given by (3-31b) as

$$\max_{\tilde{\tau}, \dot{\tilde{\tau}}} \int_0^T x_1(t) x_2(t + \tilde{\tau} + \dot{\tilde{\tau}}t) dt \quad (4-2)$$

where $x_1(t)$ and $x_2(t)$ are the signals received at the two stations and where we have changed the limits of integration to reflect an observation spanning the finite interval $0 < t < T$. In such a case we must insure that $\tau \ll T$, where τ is the actual delay; typically, $\tau \lesssim 2 \times 10^{-3}$ seconds

and $T \approx 10^2$ seconds, so that this condition is well met in practice.

Neither expression (4-1) nor expression (4-2) lends itself to direct computation. Direct computation of $X_1(\omega)$ and $X_2(\omega)$ of expression (4-1) is not practical on even the most sophisticated computing equipment. For analog recording systems, expression (4-2) could, with difficulty, be calculated since one tape may be "offset" by $\tilde{\tau}$ and "speeded up" by $\tilde{\tau}$ so that the signals on the two tapes come into precise alignment. On the other hand, digital recordings of sampled data can only be "offset" by an integral number of sampling periods, making the direct computation of expression (4-2) possible only for discrete values of $\tilde{\tau}$. Efficient computation of the ML estimates of delay and delay rate from digital recordings, then, requires that some other algorithm be developed which closely approximates the results of expressions (4-1) or (4-2).

In practice, of course, $x_1(t)$ and $x_2(t)$ are not recorded directly, but rather are recorded as bandlimited video signals, $x_1^v(t)$ and $x_2^v(t)$, after being "mixed down" from the radio frequency (see Figure 1-2 and attendant explanation).

The effects of the mixing and bandlimiting operation allow some approximations to be made which simplify the computational aspects of the ML function. Such an analysis is performed in Appendix F and we shall only quote the results here. The results of Equations (F. 14 , (F. 17), and (F. 18) of Appendix F show that the ML estimates of delay and delay rate may be computed in a basic three-step procedure:

Step 1: For each record, k, compute the "record cross-correlation function" given by

$$R_{12}^k(\tilde{\tau}'_k + m\Delta\tau_s) \equiv C_1 \sum_{\ell=1}^{L_k} R_{\ell}^k(\tilde{\tau}'_k + m\Delta\tau_s) e^{-i\omega_m^k \tilde{\tau}'_k t_{\ell}} ; \quad (4-3)$$

$$m = -3, -2, \dots, 2, 3$$

where

$R_{\ell}^k(\tilde{\tau}'_k + m\Delta\tau_s)$ is the correlation coefficient for the ℓ^{th} short segment of data of length Δt , where Δt is small enough that the change of fringe phase over Δt is small, i.e. $|\omega_m^k \tilde{\tau}'_k \Delta t| \ll 1$ (see Equation 4-7),

$\tilde{\tau}_k$ is the trial delay for the k^{th} record,

$\tilde{\tau}'_k$ is the nearest integral bit shift to $\tilde{\tau}_k$,

$\tilde{\tau}_k$ is the trial delay rate for the k^{th} record,

t_{ℓ} is the time of the middle of the ℓ^{th} data segment, starting at zero for $\ell=1$,

ω_m^k is the center radio frequency for the k^{th} record

L_k is the number of segments in the k^{th} record,

$\Delta\tau_s$ is the sampling period ($\approx 1.39 \mu\text{sec}$ for Mark I), and

C_1 is a normalization constant.

As we shall see in Equation (4-7), the form of $R_{\ell}^k(\tilde{\tau}'_k + m\Delta\tau_s)$ is somewhat different than that implied by Equation (F.17) due to the infinitely-clipped nature of the actual recorded video signals.

Step 2 : For each record, k , compute the "counter-rotated video cross-spectrum" given by

$$S_{12}^k(\omega_j^v) \equiv C_2 \sum_{m=-3}^3 R_{12}^k(\tilde{\tau}'_k + m\Delta\tau_s) e^{-i\omega_j^v m\Delta\tau_s} e^{-i\omega_0^k \tilde{\tau}_k} e^{-i\omega_j^v(\tilde{\tau}_k - \tilde{\tau}'_k)}; \quad j = 1, \dots, 7 \quad (4-4)$$

where

ω_j^v is the video frequency,

ω_0^k is the total local oscillator frequency for the k^{th} record, and

C_2 is a normalization constant.

Step 3: Evaluate the function

$$\max_{\tilde{\tau}_0, \tilde{\tau}'_0} \text{over} \left| \sum_{j=1}^7 \sum_{k=1}^K S_{12}^k(\omega_j^v) \right| \quad (4-5)$$

where $\tilde{\tau}_0$ is the trial delay for the epoch corresponding to the start time of the observation,

τ_0 is the trial delay rate for the epoch corresponding to the start time of the run, and
K is the number of records in the observation.

The actual algorithms used by VLBI-1 and VLBI-2 are close to those given in Equations (4-3), (4-4), and (4-5), although we shall see that they differ in some respects, usually for computational efficiency. The "record cross-correlation function" of Equation (4-3) is computed by VLBI-1, while the operation of Equations (4-4) and (4-5) are performed by VLBI-2. We note here that the algorithms given by Equations (4-3), (4-4), and (4-5) may be used for both single-band and frequency-switched VLBI data since the narrowband approximations made in Appendix F restrict only the instantaneous recording bandwidth, and not the total bandwidth spanned by the frequency windows.

4.4 THE VLBI 1 ALGORITHMS

VLBI-1 performs the cross-correlation and "fast fringe rotation", as given basically by Equation (4-3), on a record-by-record basis and places the results for each record onto a summary tape with no further summation. Correlation is performed for seven adjacent delays spaced one sampling period ($\approx 1.39 \mu\text{sec}$) apart and centered around the bit shift nearest to the calculated a priori delay. For each of the seven trial delays the correlation coefficient for each record is computed by a 4-step procedure:

- 1) The two bit strings, one from each of the two stations, are aligned to the proper trial delay.
- 2) Starting with the first pair of overlapping bits, the data are divided into adjacent segments of 16 bits each, corresponding to the interval Δt ($\approx 22.2 \mu\text{sec}$) in expression (4-5). The 16-bit segment length is short enough that, for fringe rates $\gtrsim 20 \text{ kHz}$, the fringe phase may be considered to a good approximation to be constant over the duration of the segment.
- 3) The correlation coefficient for each 16-bit

segment is computed and the Van Vleck correction applied. This (real) correlation coefficient is then "counter-rotated" by the a priori change in fringe phase, as computed from the beginning of the record.

4) The (complex) correlation coefficients from all of the segments are summed and normalized to yield the complex correlation coefficient for the record. We may mathematically state the cross-correlation procedure outlined above as follows: The actual "record cross-correlation function" computed¹ for the kth record of data is

$$R_{12}^k(\tau_{ap}^{k, +m\Delta\tau_s}) = \frac{1}{L_k} \sum_{\ell=1}^{L_k} R_{\ell}^k(\tau_{ap}^{k, +m\Delta\tau_s}) e^{-i\omega_m^k \dot{\tau}_{ap}^k (t_{\ell}^k - t_1^k)} ;$$

$$m = -3, -2, \dots, 2, 3$$

(4-6)

where

- $\tau_{ap}^{k,}$ = closest integral-bit shift delay to the a priori value of total delay, τ_{ap}^k , as calculated for the center of the kth record
- $\dot{\tau}_{ap}^k$ = a priori delay rate computed for the center of the kth record
- $\Delta\tau_s$ = sampling period (≈ 1.39 sec)

¹A very efficient machine algorithm for performing the bit correlation operation of Equation (4-6) was developed by Hinteregger and is described in detail in Reference 13 .

t_{ℓ}^k = epoch of the reference station at the center of the ℓ^{th} 16-bit segment of the k^{th} record. $\ell = 1$ corresponds to first segment.

ω_m^k = mid-band radio frequency for the k^{th} record.

$R_{\ell}^k(\tau_{\text{ap}}^{k, +m\Delta\tau_s})$ = correlation coefficient of the ℓ^{th} 16-bit segment of the k^{th} record for the m^{th} delay.

L_k = number of 16-bit segments correlated in the k^{th} record.

The segment correlation coefficient $R_{\ell}^k(\tau_{\text{ap}}^{k, +m\Delta\tau_s})$ is computed simply by offsetting the data bit strings by delay $\tau_{\text{ap}}^{k, +m\Delta\tau_s}$ and then counting the number of matching bits in the ℓ^{th} 16-bit segment of data. If there are q matching bits, then

$$R_{\ell}^k(\tau_{\text{ap}}^{k, +m\Delta\tau_s}) = \sin\left[\frac{\pi}{2} \cdot \frac{q-8}{8}\right] \quad (4-7)$$

where a simple Van Vleck correction has been applied [see Equation (3-107)].²

²In Appendix D we compute an additional bias in the correlation amplitude that results from applying the Van Vleck correction to a short segment of data. For 16-bit data segments, the mean correlation amplitude is reduced by the factor ~ 0.929 ; a suitable correction is therefore applied to the correlation amplitude at the completion of processing through VLBI-1 and VLBI-2.

Several points should be noted with regard to the cross-correlation algorithm of Equation (4-6):

1. The a priori total delay and delay rate for the center of the k^{th} record, designated τ_{ap}^k and $\dot{\tau}_{\text{ap}}^k$, respectively, are calculated only once for each record and are then held constant over the record. As we shall see in Section 4.6, this approximation is quite good for all observations of interest and leads only to a small, correctable bias in the estimated parameters.

2. The central trial delay, τ_{ap}^k , may differ by as much as one-half a sampling period from τ_{ap}^k . This leads to an error in fringe phase of $\omega_j^v (\tau_{\text{ap}}^k - \tau_{\text{ap}}^k)$ [see Appendix F]. This quantity is applied as a correction phase, known as the "fractional-bit correction", and is applied in VLBI 2 after the computation of the record cross spectrum [see Equations (4-4) and (4-8)].

3. The rotation phase applied to the first segment is zero and advances at a uniform rate for the succeeding segments. This means that the record cross-correlation coefficient $R_{12}^k(\tau_{\text{ap}}^k + m \Delta \tau_s)$ of Equation (4-6) must be further counter-rotated by the a priori total fringe phase, designated ϕ_{ap}^k and calculated for the time t_1^k of the k^{th} record, before any summation

of R_{12}^k or S_{12}^k over different records may be attempted [see Equation (4-5)]. In practice, ϕ_{ap}^k is calculated by VLBI 1 and passed to VLBI 2, which performs the necessary counter-rotation.

4. Since the bit string from the record of one station must be "shifted" with respect to that from the other to compensate for the delay, only the overlapping parts of the two records may actually be cross-correlated. If the station clocks are closely synchronized, the amount of data lost is just equal to the total delay. For a record length $\Delta T = 0.2$ seconds and a delay of 1.5×10^4 microseconds, about the maximum actually encountered, ~7% of the total data cannot be correlated. Also, since the delay changes at a rate of up to ~2 $\mu\text{sec}/\text{sec}$, the number of 16-bit segments correlated per record changes slowly from the beginning to the end of a single 3-minute observation. The number of segments, L_k , that can be correlated for each record must therefore be calculated carefully to account for these effects in order that $R_{12}^k(\tau_{ap}^k + m\Delta\tau_s)$ may be properly calculated and normalized.

After the data for a given record have been cross correlated, all of the relevant parameters and cross-correlation coefficients are placed in intermediate storage, usually tape, which serves as an input to VLBI 2 at a later

time. The actual parameters which are recorded on this tape for each record are:

- τ_{ap}^k , the a priori total delay computed for the center of the reference station's k^{th} record,
- τ_{ap}^k , the closest integral-bit shift delay to τ_{ap}^k ,
- $\omega_m^k \tau_{ap}^k$, the a priori total fringe rate which was used to counter-rotate the data of record k , where ω_m^k is the mid-band radio frequency for record k ,
- ϕ_{ap}^k , the a priori total fringe phase computed for the reference station epoch corresponding to the first correlated bit of the k^{th} record of the reference station tape,
- L_k , the number of 16-bit segments correlated for record k ,
- $R_{12}^k(t_{ap}^k + m\Delta\tau_s)$, the normalized complex cross-correlation coefficient for the m^{th} of seven trial delays from record k .

In addition, some other relevant parameters, such as the record lengths and parity as read by the machine, are recorded along with the above data.

4.5 THE VLBI 2 ALGORITHMS

The basic purpose of VLBI-2 is to accumulate the record-by-record correlations (from VLBI-1) with many different trial delays and delay rates, so as to find the maximum of the quantity in expression (4-5). The a priori calculations of delays and delay rates used by VLBI 1 are assumed to be of sufficient accuracy that the actual delay and delay rate may be characterized, over the duration of a 3-minute observation, simply by the addition of a small constant "residual delay" and "residual delay rate". That is to say, all higher order residuals, such as delay accelerations, are assumed to be negligible. Except under extraordinary circumstances, such as occur with observations close to the sun which are highly influenced by the solar corona, this assumption is valid.³

Processing of frequency-switched data through VLBI 2 to find the maximum-likelihood estimates of group delay and phase-delay rate is accomplished in three steps:

1. For each record of data the "counter-rotated video cross-spectrum", $S_{12}^k(\omega_j^V)$, is computed as basically given by Equation (4-4).

2. A "coarse search" in delay and delay rate is made by maximizing the sum of the correlation amplitudes of the indivi-

³If the assumption of negligible higher-order residuals is invalid, it is often possible to break an observation into several shorter time segments over which this assumption is valid. Provided the signal-to-noise ratio remains sufficiently high, each of these shorter data segments may then be processed individually.

dually processed frequency windows. The range of this search is generally about ± 5 microseconds in delay and $\pm \frac{2.5}{N}$ Hz in fringe rate around the a priori estimates, where N is the number of frequency channels. (Since one cycle of switching through the N frequency channels takes N/5 seconds, there are ambiguities in fringe rate spaced 5/N Hz apart.)

3. A "fine search" in delay and delay rate is made, centered around the values found in Step 2, by coherently summing over all frequency channels to make use of the full precision of the switched-frequency technique. The range of the fine search is typically about ± 0.5 microseconds in delay and ± 10 mHz in fringe rate around the best estimates from Step 2.

4.5-1 Cross-Spectrum Computation

The first step of the VLBI-2 processing is to compute the video cross-spectrum $S_{12}^k(\omega^v)$, as basically given by Equation (4-4), for each record. Seven discrete uniformly-spaced points, spanning $\omega_l^v = 45$ kHz to $\omega_u^v = 315$ kHz, are computed

$$S_{12}^k(\omega_j^v) = \frac{1}{7} \sum_{m=-3}^3 R_{12}^k(\tau_{ap}^{k,+m\Delta\tau_s}) e^{-i\omega_j^v m\Delta\tau_s} e^{-i\phi_{ap}^k} e^{-i\omega_j^v(\tau_{ap}^k - \tau_{ap}^k)}; \quad j = 1, \dots, 7$$

(4-8)

where k is the record number and all other quantities are defined at the end of Section 4.4. The last factor on the right-hand side of Equation (4-8) is the so-called "fractional bit-shift correction", which corrects for the fact that τ_{ap}^k is restricted to an integral number of bit-shifts.

4.5-2 "Coarse Search" in Delay and Delay Rate

All of the processing described to this point has been in preparation for an attempt to detect the signal and make estimates of the delay and delay rates. The first step in further processing is to make a coarse search in delay and delay rate, and then make rough estimates of their values once a signal has been detected. The first step of this coarse search is to compute an approximation to Expression (4-5) for each frequency channel individually

for many trial values of residual delay and delay rates, designated $\Delta\tau$ and $\Delta\dot{\tau}$, respectively. Designate the "coarse search function" for the n^{th} frequency channel as $F_n(\Delta\tau, \Delta\dot{\tau})$, defined by

$$F_n(\Delta\tau, \Delta\dot{\tau}) \equiv \sum_{k=1}^K \sum_{j=1}^7 S_{12}^k(\omega_j^v) e^{-i\omega_0^n \Delta\dot{\tau} (t_k - t_{st}) - i\omega_j^v \Delta\tau} \quad (4-10)$$

where

n = frequency channel number; $1 \leq n \leq N$

ω_0^n = the local-oscillator frequency of channel n

t_{st} = epoch of the reference station at the nominal start time of the observation

t_k = the reference station epoch corresponding to the first bit (not the first correlated bit) of the k^{th} record of the reference station tape;
 $1 \leq k \leq K$,

and where the summation over k includes only those records from frequency channel n . Note that $S_{12}^k(\omega_j^v)$ need not be recomputed for every trial residual delay rate $\Delta\dot{\tau}$ provided that accumulated residual fringe phase "error" over the duration of one record is small (i.e. $|\omega_0^n \Delta\dot{\tau} \Delta T| \ll 1$, where $\Delta T = 0.2$ seconds). The coarse search for the ML estimates of delay and delay rate is completed by maximizing the sum of $|F_n(\Delta\tau, \Delta\dot{\tau})|$ over all N frequency channels, i.e.,

$$\max_{\Delta\tau, \Delta\dot{\tau}} \sum_{n=1}^N |F_n(\Delta\tau, \Delta\dot{\tau})| \quad (4-11)$$

The estimation procedure of (4-11) is sometimes referred

to as an "incoherent search" because the phase information of the frequency channels is not used. It is basically a single-band (360 kHz wide) estimation procedure which has a very broad peak in delay compared to the peak of the multi-frequency delay resolution function (see Figure 1-1). Typically, the search range of $\Delta\tau$ is about ± 4 μsec with trial spacings at intervals of ≈ 1 μsec . For observations of moderate signal-to-noise ratio⁴ the delay can be determined to about ± 0.1 μsec by interpolating to the peak of Expression (4-11). This procedure drastically reduces the range of delays that must be searched in the "fine search", resulting in a considerable savings in computation time.

The coarse search over residual delay rate is, for computational speed and convenience, usually conducted in-

⁴We note, incidentally, that the signal-to-noise ratio of the procedure of (4-11) is reduced by a factor of $\frac{1}{\sqrt{N}}$ from the fully "coherent" estimation procedure that will be performed in the fine search [see Section 4.5-3].

stead as a search over residual fringe rate.⁵ The range of the search is $\pm \frac{2.5}{N}$ Hz quantized in steps of 4.89 mHz. (The summation over k in Equation (4-10) is done by a 1024 discrete-point FFT with the input spacing being 0.2 seconds, so that the resolution in residual fringe rate after the transform is $1/(1024 \times 0.2) \approx 0.00489$ Hz). A simple parabolic interpolation is used to reach the final "coarse search" estimate of residual fringe rate, and hence of residual delay rate.

⁵In order to save computation time, the summation over k in Equation (4-10) is often done using the so-called discrete Fast-Fourier Transform (FFT). The output of the FFT, in such a case, is the residual fringe rate spectrum for the n th frequency channel. It is convenient to compute an approximation to the summation in Expression (4-11) simply by performing an FFT for each of the frequency channels, then summing the amplitude of each residual fringe rate "box" over the N frequency channels and interpolating to the peak of this "summed spectrum". The residual fringe rate of each frequency channel is proportional to the radio frequency of that channel, so that the residual fringe rate is slightly different from channel to channel, and the peak of the "summed" spectrum may be somewhat smeared and/or broadened as a result. Provided, however, that the total fractional bandwidth spanned by the highest and lowest frequency channels is small and that the residual delay rate is also small, this broadening is not serious and this estimation procedure is suitable for the "coarse search" estimate of delay rate.

Estimation of the group delay with full multi-channel precision can proceed at this point only if the coarse search estimate of the group delay is accurate to within a "major ambiguity" spacing of the delay resolution function, which for the October 1969 experiment is 1 microsecond (see Figure 1-1). Otherwise, the final determination of the multi-channel group delay has a high probability of being wrong by one or more ambiguities.⁶

4.5-3 "Fine Search" in Delay and Delay Rate

At the conclusion of the coarse search, the delay and delay rate have been estimated to approximately the limit of single-band accuracy, and, for most purposes, the delay information in the individual channel cross-spectra has been nearly exhausted. It is therefore computationally convenient (i.e. time-saving) to perform a preliminary summation over the cross-spectrum of each record of data, fixing the delay at the coarse search estimate, before proceeding with the multi-band fine delay search. If we designate the coarse search estimate of delay and delay rate as $\Delta\tau$ and $\Delta\dot{\tau}$, respectively, then the summation, to be performed for the k^{th} record is defined by

⁶In case such ambiguities do occur, they can usually be detected by the final source-position and station-location fitting program, VLBI 3, where the postfit residuals stand out in a characteristic manner if the wrong ambiguity has been chosen.

$$D_k(\Delta\hat{t}, \Delta\hat{\tau}) \equiv \frac{1}{7} \sum_{j=1}^7 S_{12}^k(\omega_j^v) e^{-i\omega_j^v \Delta\hat{t}} e^{-i\omega_0^k \Delta\hat{\tau} t_k} \quad (4-12)$$

The amplitude and phase of $D_k(\Delta\hat{t}, \Delta\hat{\tau})$ are the correlation amplitude and residual phase for the k^{th} record, given the coarse search estimates $\Delta\hat{t}$ and $\Delta\hat{\tau}$. The amplitude and phase of $D_k(\Delta\hat{t}, \Delta\hat{\tau})$ are very weak functions of small changes in $\Delta\hat{t}$. The amplitude varies with $\Delta\hat{t}$ approximately as the broad envelope over the delay resolution function in Figure 1-1b. The phase error of $D_k(\Delta\hat{t}, \Delta\hat{\tau})$ due to a small error in $\Delta\tau$ is approximately the phase error at mid-video-band, which, for an error of 0.1 μsec at 180 kHz mid-band, amounts to a ~ 7 degrees of phase. This phase error will be common to all records (and hence all frequency channels) and hence will have a negligible effect on the fine delay search, which depends primarily on the phase as a function of frequency window.

We are now almost prepared to compute the full multi-channel delay function. First, however, we must take into account that the difference phase between the local oscillators of the two stations usually varies from channel to channel, although it remains fairly constant for any given channel as long as the instrumentation is operating smoothly and without interruption at both sites. These difference phases, which we shall designate $\Delta\phi_n''''$ are usually known as "oscillator correction phases", or "phase-calibrator correction phases" and must be determined (see Section 7.3-2) and applied as a correction to the fringe phase of each channel.

With the knowledge of $D_k(\hat{\Delta\tau}, \hat{\Delta\dot{\tau}})$ and ϕ_n^{osc} we may now construct the full multi-channel maximum-likelihood estimation function, designated $D(\Delta\tau', \Delta\dot{\tau}')$,

$$D(\Delta\tau', \Delta\dot{\tau}') = \frac{1}{K} \sum_{k=1}^K \left\{ e^{-i\omega_0^n \Delta\tau'} \sum_{k=1}^K D_k(\hat{\Delta\tau}, \hat{\Delta\dot{\tau}}) e^{-i\omega_0^n \Delta\dot{\tau}' (t_k - t_{st})} \cdot e^{-i\Delta\phi_n^{osc}} \right\} \quad (4-13)$$

where $\Delta\tau'$ and $\Delta\dot{\tau}'$ are the trial residual delay and delay rates relative to $\hat{\Delta\tau}$ and $\hat{\Delta\dot{\tau}}$, respectively, and where the superscript n is, by implication, a function of the record number k . The final estimates of delay and delay rate are made by maximizing $|D(\Delta\tau', \Delta\dot{\tau}')|$ over $\Delta\tau'$ and $\Delta\dot{\tau}'$. Typically by maximizing $|D(\Delta\tau', \Delta\tau')|$ over $\Delta\tau'$ and $\Delta\tau'$. Typically, a coarser search with spacings of about 10 nanoseconds in delay and 5 millihertz in fringe rate (corresponding to about 1 picosec/sec at 8 GHz) is made, followed by a finer search with spacings of 1 nanosecond and 1 millihertz. The final estimates of $\Delta\tau'$ and $\Delta\dot{\tau}'$, which we shall designate as $\hat{\Delta\tau}'$ and $\hat{\Delta\dot{\tau}}'$, are then made by a simple parabolic inter-

polation to the maximum of $|D(\Delta\tau', \Delta\dot{\tau}')|$. The final estimates of the total delay and delay rate, designated $\hat{\tau}_0$ and $\hat{\dot{\tau}}_0$, and referred to the epoch t_{st} of the reference station as defined after Equation (4-10), are

$$\hat{\tau}_0 = \tau_0 + \Delta\hat{\tau} + \Delta\hat{\tau}' \quad (4-14)$$

$$\hat{\dot{\tau}}_0 = \dot{\tau}_0 + \Delta\hat{\dot{\tau}} + \Delta\hat{\dot{\tau}}' \quad (4-15)$$

where τ_0 and $\dot{\tau}_0$ are, respectively, the a priori estimates of the total delay and delay rate computed (by VLBI 1) for the epoch t_{st} . The final estimate of the correlation amplitude is just $\hat{\rho} \equiv C|D(\Delta\hat{\tau}', \Delta\hat{\dot{\tau}}')|$, where C is a factor necessary to correct for the amplitude biases introduced by the processing algorithms.⁷ The final estimate of residual fringe phase, designated $\Delta\hat{\phi}$ and referred to the radio frequency ω_1 , is just the phase of $D(\Delta\hat{\tau}', \Delta\hat{\dot{\tau}}')e^{i\omega_1\Delta\tau'}$. The final estimate of the total fringe phase, referred to frequency ω_1 and epoch t_{st} , is then

$$\hat{\phi}_0 = \phi_0 + \Delta\hat{\phi} \quad (4-16)$$

where ϕ_0 is the a priori estimate of total fringe phase at epoch t_{st} . $\hat{\phi}_0$ is, of course, generally ambiguous by several 2π rotations since it is not usually possible to determine $\hat{\tau}_0$ with enough accuracy to resolve the fringe ambiguity.

⁷C includes the Van Vleck bias discussed in Footnote 2 of this chapter as well as the biases which will be discussed in Section 4,5.

It is interesting to examine $D(\Delta\tau', \Delta\hat{\tau}')$ near its maximum as a function of $\Delta\tau'$. For this purpose let us set $\Delta\hat{\tau}' = 0$ and $\phi_n^{\text{osc}} = 0$ so that we may write

$$D(\Delta\tau', 0) = \frac{1}{K} \sum_{n=1}^N \left\{ e^{-i\omega_0^n \Delta\tau'} \sum_{k=1}^K D_k(\Delta\hat{\tau}, \Delta\hat{\tau}') \right\} \quad (4-17)$$

where the inner summation includes only records from channel n .

Accordingly, let us define the complex correlation coefficient for channel n

$$\rho'_n \equiv \sum_{\text{channel } n} D_k(\Delta\hat{\tau}, 0) \quad ((4-18)$$

We may then write the magnitude of $D(\Delta\tau', 0)$ as

$$|D(\Delta\tau', 0)| \propto \left| \sum_{n=1}^N \rho'_n e^{-i(\omega_0^n - \omega_0^1) \Delta\tau'} \right| \quad (4-19)$$

where we have included a constant factor $\exp\{+i\omega_0^1 \Delta\tau'\}$ which does not affect $|D(\Delta\tau', 0)|$. The form of Equation (4-19) lends itself to a simple pictorial interpretation where the complex correlation coefficient, ρ'_n , of each channel may be considered as a simple vector. Consider the case where $|D(\Delta\tau', 0)|$ is maximized precisely at $\Delta\tau' = 0$. Then, if all instrumental phases have

been properly calibrated and removed, and if we ignore noise for the moment, the vectors ρ'_n will all have the same residual phase, ϕ_{res} , as shown in Figure 4-1a, with $|D(0,0)| = \frac{1}{K} \sum_{n=1}^N |\rho'_n|$. As $\Delta\tau'$ moves away from zero, the vector corresponding to channel n rotates by an angle $-(\omega_0^n - \omega_0^1)\Delta\tau'$. Figure 4-1b shows a pictorial representation of Equation (4-18) for $\Delta\tau' = 10$ nanoseconds with 6 channels placed at 0, +1, +4, +6, +24, +36 MHz with respect to Channel 1 (this is the actual set of frequencies used for the October 1969 experiment). We see that $|D(10 \text{ nsec}, 0)|$ is considerably reduced from $|D(0,0)|$ and that the phase angle has changed. Figure 4-1c shows $|D(\Delta\tau', 0)|$ as a function of $\Delta\tau'$ covering a range of 1 microsecond, which is the spacing of the "major ambiguity" for this case where the lowest common denominator among the frequency spacings is 1 MHz.

We should note, in regard to Figure 4-1c, that the so-called "major ambiguities" have the same amplitude because of the approximation made in Equation (4-13) that $D_k(\Delta\hat{f} + \Delta\tau', \Delta\hat{\tau} + \Delta\dot{\tau}') \approx D_k(\Delta\hat{f}, \Delta\hat{\tau})$. If this approximation were not made, i.e. if $D_k(\Delta\hat{f} + \Delta\tau', \Delta\hat{\tau} + \Delta\dot{\tau}')$ were substituted for $D_k(\Delta\hat{f}, \Delta\hat{\tau})$ in Equation (4-13), the effect would be to multiply $|D(\Delta\tau', 0)|$ of Figure 4-1c by a $\frac{\sin x}{x}$ -type envelope function produced by the finite 360 kHz bandwidth

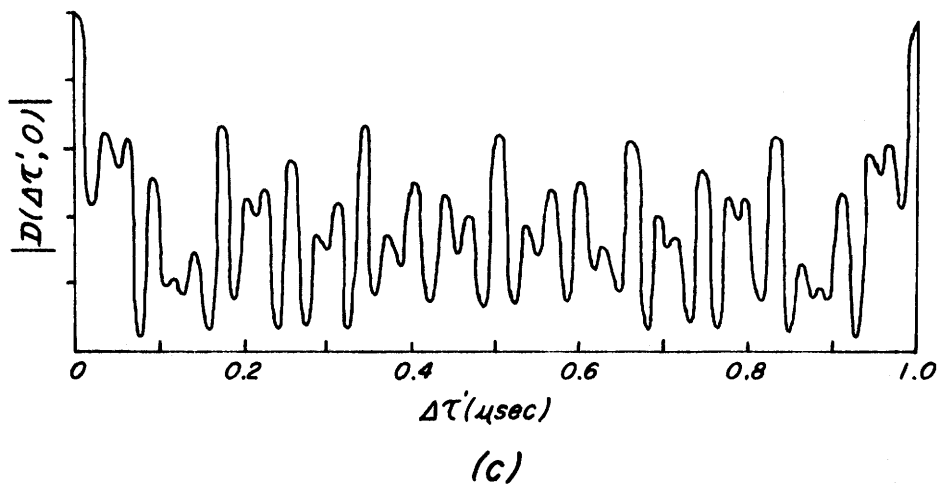
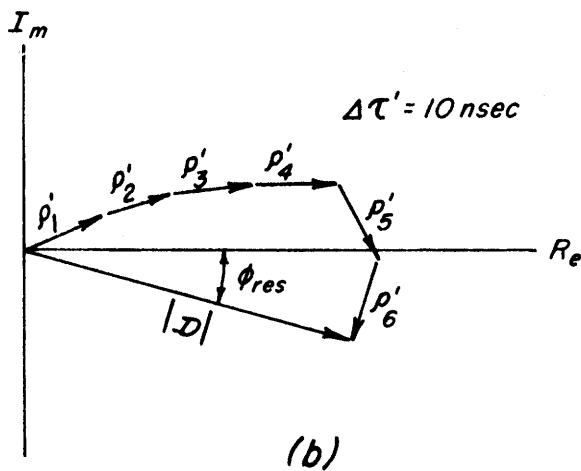
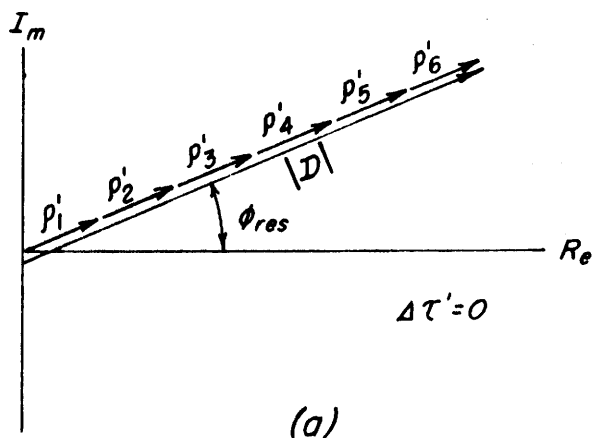


Figure 4-1

(a)-(b) Pictorial representation of Equations (4-18) & (4-19) for $\Delta\tau' = 0$ and $\Delta\tau' = 10$ nanoseconds.

(c) $|D(\Delta\tau', 0)|$ as a function of $\Delta\tau'$.

of the individual channels, as shown in Figure 1-1b, where the assumption has been made that the correlation amplitude is constant across the 360 kHz bandwidth of each channel. The reason for making the approximation used in Equation (4-13) is to vastly decrease the computation time required to compute the many trial values of $D(\Delta\tau', \dot{\Delta\tau}')$. As long as the coarse search estimates, $\Delta\hat{\tau}$ and $\dot{\Delta\hat{\tau}}$, are close to the proper delay ambiguity, as is usually the case, there is no need to perform all the extra computations.

At the conclusion of VLBI 2 processing, the estimates of delay, delay rate, correlation amplitude, and all other relevant parameters are punched onto cards to await further processing.

4. 6 ANALYSIS OF POSSIBLE SYSTEMATIC ERRORS DUE TO DATA-PROCESSING ALGORITHMS

Approximations made at various stages in VLBI 1 and VLBI 2 may contribute small systematic biases to the final estimates of the observables. As the quality of VLBI data improves over the years, it is important to know, and keep in mind, the magnitude of all these possible biases. In this section we shall discuss the approximations made in VLBI 1, and VLBI 2 examine their effects, and try to establish the limits of their applicability.

For convenience, each approximation and its consequences will be discussed separately. Also included in this discussion will be "other problems" which cannot properly be classed as "approximations", but which also may cause some bias in the estimates of the observables.

Approximation No. 1 - VLBI 1 computes the cross-correlation function for only seven discrete delays, centered around the a priori total delay.

Consequences - We may analyze this problem with sufficient accuracy by considering the ideal case of identical square bandpasses of bandwidth ω_0 at the two stations. Furthermore, suppose that the receivers are noiseless and that the actual group delay, τ , is zero. Then the cross-correlation function has the analytic form

$$R_{12}(\tau) = \frac{\sin \omega_o \tau}{\omega_o \tau} \cdot \frac{\omega_o}{\tau} \quad (4-20)$$

Now, suppose that we don't know $\tau = 0$, and are instead trying to estimate the delay τ using the algorithms of VLBI 1 and VLBI 2. Let τ_{ap} be our a priori estimate of τ ; τ_{ap} is presumably not equal to zero. Then $R_{12}(\tau)$ is computed for seven trial delays centering around τ_{ap} , and the cross-spectrum, $S_{12}(\omega_j^V)$, of Equation (4-8) is then computed as the discrete sum

$$S_{12}(\omega_j^V) = \frac{1}{7} \sum_{m=-3}^{+3} \frac{\sin \omega_o (m\Delta\tau_s + \tau_{ap})}{\omega_o (m\Delta\tau_s + \tau_{ap})} e^{-i\omega_j^V (m\Delta\tau_s)} \quad (4-21)$$

$$; j = 1, \dots, 7$$

where $\Delta\tau_s = \frac{\pi}{\omega_o}$ is the sample-bit spacing. The frequencies ω_j are customarily chosen to be 45, 90, 135, 180, 225, 270, and 315 kHz. [In Equation (4-21) we have neglected the fact that τ_{ap} generally will not fall on an integral-bit shift, and that a "fractional-bit correction" will have to be made. In a real observation the delay is constantly changing from record to record, so that any biases due to the fractional-bit correction will average very close to zero and may be neglected.] Suppose that we have calculated $S_{12}(\omega_j^V)$ of Equation (4-21) for this simple case. Then our estimate of delay, designated $\hat{\tau}$, will be found by computing (see Equation 4-11)

$$\max_{\tau'} \left| \sum_{j=1}^7 S_{12}(\omega_j^v) e^{-i\omega_j^v \tau'} \right| \quad (4-23)$$

Three questions arise:

1. How does $\hat{\tau}$, which should be zero by definition, vary as a function of τ_{ap} ?
2. How does the phase of (4-23), which should be zero by definition, vary as a function of τ_{ap} ?
3. How does the amplitude of (4-23) vary as a function of τ_{ap} ?

Question No. 1 may be answered empirically by computing (4-23) as a function of τ_{ap} . The results of such a calculation are shown in Figure 4-2a. Question No. 2 is answered simply by computing the phase of (4-23) as a function of τ_{ap} and is plotted in Figure 4-2b. Question No. 3 may be similarly answered by computing the amplitude of (4-23) as a function of τ_{ap} ; the results are shown in Figure 4-2c.

It is important to note that the results of Figure 4-2a apply only to group delay estimation over a single frequency channel. The estimate of switched-frequency group delay will not be biased because the phase (Figure 4-2b) of every channel is affected similarly. The

switched-frequency estimates of total and residual fringe phases will, however, be biased by the small amounts shown in Figure 4-2b. The effect on correlation amplitude of Figure 4-2c will, of course, be common to both switched and unswitched observations. Finally, we should note that the biases shown in Figure 4-2 are due only to inaccuracies in the a priori model. If desired, the a priori model may be upgraded and the data reprocessed, but this is not generally necessary.

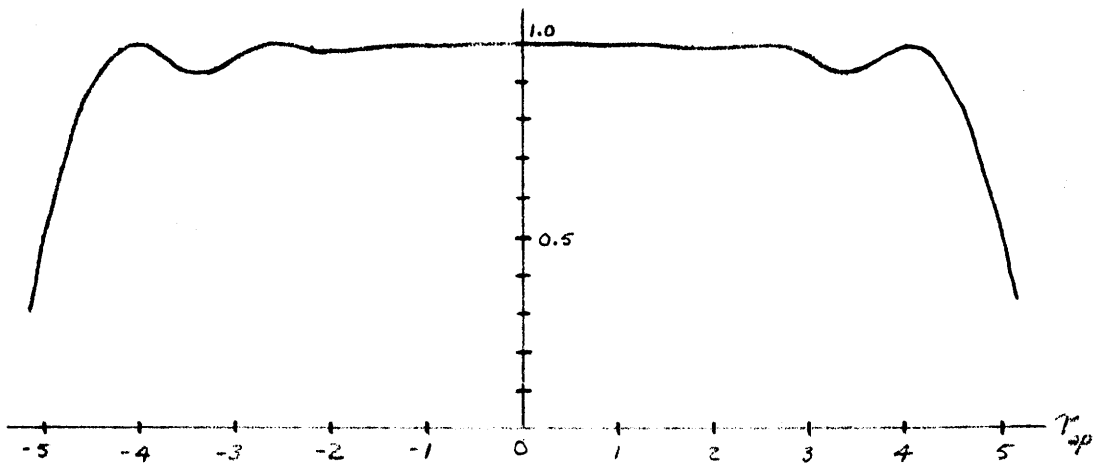
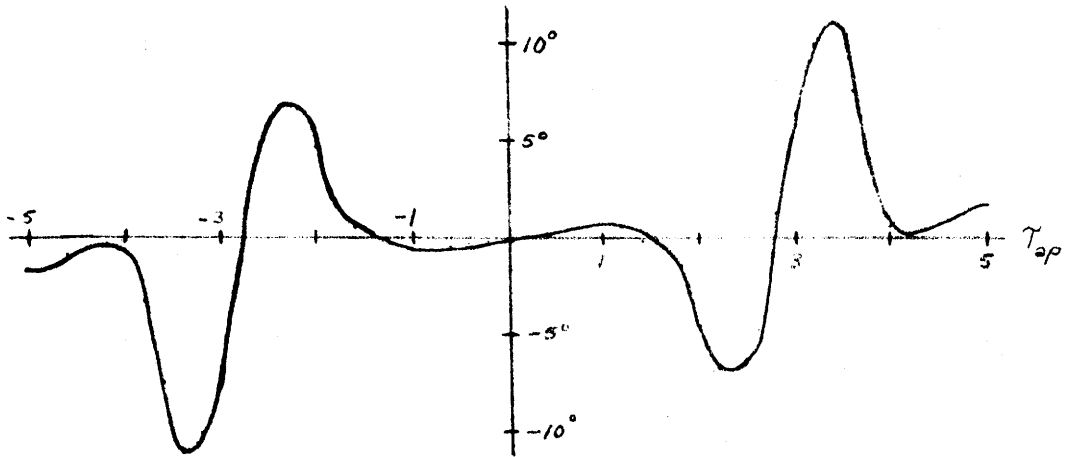
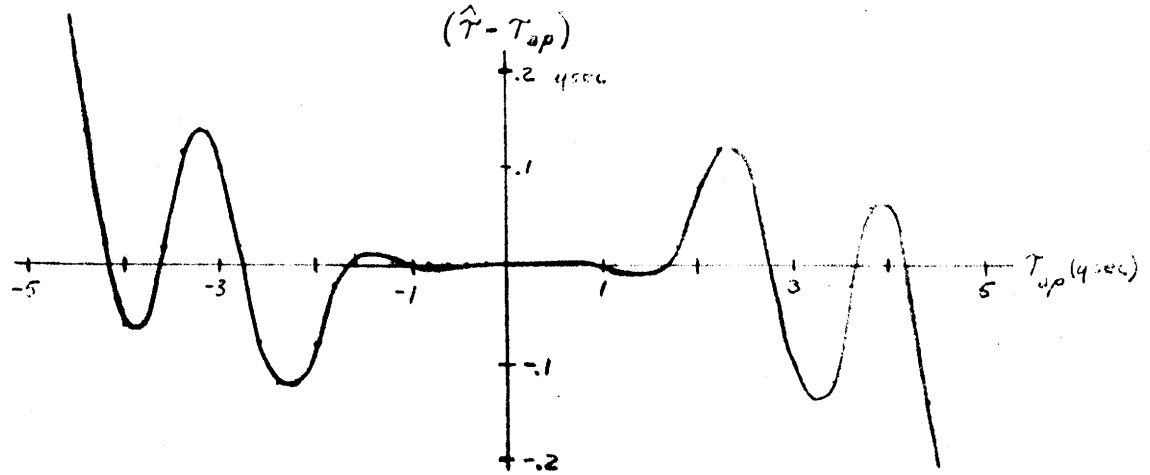


Figure 4-2
Consequences of Approximation No. 1

Approximation No. 2 - The a priori delay is calculated for the epoch corresponding to the middle of the record at the reference station, and the bit shift determined by this delay is held constant over the duration of the record.

Consequences - Because the delay is actually changing somewhat over the duration of the record, some loss of coherency, or "phase smear", will occur. This smear will result in a somewhat lowered correlation amplitude and a possible phase bias. We can easily estimate the maximum possible phase smear by examining the case of maximum expected delay rate. The maximum delay rate encountered in any real VLBI experiment is $\dot{\tau} \approx 2 \mu\text{sec}/\text{sec}$, so that the maximum delay change over the 0.2-second duration ΔT of a single record is $\dot{\tau}\Delta T \approx 0.4 \mu\text{sec}$. The resulting accumulation of differential phase error over the record at video frequency ω_v will then be about $\omega_v \dot{\tau}\Delta T$; for $\omega_v = 360 \text{ kHz}$, $\dot{\tau} = 2 \mu\text{sec}/\text{sec}$ and $\Delta T = 0.2 \text{ sec}$, the phase smear is about 50° . However, since the phase is corrected with respect to the delay at the center of the reference station record, the phase smear actually accumulates to only about $\pm 25^\circ$ (at $\omega_v = 360 \text{ kHz}$) at the beginning and end of the record and is very nearly sym-

metric about the center of the record.⁸ The phase bias is therefore negligible and, in any case, common to all channels so that the multi-channel group delay determination will be unaffected.

The reduction in cross-correlation amplitude due to this "phase smear" is easily calculated. At any given video frequency, ω_v , the phase smear leads to a reduction in correlation amplitude by a factor of

$$\frac{1}{\omega_v \dot{\tau} \Delta T} \int_{-\frac{\omega_v \dot{\tau} \Delta T}{2}}^{\frac{\omega_v \dot{\tau} \Delta T}{2}} e^{i\theta} d\theta \quad (4-24)$$

where $\dot{\tau}$ is the delay rate and ΔT is the record duration.

⁸We are neglecting here the fact that, due to the shifted bit strings, the actual data correlated for a given record are not exactly centered around the middle of the reference station record. The effect is to produce a slight asymmetry in the phase smear around the center of the reference station record, so that a slight phase bias may be introduced. Since the maximum asymmetry of actual correlated data about the center of the reference station record is about 10^4 μ sec, the maximum phase bias due to the effect, which will be common to all channels, is a few degrees at most.

Then, if the actual correlation amplitude is independent of ω_v , the correlation amplitude taken coherently over the video bandwidth ω_o is reduced by the factor

$$\frac{1}{\omega_o} \int_0^{\omega_o} d\omega_v \frac{1}{\omega_v \dot{\tau} \Delta T} \int_{-\frac{\omega_v \dot{\tau} \Delta T}{2}}^{\frac{\omega_v \dot{\tau} \Delta T}{2}} d\theta e^{i\theta} \quad (4-26)$$

This integral may be evaluated by noting that the inner integral is symmetric about $\theta = 0$ so that $e^{i\theta}$ may be replaced with $\cos \theta$. Then, for $\theta \ll 2\pi$, $\cos \theta \approx 1 - \frac{\theta^2}{2}$ and (4-26) may be evaluated directly to yield approximately

$$1 - \frac{\omega_o^2 \dot{\tau}^2 \Delta T^2}{72} \quad (4-27)$$

Substituting the "worst case" $\dot{\tau} = 2 \mu\text{sec}/\text{sec}$, and with $\omega_o = 360 \text{ kHz}$ and $\Delta T = 0.2 \text{ sec}$, we find a value of ~ 0.989 , indicating a 1% reduction in correlation amplitude.

Approximation No. 3 - The a priori delay rate is calculated for a time corresponding to the middle of the record at the reference station and the fringe rate determined from this delay rate is held constant over the duration of the record.

Consequences - There are two sources of possible systematic bias due to this approximation. As we shall see, only the estimate of correlation amplitude is substantially affected by them; group delay estimates are negligibly affected.

Firstly, since the a priori calculation of fringe rate is not exactly correct, a phase error will develop over the duration of a record. For the moment, assume that the actual fringe rate is constant over a record. Then any error in the a priori fringe rate will lead to a phase error which develops linearly over the duration of a record. If the center of correlated data (for this record) is used as a reference epoch for this record, the accumulation of phase error will be symmetric around this reference, and the resultant estimate of fringe phase will be unbiased. In VLBI 2, however, the a priori fringe phase is started at zero at the beginning of each record [see Equation (4-6)]

and then is later rotated to the total a priori fringe phase calculated for a time corresponding to the beginning of the record [see Equation (4-8)]. Thus the final estimate of fringe phase for that record will be biased by an amount equal to half the fringe phase error developed over the record. Suppose the a priori estimate of fringe rate is in error by $\Delta\omega$. Then for a record of length ΔT , the error in the estimate of fringe phase (for that record) will be $\Delta\omega\Delta T/2$. If $\Delta\omega$ remains relatively constant over an entire observation (see

Section 4.5), the final estimate of fringe phase for the observation will be biased by $\Delta\omega\Delta T/2$. Usually $|\Delta\omega| \approx 1$ Hz so that the largest fringe phase error due to the cause is about 36° . Attendant to this phase error will be a "phase smear" and a corresponding reduction in correlation amplitude which is exactly analogous to the "phase smear" analyzed in Approximation No. 2. Correlation amplitude will be reduced by a factor of approximately

$$\frac{1}{\Delta\omega\Delta T} \int_{-\frac{\Delta\omega\Delta T}{2}}^{+\frac{\Delta\omega\Delta T}{2}} e^{i\theta} d\theta \approx 1 - \frac{\Delta\omega^2 \Delta T^2}{24} \quad (4-28)$$

For $\Delta\omega = 1$ Hz and $\Delta T = 0.2$ sec this factor is about 0.934, a reduction in correlation amplitude of about 7%.

How will this error in a priori fringe rate affect the estimate of multi-frequency group delay? The estimate of group delay will, of course, be unaffected if the phase biases are common to all frequency channels. This will not quite be the case, however, since $\Delta\omega$ is proportional to the radio frequency of each particular channel. If the total frequency spread of the sampled bands is ΔF around some mean frequency F , the total spread in $\Delta\omega$ between the highest and lowest radio frequency will be approximately $\frac{\Delta F}{F} \Delta\omega$, apart from any dispersive effects such as the ionosphere, etc.

The corresponding spread in phase biases over the sampled channels will then be approximately

$$\Delta\phi_{\text{bias}} \approx \left(\frac{\Delta F \Delta\omega}{F}\right) \left(\frac{\Delta T}{2}\right) \quad (4-29)$$

and the corresponding bias in the estimate of the group delay will then be approximately

$$\Delta\tau_{\text{bias}} \approx \frac{\Delta\phi}{\Delta F} \approx \frac{\Delta\omega \Delta T}{2F} \quad (4-30)$$

For $F = 8000$ MHz, $\Delta T = 0.2$ sec, and $\Delta\omega = 1$ Hz we find $\Delta\tau_{\text{bias}} \approx 12.5$ picosec, which is entirely negligible.

Again, we should emphasize that the biases indicated in Equations (4-28) and (4-30) are due to inadequacies of the a priori model and may be reduced to an arbitrarily small value simply by improving the model. These biases are in no way inherent in the data.

The second possible source of bias due to Approximation No. 3 results from the assumption that the fringe rate is constant over the duration of a record, when in fact it is not constant. Some phase smearing will result, regardless of the accuracy of the a priori fringe rate calculation. The maximum rate of fringe rate change on an intercontinental baseline at an 8000 MHz radio frequency is about 0.5 Hz/sec. This is equivalent to a change of 0.1 Hz over a record, which corresponds to an "average"

fringe rate error over a record of about 0.05 Hz if the a priori calculation of fringe rate at the center of the record is exactly correct. The resultant phase smearing over a record is less than 2° which is usually negligible. This smearing is a much much smaller effect than that due to an error in the a priori fringe rate; hence, the analysis resulting in equations (4-28) and (4-30), where we assumed a constant fringe rate over a record, is justified.

Approximation No. 4 -- When performing the "fringe rotation" of Equation (4-6), VLBI 1 updates the fringe phase in a step-wise fashion once every 16 bits (corresponding to ≈ 22.2 microseconds of data).

Consequences -- The first consequence of this approximation is a "phase smear" exactly analogous to that of Approximation No. 3. Let $\dot{\tau}$ be the delay rate, ω be the radio frequency, and $\Delta t \approx 22.2$ microseconds be the time interval over which the fringe phase is held constant. Then there will be a phase smear of $\pm \frac{\omega \dot{\tau} \Delta t}{2}$, symmetric around the center of the 16-bit segment. Since all phases are referred to the center of the 16-bit segment [see Equation (4-6)], there will be no phase bias or effect on group delay determination, but there will be a reduction in correlation amplitude by a factor of

$$\frac{1}{\omega \dot{\tau} \Delta T'} \int_{-\frac{\omega \dot{\tau} \Delta T'}{2}}^{+\frac{\omega \dot{\tau} \Delta T'}{2}} e^{i\theta} d\theta \approx 1 - \frac{(\omega \dot{\tau} \Delta t)^2}{24} \quad (4-31)$$

For the "worse case" of $\omega = 8000$ MHz, $\dot{\tau} = 2$ μ sec/sec, $\Delta t = 22.2$ sec, the fringe rate is 16 kHz and the factor of Expression (4-31) is ≈ 0.794 , more than 20% reduction in correlation amplitude! This, then, is a significant bias which must be taken into account. Higher fringe rates suffer increasingly large reductions in correlation amplitude until the useful limit of the algorithm of Equation (4-5) is reached, approximately when the phase change over a 16-bit segment reaches 180°. This limit is at a fringe rate of about 22 kHz, where the correlation amplitude is reduced by almost 40%!

We note here again that the correlation amplitude is also biased due to the fact that a simple Van Vleck correction is applied to the correlation coefficient of each 16-bit segment [see Equations (4-6) and (4-7)]. We can use Equation (4-8) with $K = 16$ to compute that the correlation amplitude is reduced by a factor of about 0.929 due to this effect, more than a 7% reduction! It is important, then, that this factor must also be applied as a correction when computing correlation amplitudes.

Approximation No. 5 -- VLBI 2 estimates only a constant delay residual and a constant delay-rate residual for each 3-minute observation. No effort is made to fit any higher-order terms (see Section 4.5).

Consequences -- Clearly, the effects of this approximation are entirely dependent on the a priori model. If the model is accurate enough so that there is no need to fit higher-order residuals, then the model is sufficient. If, on the other hand, significant higher-order residuals are present, but not estimated, the resulting estimates of delay and delay rate may be biased. Let us examine the models used in VLBI 1 so that we can place upper limits on such biases.

First let us assume only the simple earth-star model analyzed in Chapter 2, and neglect any effects due to instrumentation or propagation (i.e. atmosphere, ionosphere, solar corona, etc.). We can easily calculate the consequence of unestimated higher-order residuals by examining the change in residual fringe rate over the course of a 3-minute observation. From Equation (2-13c) we see that the geometric fringe rate, v_g , is affected the same way by an error in baseline coordinates α_b, δ_b , as by an error in source coordinates α_s, δ_s . Therefore, we shall just consider the effect of an error in source coordinates. Let α_s, δ_s be the actual source coordinates

and let $\Delta\alpha_s$, $\Delta\delta_s$ be the errors, assumed small. Then from Equation (2-13c) we can easily calculate the corresponding error in the total fringe rate to be

$$\begin{aligned} \Delta v_g \approx & - \frac{\omega b}{c} \frac{\partial \alpha_G}{\partial t} \cos \delta_b [\sin \delta_s \sin(\alpha_b - \alpha_s) \Delta \delta_s \\ & + \cos \delta_s \cos(\alpha_b - \alpha_s) \Delta \alpha_s] \end{aligned} \quad (4-32)$$

where ω is the radio frequency, b is the baseline length, and $\frac{\partial \alpha_G}{\partial t}$ is the rotation rate of the earth. It is simplest to consider the effects of $\Delta\alpha_s$ and $\Delta\delta_s$ separately. Suppose that $\Delta\delta_s = 0$, then

$$\Delta v_g \approx - \frac{\omega b}{c} \frac{\partial \alpha_G}{\partial t} \cos \delta_b \cos \delta_s \cos(\alpha_b - \alpha_s) \Delta \alpha_s \quad (4-33)$$

and the time rate of change of Δv_g is

$$\Delta \dot{v}_g \approx \frac{\omega b}{c} \left(\frac{\partial \alpha_G}{\partial t} \right)^2 \cos \delta_b \cos \delta_s \sin(\alpha_b - \alpha_s) \Delta \alpha_s \quad (4-34)$$

which has a maximum magnitude of

$$|\Delta \dot{v}_g|_{\max} \approx \frac{\omega b}{c} \left(\frac{\partial \alpha_G}{\partial t} \right)^2 \Delta \alpha_s \quad (4-35)$$

regardless of the values of α_s , δ_s , α_b , δ_b .

Similarly, if $\Delta\alpha_s = 0$ we find that

$$\Delta \dot{v}_g \approx - \frac{\omega b}{c} \left(\frac{\partial \alpha_G}{\partial t} \right)^2 \cos \delta_b \sin \delta_s \cos(\alpha_b - \alpha_s) \Delta \delta_s \quad (4-36)$$

which has a maximum magnitude of

$$|\Delta \dot{v}_g|_{\max} \approx \frac{\omega b}{c} \left(\frac{\partial \alpha_G}{\partial t} \right)^2 \Delta \delta_s \quad (4-37)$$

Usually, $\Delta \alpha_s$ and $\Delta \delta_s$ are less than about 10 arc-sec.

If we use $\Delta \alpha_s, \Delta \delta_s = 10$ arc-sec in Equations (4-35) and (4-37) with $\omega = 8$ GHz, $b = 4 \times 10^6$ m and $\frac{\partial \alpha_G}{\partial t} = 7.3 \times 10^{-5}$ rad/sec, we find

$$|\Delta \dot{v}_g|_{\max} \approx 1.8 \times 10^{-4} \frac{\text{rad}}{\text{sec}^2} \approx 2.9 \times 10^{-2} \frac{\text{mHz}}{\text{sec}} \quad (4-38)$$

Over a 3-minute observation, then, the residual fringe rate may change by as much as about 5 mHz, which corresponds to a delay rate error of about 0.6 picosec/sec.

This error will lead to a bias in the total fringe-rate estimate of about 2.5 mHz since, historically, the reference epoch of an observation is at the beginning of the observation. If the reference epoch were at the center of the observation there would be a far smaller bias since $\Delta \dot{v}_g$ will be very nearly symmetric around the center of observation.

Generally, however, $\Delta \alpha_s$ and $\Delta \delta_s$ are on the order of 1 arc-sec, much smaller than the 10 arc-sec error we assumed above, so that the fringe-rate estimate bias will be

$\lesssim 0.25$ mHz, or $\lesssim 0.03$ picosec/sec in delay rate.

The possible biases that we have just calculated above may or may not be important depending on what other unmodelled systematic effects might contribute even larger values of higher-order derivatives of the residual fringe rate (i.e. delay rate). Some effects, such as instrumental drifts, cannot be adequately modelled and hence are simply left out of the a priori model. Others, such as the neutral (dry) atmosphere, wet atmosphere, and ionosphere are usually modelled with some difficulty, if at all. Of these, the neutral atmosphere is the most important (at least at the high observing frequencies) and the most amenable to simple modelling. If left unmodelled, the neutral atmosphere may introduce a large change in the residual fringe rate during the course of one observation. For example, consider the case of a long east-west baseline in which the source is setting at the easterly antenna but is still relatively high in the sky of the westerly antenna. At the westerly antenna there will be a small delay rate acceleration due to the atmosphere, but the easterly antenna will be looking through a rapidly increasing atmospheric thickness as it approaches the horizon. If the source elevation is $\approx 10^\circ$ at the easterly antenna and decreasing at the rate of about 0.25 deg/min the "atmospheric delay rate" (due to the atmosphere above the easterly antenna alone) will increase by ~ 2 picosec/sec during a single 3-minute observation (Reference 36). At

$\approx 5^\circ$ elevation the increase may be as much as about 10 picosec/sec. Clearly, large systematic biases of delay-rate estimates will occur in such a situation if the neutral atmosphere is left unmodelled.

Early versions of VLBI 1, including the version which was used to process the data from the October 1969 experiment, did not include an atmospheric model. However, mostly due to the fact that two of the three antennas which were used were of the equatorial-mount type with limited hour-angle coverage, most observations were at elevations above 20° , where the maximum bias of the delay-rate estimate is ~ 0.1 ps/sec. Later versions of VLBI 1 now incorporate a semi-empirical analytical a priori model of the neutral atmosphere (Reference 36) and, except perhaps for some extremely low-elevation observations, the biases of the delay rate estimates due to the neutral atmosphere are negligible (i.e. $\lesssim 0.01$ picosec/sec). In any case the atmospheric model may be improved a posteriori in a bootstrap manner to reduce any biases to an arbitrarily small value.

Other Possible Systematic Errors -- There are, of course, other possible sources of systematic error which we have not discussed, but they are all believed to be negligible. These include such things as machine-precision limitations in calculation of the a priori parameters (most critical in VLBI 1) and errors in interpolating to the peak of a discretely-computed function, for example.

In regard to machine precision we note that, at 8 GHz, the calculation of a priori total fringe phase must be precise to about 12 decimal digits in order for the error to be below about 0.1 degrees of phase; this means that all of the intermediate calculations, including the trigonometric functions, must be somewhat more precise if 12 decimal digits of precision are to be maintained in the final result. In this regard, double precision calculations on the IBM 360 system are barely sufficient.

4.7 IMPLEMENTATION AND TESTING OF VLBI 1 and VLBI 2

VLBI 1 and VLBI 2 were written, debugged and tested at the NASA Goddard Space Flight Center during the period from about November 1969 through May 1970.

Almost all of the work was done using the IBM 360 Model 91 machine located at Goddard.

4.7-1 Implementation of VLBI 1 and VLBI 2

Because of the extremely large volume of data to be processed, it was imperative that every effort be made to make programs extremely machine-efficient. This is particularly true of VLBI 1, by far the most time-consuming of the processing operations. Several of the "efficiency-critical" sections of VLBI 1, including the actual bit-by-bit correlation and "fast-fringe rotation", were written in assembly language using high-efficiency algorithms developed specially for this application. The exact algorithms used to obtain efficient bit correlation and fringe rotation are described fully elsewhere (Hinteregger, Reference 13), and we shall not describe them in detail here. All "precision-critical" computations were performed using full machine double-precision (approximately 14 significant decimal digits) to ensure that the effects of computation errors were much smaller than

algorithm-dependent errors. Special care was taken to ensure that a priori parameter estimates varied in a smooth way from observation to observation, as well as within a single observation, so that "residual observables" could be interpreted meaningfully over a set of several observations.

A major factor in increasing the efficiency of VLBI 1 was gained by writing it to simultaneously process all three baselines of a 3-station observation. This complicated the bookkeeping aspects of the program but increased the processing efficiency several-fold over a program only able to process a single baseline at a time. Also as an aid to efficiency, input and output were streamlined as much as possible. All critical a priori parameters such as station locations, source positions, observing frequencies, and clock-offset parameters were permanently stored in large tables and were called up by codes. This not only streamlined the processing-setup procedure but also helped to reduce the likelihood of input parameter mistakes. Several separate bookkeeping and sorting programs were written to help support the library aspects of the project and help expedite the data flow to and from the main processing programs.

When VLBI 1 and VLBI 2 reached their final working versions, they alone contained nearly 7000 punched cards of programming. A 3-minute observation on a single baseline can be processed through VLBI 1 with about 90 seconds

of machine computation time, although the actual time on the machine is longer since a minimum of three minutes is required to physically read an entire tape; processing of a single-baseline observation through VLBI 2 takes about 15 seconds. The corresponding times for a 3-baseline observation are 270 seconds through VLBI 1 and 45 seconds through VLBI 2, although minimum tape reading time through VLBI 1 is increased to 6 minutes because only two tapes may be read simultaneously on the Model 91. Including time to mount and demount tapes on tape drives, each 3-baseline observation typically requires about 10 minutes of physical processing time.

Because of the very large volume of data to be processed, VLBI 1 was also implemented on three other machines at Goddard, two IBM Model 75 machines and one IBM 360 Model 95. The Model 75 is roughly one-third as fast as the Model 91, and the Model 95 is nearly twice as fast as the Model 91.

4.7-2 Testing of VLBI 1 and VLBI 2

Testing of VLBI 1 and VLBI 2 poses somewhat of a special problem because of the difficulty of generating artificial test data which adequately models real data in all respects. Artificial-data tapes of correlated noise may be fairly easily generated, but it is very difficult to simulate a smoothly changing delay of the signal

on one tape with respect to the other. In practice, therefore, one is forced to use real VLBI data to do the final testing of the VLBI data-reduction programs.

Testing of VLBI 1 and VLBI 2 with real data can be broadly broken into two phases: 1) comparison of results with other programs (i.e., those already implemented at Haystack) and 2) self-consistency tests within the programs under test. Comparison of test results with the results from other programs was useful only for single-band data since the available programs at Haystack had been fully tested only for that type of data. When good agreement with these programs was reached, program testing moved into the second test phase, that of internal consistency tests.

The basic premise of the internal consistency tests is that the estimates of the observed parameters should be independent of errors in the a priori model to within the expected noise, except for small possible biases that we have indicated in Section 4.4. These tests are easily performed by slightly changing one or more of the a priori parameters and verifying that the estimates of delay, delay rate, fringe phase and correlation amplitude remain unchanged to within the noise. A further valuable

check of internal consistency is provided by processing data from a 3-station experiment, since simple geometrical constraints require that the sum of delays and the sum of the delay rates around the three baselines meet a simple "closure criterion" (see Section 7.3-5). All of these tests were performed and carefully checked to insure proper operation of the programs.

4.8 BRIEF DESCRIPTION OF VLBI 3

After the completion of processing through VLBI 1 and VLBI 2, the estimates of delay and delay rate for each observation are used as input data to a final parameter estimation program, dubbed VLBI 3, which estimates many model parameters using a least-squares algorithm. Most of the work on this program has been done by others, primarily D. S. Robertson, and we shall discuss it only briefly here. The model parameters that can be estimated by VLBI 3 include 1) the 3-dimensional baseline vectors between the observing stations, where one station is usually defined as an origin of coordinates, 2) the source positions, where the right-ascension origin may be defined by fixing the right-ascension of a single source or, alternatively, a priori covariances may be assigned to the right-ascensions of a set of sources to constrain the weighted mean-square right-ascension adjustments to a minimum, 3) the coefficients of a power series that describe clock synchronizations as a function of time, with provisions for handling clock "breaks" as needed, 4) the zenith excess-electrical path-length due to the atmosphere based on a simple model by Chao (Reference 36), 5) the polar motion of the earth, and 6) a parameter usually denoted by γ , related to the relativistic "bending" of electromag-

netic waves in a gravitational field. Any or all parameters may either be fixed or estimated, although, of course, the data must be sufficient to allow for the solution of a given set of parameters (see Section 2.3-2).

The theoretical model used by VLBI 3 takes into account, in as complete a way as possible, the effects of precision, nutation, polar motion, earth tides, and deviations of UT 1 from atomic time. A crude ionospheric model is also provided, although in practice the ionospheric contributions to delay and delay rate are sometimes estimated a priori and applied as fixed corrections to the observed delays and delay rates. Such estimates may be implied, for example, from observations of the Faraday rotation of radio signals transmitted from a satellite in synchronous orbit. At L-band the corrections due to the ionosphere are typically in the range from about 0.1 to 3 nanoseconds for (differential) delay measurements and from about 0.05 to 1.0 picosec/sec for delay rate measurements. At X-band the corresponding ionospheric corrections are reduced by a factor of approximately 25.

Typically, processing of data through VLBI 3 takes place in a multi-step procedure. First, all parameters except clock synchronization errors are fixed at the best a priori estimates as an aid to identify any obviously bad data which would otherwise badly distort any

more complete set of solution parameters. As the bad points are weeded out (or explained) more and more solution parameters may be used, until the information in the observation set is exhausted. Also, studies are usually conducted to measure the sensitivity of particular parameters to slight changes in the data or solution parameter set so as to estimate the real accuracies of the solution parameters.

CHAPTER 5

DESIGN OF A 3-STATION FREQUENCY-SWITCHED VLBI EXPERIMENT

5.1 GENERAL OUTLINE AND GOALS OF THE EXPERIMENT

In the spring of 1969, with the lessons of three frequency-switched VLBI experiments in hand, planning began for another, more ambitious such experiment to be conducted during the first two weeks of October 1969. Three antennas would be used: the 120' Haystack antenna in Tyngsboro, Mass., the 140' NRAO antenna at Green Bank, W. Va., and one of the 90' antennas at the Owens Valley Radio Observatory near Big Pine, California. As conceived, the experiment had three primary and interrelated goals:

- 1) to determine the three components of the baseline vector for each of the three baselines to an accuracy of the order of one meter, 2) to measure the positions of about a dozen widely separated radio sources to an accuracy of about a tenth of an arc-second, and 3) to measure the gravitational bending of radio waves as they passed near the sun, to be accomplished by monitoring the apparent change in the position difference between 3C273B and 3C279 as the sun passed through their vicinity and occulted 3C279.

If accomplished, all of these measurements would represent significant advances in the application of VLBI to precision geodesy and astrometry. Goals 1 and 2 are, of course, very closely related since the baseline vector components and source positions must all be estimated

simultaneously. Goal 3, however, is relatively independent since it involves observing only an apparent change in the positions of two radio sources and is relatively insensitive to small errors in baseline vectors or source positions.

The decision to use three stations in this experiment was primarily motivated by a desire for data redundancy. Since data from a single baseline is sufficient, with enough observations, to determine all baseline components and source positions, the inclusion of three antennas provides automatic data redundancy. A powerful test of the internal consistency of the data can be performed by solving for baseline components and source positions from data collected on each individual baseline. Source positions should be consistent to within expected errors from baseline to baseline, and the baseline vector components should sum to zero around the three baselines. Any inconsistencies in these respects would lead one to suspect that unmodelled systematic biases existed in the data. The magnitude of these inconsistencies would give a good idea of the real accuracy of the data. These tests are important because no independent non-VLBI method presently exists against which checks could otherwise be made.

5.2 CHOICE OF RF OBSERVING FREQUENCIES

In order to attempt to measure and correct for the frequency-dependent effects of the ionosphere and solar corona and to separate these effects from the frequency-independent effects of the gravitational bending, observations were scheduled to be made at two widely-separated basic observing frequencies. The first frequency was chosen to be near 7800 MHz (X-band) at a wavelength of ~ 3.8 cm and the second was chosen to be near 1600 MHz (L-band) at a wavelength of ~ 19 cm.

Due to the strong variability of both the ionosphere and solar corona, it was possible in advance only to make rough estimates of their expected effects. Day-time ionospheric contributions to the radio path length in the zenith direction were estimated to be as large as ~ 1.5 nanoseconds at X-band and ~ 30 nanoseconds at L-band (Reference 37). Depending on the exact conditions during an observation (projected baseline, ionospheric irregularity structure, position of day-night boundary, etc.), the effect of the ionosphere on the VLBI group-delay measurements could be widely varying up to those approximate limits. Likewise, the effect of the solar corona could only be roughly estimated. For a ray path with closest solar approach equal to 10 solar radii, the refraction due to the average solar corona was estimated to be ~ 0.01 arc-seconds at X-band and ~ 0.25 arc-seconds at L-band. For a 1000 km projected baseline, these

deflections correspond to delay changes of approximately 0.4 nanoseconds at X-band and 10 nanoseconds at L-band. As would be expected, the refraction is a strong function of the close-approach radius to the sun.

The gravitational bending effect, unlike the effects of the ionosphere and the solar corona, is independent of both time and frequency. Based on Einstein's predictions, a ray with closest solar approach of 10 radii will be deflected by approximately 0.175 arc-sec, which corresponds to ~7 nanoseconds of delay change on a 3000 km projected baseline. The bending is inversely proportional to the closest approach of a ray to the sun and is approximately 1.75 arc-sec at the limb of the sun. No data existed which could adequately predict the closest distance to the sun for which useful VLBI observations could be made; therefore observations were scheduled right up to the time of the 3C279 occultation and immediately following.

Another reason for the choice of the X-band and L-band frequencies was that the necessary parametric amplifiers and mixers necessary to operate at these frequencies were either in-hand or available for all three sites. And also, not incidentally, the X-band frequency is near the limit of the operating capability of the 90' antenna at OVRO.

Ideally, it would have been possible to monitor X-band and L-band simultaneously, but equipment restrictions prevented this possibility. In fact, a 15-minute changeover

period was necessary at Haystack. Provided that the ionosphere and the solar corona remained relatively constant for periods longer than about 15-30 minutes, this restriction would not be a problem. It was not expected that the ionosphere would pose any particular problems in this respect, but very little, if any, data were available concerning irregularity scale-size statistics for the solar corona for periods of minutes or hours, particularly for scale sizes on the order of the projected baseline lengths.

5.3 CHOICE OF SWITCHED-FREQUENCY WINDOWS

Due to the recent availability of wide-bandwidth IF systems, and in view of the many problems encountered in earlier experiments in which the high-frequency phase-locked first local oscillators were synchronously switched, it was decided not to attempt such a procedure in this experiment. Instead, the first local oscillator would be fixed and the bandwidth of the first IF would be broadened as much as possible; then, only the much lower frequency second local oscillator would be switched (see Figure 1-2). This procedure considerably simplifies the switched-frequency procedure, although available IF-amplifier bandwidths restricted the total "switched-bandwidth" to ≈ 40 MHz.

The choice of the number and spacing of the frequency windows to be observed, as we indicated in Section 1.2, is a compromise between high delay resolution and tolerable sidelobe levels of the delay resolution function. To achieve the highest delay resolution, the choice is clear: choose two frequency windows, placed at the upper and lower extremities of the receiver frequency band. This choice maximizes the quantity $\Delta\omega_{\text{rms}}$ of Equation (3-154) and hence minimizes the expected error in the measured delay. Unfortunately, this choice also creates the worst sidelobe (or ambiguity) problems. Perhaps the sidelobe

problems may be tolerable for a particular set of baselines and/or sources where the a priori knowledge is very good, but unacceptable for others where a priori knowledge is relatively poorer. Since it is undesirable from an operational point of view to use a different set of frequency windows for each baseline and/or source, a single compromise set is usually chosen. One logical choice of a set of frequency windows is a set which has "non-redundancy" in the spacing of the windows, which tends to minimize the sidelobe levels. Such an array is a so-called "Arsac" sequence¹, which has four frequency windows placed at relative frequencies 0, 1, 4, 6. Every unit spacing between 1 and 6 is represented in this sequence and none are duplicated. In practice, it is usually desirable to choose the "unit" spacing to be 1 MHz so that such an Arsac sequence would cover a total of only 6 MHz. In order to span a wider bandwidth some other sequence must be used, but there is no known sequence of more than four frequencies which has non-redundant spacings. Therefore, a "minimally-redundant" sequence of frequencies is usually

¹Named after J. Arsac (Reference), an "Arsac array" originally pertained to antenna spacings in a linear array.

chosen for a sequence of more than four frequencies. Such a sequence may be determined empiracally by a search on a digital computer or, in practice, well-approximated by an educated judgment.

For the October 1969 experiment a choice of six frequency windows was judged to be suitable for the anticipated operating conditions. The set of relative frequencies chosen was 0, 1, 4, 6, 24, 36 MHz. This selection constitutes two overlapping Arsac arrays, one which has a unit spacing of 1 MHz and the other which has a unit spacing of 6 MHz. The arrays have two elements, 0 and 6 MHz, in common. If that the correlation amplitude in each frequency window is the same, the delay resolution function associated with this set of frequencies is the same as that shown in Figure 1.1b. The highest sidelobes have an amplitude of 67% of that of the main lobe, with the spacing to the nearest major sidelobes being ± 1 microsecond. For observations where the a priori knowledge of the group delay is poor, there is a chance that the wrong lobe will be chosen. This was judged not to be a serious problem, however, since fitting of the group-delay data to an earth-star model should easily reveal the existence of such discrepancies and allow them to be corrected, i.e. allow the proper lobe to be chosen. This did, in fact, turn out to be the case.

5.4 EXPECTED SYSTEM PERFORMANCE AND MEASUREMENT ACCURACIES

Based on the specifications of the equipment available for the experiment, estimates were made of the expected performance of the interferometer system as a whole. Table 5-1 shows the a priori estimates of system temperature and antenna efficiencies for each of the three participating stations at each of the two basic observing frequencies. Using the data of Table 5-1 we can calculate the approximate expected correlation amplitudes, signal-to-noise ratios, rms group-delay errors, and rms phase-delay-rate errors from Equations (3-148), (3-151), (3-154), and (3-156). Table 5-2 summarizes the results of these calculations for correlated flux densities of 5 f.u. and 15 f.u., which represent approximately the range of correlated fluxes expected from the set of "major" sources to be observed. These sources include 3C273B, 3C279, 3C84, 3C345, 3C454.3, 4C39.25, 3C120 , 2134+00, and VR042.22.01. It was not anticipated that variations in correlated flux densities due to small-scale source structure would be of significant concern; previous experiments, at least, had not indicated that such a concern might be warranted, although admittedly most of the earlier measurements had generally been made on shorter baselines, or at lower frequencies, or had not been

carefully calibrated to measure fringe visibilities.

If the measurement accuracies stated in Table 5-2 were indeed achieved and the results were free from unknown sources of bias or error, the goals of the experiment should easily have been achieved. Previous experience with earlier frequency-switched experiments, however, had indicated that these numbers were somewhat optimistic, particularly the delay-rate-error estimates, since delay-rate estimation is vitally dependent upon stable local oscillators and these had always proved troublesome. In practice, delay-rate errors were usually dominated by systematic effects which appeared to be mainly due to local-oscillator drifts. Group-delay estimates, on the other hand, are nearly independent of local oscillator drifts provided the drifts are the same for each frequency window. This had usually been the case in practice, and previous experience had indicated that random errors in the group-delay estimates did usually not exceed the theoretical errors by much more than a factor of two or three. Even with these increased errors, provided they are not systematically biased, the goals of the experiment should have been attainable.

Baseline	Freq.	Corr. Amp. (%)		SNR		RMS Delay Error (nsec)		RMS Delay Rate Error (psec/sec)	
		5 f.u.	15 f.u.	5 f.u.	15 f.u.	5 f.u.	15 f.u.	5 f.u.	15 f.u.
NRAO-HAYSTACK	X	0.54	1.62	37	110	.32	.11	.012	.0041
	L	0.49	1.47	33	100	.36	.12	.068	.023
NRAO-OVRO	X	0.22	0.66	15	45	.79	.26	.030	.010
	L	0.34	1.02	23	69	.51	.17	.098	.033
Haystack-OVRO	X	0.17	0.51	12	35	1.0	.34	.037	.013
	L	0.14	0.42	9	28	1.3	.42	.250	.080

Table 5-2

Expected signal parameters based on correlated fluxes of 5 f.u. and 15 f.u. These estimates are based on a (digital) recorded bandwidth of 360 kHz with frequency-window spacings 0, 1, 4, 6, 24, 36 MHz, and an observation length of 160 seconds. They implicitly assume perfect oscillator stabilities and proper antenna pointing.

		X-Band				L-Band			
	Type	Diam (ft.)	T _s (°K)	Effic. (%)	T _a (°K/f.u.)	T _s (°K)	Effic. (%)	T _a (°K/f.u.)	
Haystack	Cassegrain	120	170	35	0.14	250	27	0.10	
NRAO	Prime focus	140	170	45	0.24	110	50	0.26	
OVRO	Prime focus	90	300	20	0.04	400	40	0.08	

Table 5-1

Approximate expected parameters for the individual antenna sites

CHAPTER 6
INSTRUMENTATION FOR EXPERIMENT

6.1 INTRODUCTION

The October 1969 experiment was the first VLBI experiment done by the MIT group which used an antenna at the CalTech Owens Valley Radio Observatory. At the time of conception of the experiment no suitable instrumentation existed at this site for either of the proposed observing frequencies (~7840 MHz and ~1660 MHz). The entire receiving system from feed horns to back-end video converters had to be engineered and constructed in a period of about four months. This project was undertaken by the author in the summer of 1969.

The instrumentation situation at the Green Bank antenna was similar to that at Owens Valley, although at Green Bank appropriate feed horns and a temperature-controlled front-end box were already available. Curt Knight undertook the project to complete all of the Green Bank instrumentation. At Haystack, most of the basic instrumentation already existed since several switched-frequency VLBI experiments at both X-band and L-band had already been done and most of the instrumentation was intact.

Apart from some differences in detail, the receiver systems at each of the three stations were quite similar. Therefore, in this chapter we will discuss only the

OVRO instrumentation. After first examining a basic block diagram of the system, we shall proceed to discuss more detailed design goals, implementation, and testing of the receiver system.

6.2 BASIC SYSTEM DESCRIPTION

Figure 6-1 shows the basic block diagram of the entire VLBI receiving and recording system. Two radio-frequency bands are covered, one from ~1610 to ~1650 MHz (L-band) and the other from ~7790 to ~7840 MHz (X-band). A dual-frequency, concentric-element feed horn collects the energy concentrated at the focus of the dish. The L-band signal is first amplified by an uncooled parametric amplifier, followed by an uncooled tunnel-diode amplifier, before the RF signal is mixed with a phase-stable oscillator signal at 1575 MHz. Similarly, the X-band signal is amplified by an uncooled two-stage parametric amplifier before being mixed with a phase-stable local-oscillator signal at 7875 MHz. The resulting IF bands from both X- and L-band cover a range of approximately 40-80 MHz. The L-band IF signal is from the upper sideband and the X-band IF signal is from the lower sideband. At this point, a switch is used to select either L-band or X-band to be processed through the remainder of the receiver system. The "back-end" part of the receiving system, i.e. the part of the system after the first mixing operation, is

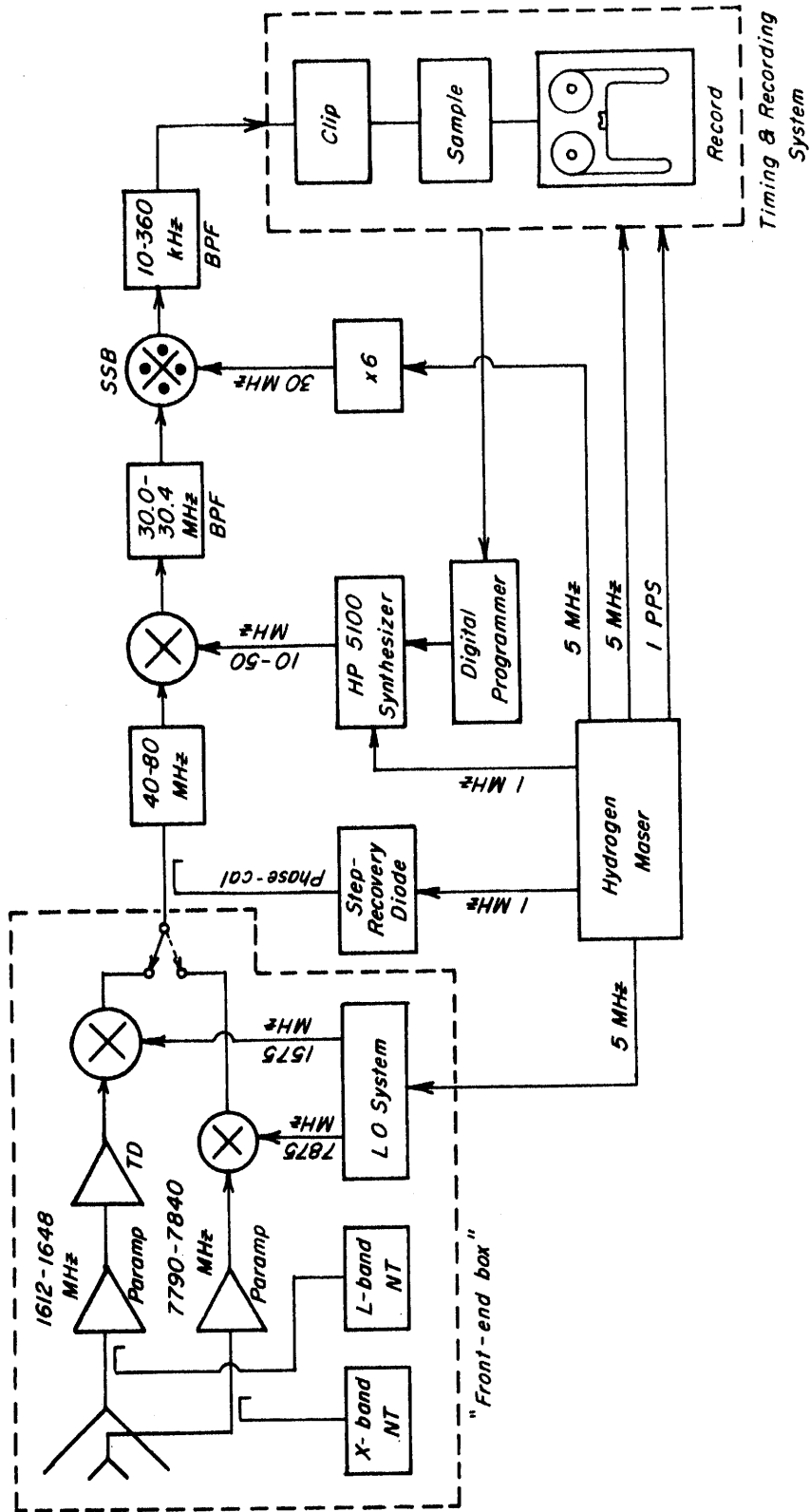


Figure 6-1

Block diagram of OVRO receiving and recording system

common to both the X- and L-band systems. The 40-80 MHz IF signal is upper-sideband-mixed with a programmable oscillator whose output frequency is controllable over the range ~10-50 MHz. The resulting IF signal is filtered through a bandpass filter covering the range ~30.0-30.4 MHz, and a final mixing with 30 MHz takes place in a single-sideband mixer which translates the signal to the 0-360 kHz video band. This video signal is infinitely clipped, one-bit sampled and recorded on magnetic tape.

All local-oscillator signals are derived from a hydrogen-maser frequency standard (and clock) which has a basic output at 5 MHz. The first (high-frequency) local-oscillator signal is derived by locking the frequency of a klystron oscillator, with a phase-lock loop, to a multiple of the maser 5 MHz signal. The programmed local oscillator is a commercially available frequency synthesizer which can be electronically programmed, and which has the special property of returning to the same phase it would have had without frequency switching, after having been switched to another frequency and then switched back again. The output frequency of this synthesizer is controlled by a digital programmer which is in turn controlled by the reading of the time on the hydrogen-maser clock.

A phase-calibration system is used to calibrate the phases of the output frequencies of the programmable

oscillator relative to the hydrogen-maser reference. This calibration is accomplished by injecting a maser-controlled 1 MHz signal into a step-recovery diode which generates extremely short pulses at a 1 MHz rate. The frequency spectrum of this train of pulses is a "comb" of harmonics spaced at 1 MHz intervals spanning ~1-100 MHz. The phase of each harmonic is well-defined (e.g., zero) at every other zero-voltage crossing of the 1 MHz driving waveform, thereby providing a reference against which the phase of each local-oscillator frequency in the programmed sequence may be calibrated. The phase-calibration system was programmed to be activated for only a few seconds at the beginning of each observation. Note that this system does not calibrate the high-frequency (first) local oscillators.

6.3 FEED HORN SYSTEM

In order to be compatible with existing feed-horn systems at Haystack and NRAO, a feed system was required at OVRO which could receive linear polarization at both X-band and L-band. A concentric feed design was chosen so that both X-band and L-band signals could be received simultaneously, if desired, and also so that the antenna pointing would be the same at both frequencies.

Since time was at a premium, it was decided to copy the basic design of an existing dual frequency X-L-band

feed already in use on the NRAO 140' antenna. The focus-to-diameter ratio (commonly called "f/D ratio") of the NRAO antenna is 0.42857 and is nearly the same for the OVRO antenna, so that the same feed horns may be used on both antennas. The only change in design of the feed for the OVRO antenna was to scale the size of the L-band horn from the original design frequency of ~1450 MHz of the NRAO horn to a design centered around ~1630 MHz. The horn was fabricated at the MIT Research Laboratory of Electronics and then tuned and tested by the author in a large microwave-anechoic chamber at the MIT Lincoln Laboratory.

6.4 RF AMPLIFICATION SYSTEM

The choice of an RF amplification system for microwave VLBI depends on several factors including noise figure, gain, gain stability, bandwidth, and cost. The significance of each of these factors depends on the particular requirements of a given experiment, but some general comments can be made regarding each.

The noise figure (i.e., noise temperature) specification depends primarily on the signal-to-noise requirement for observations on the weakest radio sources. This requirement not only depends on the parameters of the station in question, but also on the performance of all other participating stations. For example, the approximate

minimum usable correlation coefficient for a 3-minute frequency-switched VLBI observation is $\sim 0.15\%$. Let us estimate the maximum allowable system temperature T_s at OVRO for the case where the minimum source flux-density to be observed is ~ 4 flux units. At L-band the poorest performing antenna is Haystack where the system temperature is $\sim 250^\circ\text{K}$ and the antenna efficiency such that the antenna temperature T_a is $\sim 0.4^\circ\text{K}$ for a 4 f.u. source. The performance of the OVRO antenna is such that $T_a \sim 0.4^\circ\text{K}$ (based on an antenna efficiency of 40%) for a 4 f.u. source. Therefore, the minimum acceptable system temperature at L-band at OVRO is approximately

$$T_{s_{\min}} \approx \frac{0.4 \times 0.4}{250 \times (.0015)^2} \approx 290^\circ\text{K}$$

Similarly at X-band, Haystack performs more poorly than NRAO and hence dictates the maximum tolerable T_s at OVRO. At Haystack a 4 f.u. X-band source will yield $T_a \sim 0.8^\circ\text{K}$; at OVRO the same source will yield $T_a \sim 0.2^\circ\text{K}$ (based on an antenna efficiency of 25%). We can then easily calculate that the minimum acceptable T_s at OVRO is again $\sim 290^\circ\text{K}$.

The gain of the RF amplifier is important primarily for two reasons. The first is to reduce the contributions to the total system temperature, T_s , of noisy devices, such as mixers, which follow the first stage of amplification. (Recall that the contribution to T_s of any element

M is T_M/G , where T_M is the noise temperature of the element and where G is the total gain preceding element M.) The second purpose of the RF amplifier is to reduce the noise contributions from an unwanted sideband by adjusting the amplifier for high gain over the band of interest and for a low gain over the undesired band which will unavoidably be imaged into the IF after mixing. The difference in gain between these two bands is known as the "sideband rejection" and is typically 10-20 db for the parametric amplifiers used. We should note also that some interferometer systems are double-sideband systems where it is desired to include both sidebands and in which "sideband rejection" is not wanted.¹

The bandwidth requirements of an amplifier vary in importance, obviously, with the goals of a particular experiment. The inherent bandwidth of low-noise microwave amplifiers varies considerably between different types and also depends, in most cases, on the details of adjustment. Maser amplifiers, though very low noise, are generally limited to bandwidths of a few tens of MHz. This limitation is not fundamental, but is usually imposed by problems of relatively narrow bandwidth of the so-called "slow-wave" structures internal to the maser. Parametric amplifiers have instantaneous bandwidths ranging

¹For several reasons, double-sideband systems are not particularly appropriate for VLBI work (see, for example, Reference 13).

from a few MHz to a few hundred MHz, depending on the details of the design. The bandwidth limitations generally stem from the bandwidth limitations of the actual microwave structure design and are not inherent in the amplifying element. Indeed, very wide-band parametric amplifiers may be fairly easily built simply by "stagger-tuning" multiple stages since detuned stages appear nearly transparent (i.e., no loss or gain) to out-of-band signals. Tunnel-diode microwave amplifiers, though generally not competitive with masers or parametric amplifiers for low-noise characteristics, generally have a very large fractional bandwidth. A fractional bandwidth of nearly 50% is not uncommon.

Gain stability of microwave preamplifiers may or may not be important, again depending on requirements of a particular experiment. Good gain stability is primarily important for a VLBI experiment if correlation amplitudes are to be accurately normalized to obtain meaningful fringe amplitudes. Good gain stability is also a very useful, if indeed not absolutely necessary, to aid in determining when the antenna is properly pointed toward a weak source. Dicke switching may be employed to overcome both of these problems, but only at the expense of halving the integration time on a source, and is therefore not generally employed.

After all the technical factors have been considered, cost is the final consideration that must be taken into account. Prices of maser amplifiers and cooled parametric amplifiers tend to start at many tens of thousands of dollars. Uncooled parametric amplifiers generally have a price tag of from several thousand to several tens-of-thousands of dollars. Uncooled tunnel-diode amplifiers are usually priced in the several-thousand-dollar range.

The RF amplifier instrumentation for the OVRO site for the October 1969 experiment was chosen to meet several specific criteria: 1) The operating frequencies had to match those of the other sites, with instantaneous bandwidths of at least 50 MHz; 2) It was preferred, obviously, to keep noise temperatures as low as possible, with 300°K set as a practical maximum; 3) Small physical size was of utmost importance because the entire two-band receiver system had to be fitted into less than 3 cubic ft. of space; and 4) The amplifiers had to be available for delivery and use almost immediately. To be sure, the experiment would not have been undertaken had not equipment meeting the necessary specification been available. Uncooled parametric amplifiers were chosen for both the L-band and X-band preamplifiers. In addition, a following tunnel-diode amplifier was used for L-band. We shall describe these items more fully below.

The L-band parametric amplifier chosen was a two-stage commercial unit with a gain of about 15 db over a bandwidth of ~50 MHz and with a noise temperature specification of about 250°K (although in actual operation it was believed to be more nearly ~600°K). A klystron was employed to generate the pump signal for the paramp. Input and output L-band signals were connected through standard type-N coaxial connectors. The L-band tunnel-diode amplifier was also of commercial manufacture; gain was ~ 10 db over a bandwidth of ~ 300 MHz with a noise temperature of ~1000°K. For a gain of 15 db for the L-band parametric amplifier, the contribution of the tunnel-diode amplifier to the total system temperature was ~32°K.

The X-band parametric amplifier was an uncooled two-stage unit designed and built at MIT Lincoln Laboratory (see Getsinger & Kessler, Reference 39). The amplifier was unique in that it was designed by a computer program which was given the desired specifications of the amplifier and then optimized a practical design to meet these specifications as closely as possible. The amplifiers built from these designs have proven to closely meet the expected performance criteria and, at the time they were built, were among the best available. The units which were obtained for VLBI use had originally been built for use in artificial earth-satellite research,

but had become surplus when they were no longer required for satellite observations. The design center-frequency is 7840 MHz with ~ 20 db gain over a bandwidth of ~ 50 MHz (the gain-bandwidth product depends on the exact details of adjustment). Amplifier noise temperature was ~150°-200°K.

6.5 LOCAL-OSCILLATOR SYSTEM

6.5-1 General Comments

The performance characteristics of the local oscillator system in a VLBI experiment in large part determine the success or failure of an experiment. The performance requirements of the LO system can be qualitatively stated in one word -- stability. Perfect oscillators, of course, do not exist, and we must examine the effects of imperfections in practical systems. All of the imperfections of practical oscillators may be broadly classed as phase noise -- that is, the deviations of oscillator phase from the phase of a perfect oscillator. In the context of VLBI observations made with the Mark I recording system, phase noise may be separated into three categories:

1. Short-term phase noise with a time scale ≈ 0.2 sec (the length of one Mark I record);
2. Medium-term phase noise with a time scale ≈ 3 minutes (the length of a standard observation); and

3. Long-term phase noise with a time scale greater than a few minutes.

We shall briefly examine the consequences of each of these.

Short-term phase noise -- Phase noise of this type will cause a reduction of correlation amplitude (even, conceivably, the complete loss of signal) over one record of data, which is the smallest basic integration time. Consider a simple case where the local-oscillator phase noise, $\Delta\theta$, is Gaussianly distributed with a variance $\sigma_{\Delta\theta}^2$. At any given instant the correlation amplitude will be reduced by a factor of $\cos \Delta\theta$ due to this phase noise. If $\sigma_{\Delta\theta} \approx 1$ radian, as it generally must be, then $\cos \Delta\theta \approx 1 - \frac{\Delta\theta^2}{2}$ and the average reduction in correlation amplitude will be simply $\sim (1 - \frac{\sigma_{\Delta\theta}^2}{2})$. As long as the mean short-term phase noise is zero, it will cause no bias in the estimate of either the group delay or the phase-delay rate.

Short-term phase noise is often very difficult to detect experimentally. It can be detected during actual operations only if a more stable calibration signal exists, which is usually not the case. It is usually nearly impossible to detect short-term local-oscillator phase noise a posteriori because receiver noise almost always dominates the correlation coefficient over integration times shorter than one record.

Medium-term phase noise -- This type of noise, with a time-scale of longer than one record but shorter than the length of an observation, again acts to reduce the correlation amplitude. However, it usually is quite detectable a posteriori on a high signal-to-noise observation where fringe phase is well-defined over a few-second integration. As long as the phase noise is small and independent of frequency channel there will be no effect on the estimate of group delay. However, the estimate of phase-delay rate (derived from the observed fringe rate) may be biased if the mean phase noise over the 3-minute observation is non-zero.

Long-term phase noise -- This type of phase noise, if frequency-channel independent, will tend to bias only the phase-delay rate estimates and will have no effect on group-delay estimates. It can generally be detected only by examining a series of observations on the same source taken over a period of a few hours. Long-term phase noise will show up as an inconsistency between the phase-delay rates and the time development of the group delay (after any frequency-dispersive effects have been removed).

6.5-2 High-Frequency Local Oscillators

Of the various local oscillator signals needed to operate VLBI hardware, the high-frequency oscillator is almost always the most critical. The mildest requirements on its performance for any VLBI experiment are that short-term and medium-term phase noise be smaller than ap-

proximately one radian. And, if phase-delay-rate data are to be used for geodetic or astrometric purposes, comparable long-term stability is demanded. Here again we note that such long-term stability must be defined with respect to the reference source which is ultimately keeping the station time.

High-frequency VLBI local-oscillator systems are generally of one of two types. The first, and less common, type is direct multiplication from a lower-frequency reference source at (generally) 1 MHz, or 5 MHz. There are several problems with this type of system. Any phase noise at the reference frequency will produce greatly magnified phase noise on the local-oscillator signal. Required multiplication factors are often as high as 1000 or more, so that even a very small phase noise on the reference signal will be magnified to unacceptable proportions in the local-oscillator signal. Furthermore, very phase-stable multiplier chains with large multiplication factors are very difficult to build. Further difficulty is often encountered in extracting sufficient power from such a multiplier chain to drive the mixer. Recently, some reference frequency standards have been built with a 100 MHz output available to the experimenter, relieving many of the problems of direct multiplication. At least one regularly-used VLBI station (Goldstone) successfully

uses a direct-multiplication system, multiplying a 100-MHz reference signal to generate a high-frequency local-oscillator signal at 7800 MHz.

A more common type of system used to generate the high-frequency local-oscillator signal is the so-called "phase-locked loop". This is effectively an independent oscillator which is allowed to oscillate freely over a short time period (typically a few milliseconds, $\ll 0.2$ sec, the record length) during which its stability is high, but which is "locked" to an exact multiple of the reference frequency over a longer period of time. In this way the effect of short-term noise on the reference frequency is minimized.

Such a phase-locked loop was used to generate the L-band local-oscillator signal for the October 1969 experiment. A basic block diagram of the system used is shown in Figure 6-2. The 5-MHz reference signal from the hydrogen maser is first multiplied by 9 through a broadband multiplier and then subsequently by 6 to generate a 270 MHz signal. This signal enters a harmonic mixer which effectively multiplies the 270 MHz signal by a factor of 6 and mixes it with a sample of output from the 1575-MHz oscillator frequency. A filter at the output of the harmonic mixer allows only components near 45 MHz to pass (i.e. only those components which have resulted from

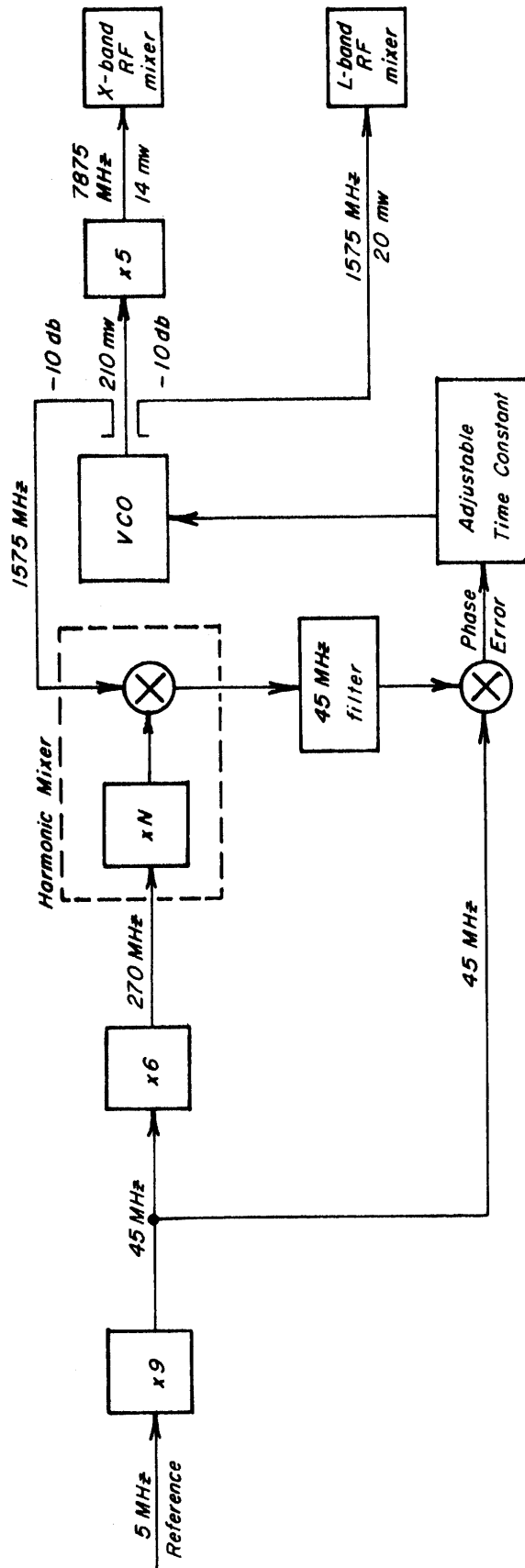


Figure 6-2

Block diagram of local oscillator system

mixing 1575 MHz with the sixth harmonic of 270 MHz). The phases of this signal are compared with the reference 45 MHz signal (by simple mixing) to generate a voltage proportional to their phase difference. This "error" signal is then filtered through a circuit with an adjustable time constant before being applied as a correction signal to adjust the voltage-controlled oscillator (VCO) to the proper frequency. A small part of the output from this oscillator is "leaked-off" to drive the L-band RF mixer, while the major portion of it is used to drive a x5 multiplier to generate the 7875 MHz X-band local oscillator signal. The inefficiency of direct multiplication is well illustrated here; with 210 mw of L-band power into the x5 multiplier, only ~14 mw of useful X-band signal emerges from the multiplier.

In practice, the phase-error signal is passed through a filter with a time constant of several milliseconds before being applied to the VCO in order to eliminate as much as possible the effects of noise picked up on the 5-MHz reference signal on the several-hundred-foot long cable from the control room to the front of the antenna. In the case of the present instrumentation was installed which, in case of a power

oscillator drop-out for any reason, would attempt an automatic search and oscillator-relocking procedure.

Early experiments in switched-frequency VLBI attempted to rapidly switch the high-frequency local oscillator as well as the low-frequency local oscillator. In principle this switching should be possible, but in practice there are considerable technical problems. These primarily stem from the fact that the VCO must be retuned and the phase-locked loop restabilized in less than the 5 millisecond gap between data records. Switching the first local oscillator was attempted in experiments performed in October 1968 and January 1969 but subsequently abandoned when it became possible to avoid the complications of rapid phase-locked-loop switching of the first local oscillator simply by widening the bandwidth of the first IF. The IF bandwidth had been previously limited by the relatively narrow bandwidth of available IF amplifiers, but amplifiers are now available with bandwidths well in excess of 100 MHz so that switching of the first local-oscillator frequency is no longer necessary and has been discontinued.

6.5-3 Low-Frequency Local Oscillators

After the initial signal conversion from RF to IF, the remainder of the local-oscillator signals may be classified as "low frequency". Although phase noise in these oscillators has the same effect on the data as that of the high-frequency oscillators, these low-frequency oscillators are relatively easy to construct and control. In general,

they are responsible for far less phase noise than the high-frequency RF oscillators.

The second local oscillator, as indicated in the block diagram of the system in Figure 6-1, is the frequency-switched oscillator. In principle, the easiest way to switch the oscillator frequency is probably to build one fixed oscillator for each frequency channel and then connect them sequentially to the mixer in the proper time order. In practice this is not difficult to do, but it is an expensive and time-consuming task to design and build all of the necessary oscillators (particularly when the frequency selections seemed to be changing from experiment to experiment). An easier solution is simply to adopt a ready-made commercial unit with the proper characteristics for the job.

The Hewlett-Packard Model 5100 frequency synthesizer was chosen for the job of switched-frequency oscillator. This instrument is a general-purpose synthesizer which can generate any frequency from 0 to 50 MHz in 0.01 Hz steps, allowing almost complete flexibility in the choice of frequency selection. In addition, it has two characteristics which are demanded for its use as a switched-frequency local oscillator. The first is that the output frequency can be electronically selected, allowing it to be controlled by an external digital controller. And secondly, the

internal electronics of the synthesizer are such that, upon switching from one frequency to another and then back again, the phase returns to the same value it would have had, had the frequency never been changed. In other words, changing the frequency of the synthesizer is equivalent to simply connecting the output terminal of the synthesizer to a fixed oscillator of the desired frequency which has been running uninterrupted for all time, regardless of the previous history of frequency selections. The only restriction is that, after the selection of some initial frequency has been made, all further changes must be in exact multiples of 100 kHz. The reason for this restriction lies in the method that is used to generate frequencies lower than 100 kHz within the synthesizer and need not concern us here. In any case this restriction is of no serious concern to us since the minimum frequency step size is usually 1 MHz or more.

There is, however, one cause of concern in the use of the HP5100 synthesizer as the switched-frequency oscillator, a problem which was not fully recognized until several experiments, including the October 1969 experiment, had been completed. This concern relates to a problem of "phase-settling transients" when the device has been set to one frequency for a relatively long period of time (such as the few minutes between observations) and then is suddenly switched to another frequency. If the synthesizer output is compared to an ideal stable oscillator immediately after the synthesizer frequency has been changed, the synthesizer phase may drift through a large fraction of a rotation over

a period of a minute or more before "settling" to a final constant phase. This "settling" process seems to be more or less exponential, most likely indicating internal thermal transients. Because the effects are fairly frequency-independent, it is not believed that group-delay measurements were significantly affected, however phase-delay rates may possibly show some bias due to this problem. Unfortunately, since the problem was not discovered until some time after the October 1969 experiment was completed, it was not possible to reassemble the necessary equipment to make more precise measurements of the effects. More recent experiments have reduced the "phase settling" problem by leaving the HP5100 synthesizers switching continuously even between observations.

The final stage of IF conversion takes place with a local oscillator of 30 MHz which is derived by direct multiplication from the maser 5 MHz signal. Problems of transients or drifts affecting this oscillator are negligible compared to those introduced by the first and second local oscillators.

6.6 FREQUENCY CONVERTERS AND IF SYSTEM

Frequency conversion to video, as shown in Figure 6-1, takes place in three conversion steps. The primary requirement for all the elements in the frequency conversion and IF system is wide-band performance so as to reduce to a minimum any undersirable phase-dispersion

characteristics.

Both the L-band and X-band mixers are of commercial manufacture (both by RHG Electronics Lab., Inc.) and are followed by built-in wide-band IF preamplifiers with bandwidth 10-150 MHz. The noise figures of these mixers are 8.3 db ($\sim 1700^\circ\text{K}$) and 10.5 db ($\sim 3000^\circ\text{K}$), respectively. A 40-80 MHz bandpass filter follows the first stage of mixing in order to prevent the mixing of harmonics of the second local-oscillator with frequencies in the higher part of the 10-150 MHz IF passband.

The second conversion takes place in a standard commercial mixer of bandwidth ~ 10 -100 MHz. This mixing is immediately followed by a narrow-band (30.0-30.4 MHz) image rejection filter.

The final frequency conversion to video is accomplished in a single-sideband converter of unique design. (Rogers, Reference 40). In this converter, the local oscillator is separated into two quadrature components by a quadrature hybrid and mixed with the IF signal. Two very-wide-band 45° phase shifters are then used to form quadrature components before adding. The uniqueness of the design stems from the very wide bandwidth of the 45° phase shifters, allowing the local oscillator to be shifted over a large range and requiring only a low-pass filter to select the video bandwidth.

6.7 RECORDING SYSTEMS

6.7-1 General Comments

Some of the very earliest VLBI experiments employed analog recording systems (References 1 and 2), and in fact, the present-day recording system developed by the Canadians (Reference 3) is basically an analog system. On the other hand, all of the recording systems developed specifically for VLBI in this country, outside of the very early work, are exclusively digital in nature. There are advantages and disadvantages to both techniques, which we shall briefly review.

The advantages of analog recording are primarily the ready availability of low-cost off-the-shelf recorders and the relatively low-cost of constructing a processing terminal. A mild advantage is that the usable data bandwidth, for a given recording bandwidth, is about twice that of the clipped and sampled digital systems. The primary disadvantages of analog recording systems are the difficulty of maintaining time-base stability during playback and, in general, the relative inflexibility of the data-processing procedure. Neither of these disadvantages needs be serious if one is interested only in obtaining fringe amplitudes, delays, and fringe rates over an instantaneous bandwidth of no more than a few MHz, where playback time-base stability need be good to no better than

several tens of nanoseconds in order to obtain results which are not limited by the recording or playback system. Switched-frequency delay-interferometry, on the other hand, demands time-base stability on playback (as well as during recording) of a very small fraction of a nanosecond, which is more difficult to achieve with analog recording systems.

The primary advantage of digital recordings for delay interferometry is that the playback time-base stability is absolutely faithful to the recording time-base stability. The position of a bit in the sampled-data stream unambiguously labels the time at which it was taken, completely eliminating problems of tape stretch, wow, flutter, or alignment in assigning a time-tag to any bit of data. A further major advantage of digital recording is the extreme flexibility of data processing with general-purpose digital computers. The disadvantages are the complexity and high cost of special recording hardware and the high costs of data processing.

Another consideration in the choice of data recording systems is the duration of a recording on a single reel of tape. This duration varies widely from system to system and is not generally dependent on whether the system is analog or digital. The NRAO Mark I digital system (720 kbit/sec sample rate) fills a standard 2400-ft reel of 1/2-inch-wide computer tape in 3 minutes. The Canadian analog recording system (4 MHz bandwidth) fills a 1000-ft reel

of 1-inch wide video tape in 15 minutes. And the new NRAO Mark II system (4 Mbit/sec sample rate) fills a 1000-ft reel of 2-inch wide video tape in one hour. Obviously, greater convenience is afforded by longer recordings. The relative shortness of the 3-minute Mark I recordings is somewhat compensated for, however, by its direct compatibility with standard computer tape-reading facilities.

6.7-2 The NRAO Mark I Recording System

The NRAO Mark I recording system was chosen for use in the October 1969 experiment. At the time, there were no other viable alternatives and, furthermore, the Mark I system was a proven recording system with which we had had some previous experience. The main drawback was the very large number of magnetic tapes needed to perform a large experiment.

In principle, the Mark I recording system is quite simple. The bandlimited video signal (0-360 kHz) from the output of the video converter is first infinitely clipped (i.e. the output of the clipper takes on one of two values depending on whether the video output voltage is above or below zero), then is sampled at a rate of 720,000 samples per second. The sampling rate is controlled by internal digital logic, which is in turn driven by a 5 MHz signal from the hydrogen maser station reference. Uniformity of sampling rate with respect to the station reference clock is essential in obtaining an accurate switched-frequency group-delay estimate. Any internal

drifts or changes in the digital logic controlling the sampling times will directly affect the group-delay estimate by the amount of the drift.

After sampling, the data are formatted for output onto a standard 7-track computer tape drive operating at a tape speed of 150 inches per second and writing at a density of 800 characters per inch. The data stream is interrupted for ~5 milliseconds out of every 0.2 sec in order to generate a "record gap" which separates physical records on the tape. This data-blanking is precisely timed so that the first data-bit of each record is separated by exactly 144,000 samples (corresponding to 0.2 sec of real time) from the first data-bit of the adjacent records. In addition, the sample-time of the first data-bit of the first record is precisely controlled. By simply correcting records and bits within a record, then, the sample-time of any given bit is easily determined. Occasionally, due to errors occurring during playback, some "bit slippage" may occur so that calculation of the sample-time within a single record will be in error. This error is usually not serious because the bit count is reinitialized at the beginning of each record. Generally, when "bit-slippage" occurs, it is in multiples of 6 bits (the length of a tape "character") so that any correlation "signal" is almost entirely lost (a shift of 6 bits corresponds to more than 8 μ sec, several times the width of the main lobe of the cross-correlation function). The only effect is then to add a small amount of

noise to the observation. Another error that can occur during playback is a loss of record count. This loss can occur if a very bad section of tape is encountered such that several records are read as one or vice versa. A complete loss of signal, of course, results. And since no identifying data are written on a record by the Mark I system, this type of error is generally irrecoverable, although it happens infrequently. Newer VLBI recording terminals, including the terminal used at Haystack for the October 1969 experiment, includes time identification as an integral part of each record, thus eliminating possible problems of this sort.

Time synchronization of the Mark I recording system is basically a simple procedure. An external time reference from Loran C or a "traveling" rubidium clock, for example, is compared to a periodic pulse generated within the recording system. This pulse is adjusted forward or backward to match as closely as possible the epoch of the external reference source. The precision of this adjustment is typically about one microsecond. The rate at which site time progresses is then determined by the site frequency standard, which is not necessarily the same device from which the site epoch is derived. Typically, for example, the site epoch is set by comparison to a "traveling" cesium standard, but once set, site time is maintained by a hydrogen maser.

6.8 FREQUENCY STANDARDS

VLBI has become synonymous with "independent-standards" interferometry; that is, the participating sites take data completely independently, with no communication of any sort passing between them (except, of course, for the usual frantic phone calls). Of all the technical developments that have made VLBI possible, modern frequency standards are among the most important. Here we shall first briefly examine the characteristics of currently available devices. Then we shall discuss the requirements of various types of VLBI experiments and the corresponding suitability of particular types of frequency standards.

6.8-1 Characteristics of Existing Standards

Figure 6-3 shows a plot of the RMS fractional frequency stability vs. averaging time for a variety of commonly used standards. (Reference 41). It must be noted that there exists considerable controversy over the estimates of some of the characteristics indicated in this plot, particularly in regard to the hydrogen maser. Data published as early as 1968 (Reference 42) indicate measured stabilities of 7×10^{-15} over a 1 minute averaging period. A study made by NRAO in 1970 (Reference 43) measured the fractional stability of 1×10^{-12} over the same averaging period between a specific pair of masers. On the other hand, data taken between Haystack and Goldstone in 1972 is not inconsistent with fractional stabilities of several parts in 10^{14} over one day. Emerging from this controversy

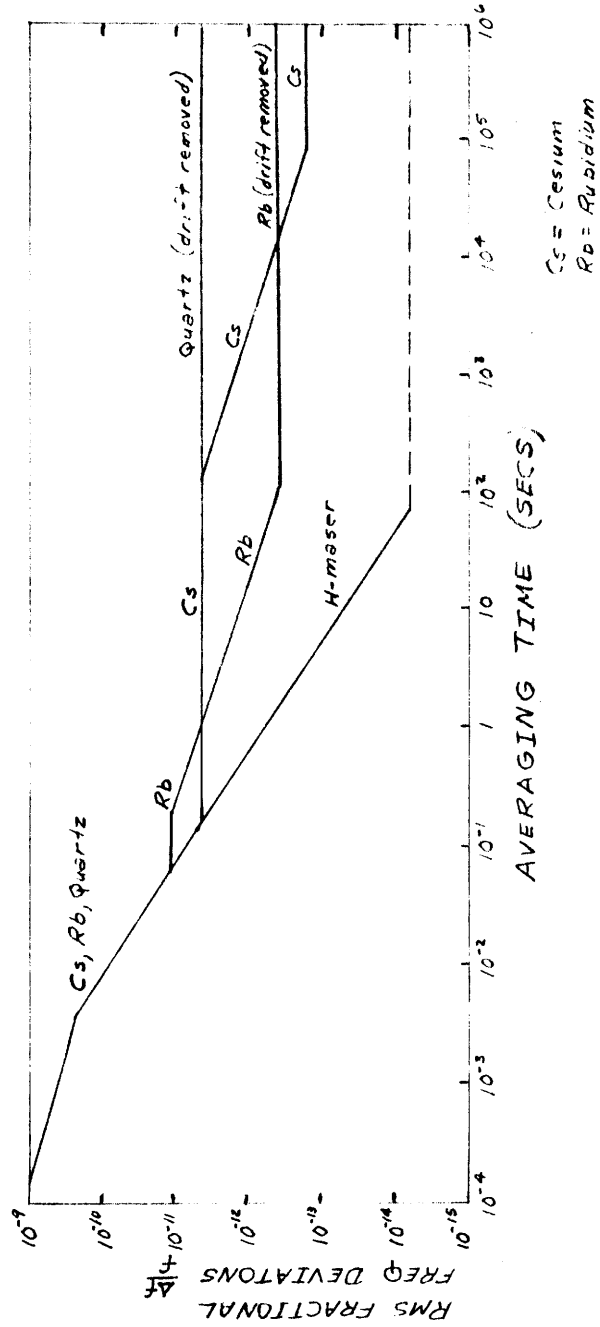


Figure 6-3
RMS Fractional Frequency Stability for a
Variety of Commonly-Used Frequency Standards

seems to be the simple facts that 1) estimates of hydrogen maser stability vary widely, and 2) hydrogen masers themselves seem to vary widely in their stability. Based on our own experience, however, the stability of a well-engineered and properly-running hydrogen maser is at least 1×10^{-13} over a day.

In contrast to the maser-stability uncertainties, the stability performance of other types of standards is quite well known. Here again, however, the plots of Figure 6-3 represent only a general indication of the average performance to be expected. The author's experience and that of others, however, warns that some of these devices, as with hydrogen masers, show significant variation between different models of the same type of device and even from serial number to serial number within a given model.

For averaging periods of up to ~ 1 second, Sulzer crystal oscillators (not shown in Figure 6-3) appear to be the best available, and it is logical, therefore, that for best stability, the Sulzer oscillator should be phase-locked to the hydrogen maser with a phase-locked loop time-constant of a few seconds. In fact, the short-term stability of all the atomic-frequency standards is essentially limited by the short-term stability of their associated phase-locked crystal oscillators and all currently available atomic standards could be improved by employing state-of-the-art Sulzer oscillators. Such a feature is not yet generally commercially available, but probably

will be if demand increases.

6.8-2 Frequency-Stability Requirements

Frequency-stability requirements vary considerably from one VLBI experiment to another. The stability requirements for simply obtaining fringes are, in general, much less stringent than requirements for doing good astrometric or geodetic work. Almost always a fairly large constant frequency offset can be tolerated, as long as it is not so large that observed fringe rates fall outside the limits of a practical search, usually about ± 2.5 Hz for the Mark I recordings. In order to simply obtain good fringe-amplitude results, random fluctuations in the phase of primarily the first local oscillator must be less than about one radian over the integration period. Assuming that the fractional stability of the first local oscillator closely reflects the fractional stability of the frequency standard, this requirement can be approximated by

$$\left(\frac{\Delta f}{f}\right) \omega T < 1 \quad (6-1)$$

where $\frac{\Delta f}{f}$ is the fractional stability of the frequency standard, ω is the radio frequency of the observations, and T is the integration time. In cases where the frequency standard is the major limitation in fringe-phase stability, expression (6-1) can be used to estimate the maximum

usable observing frequency for a given integration time, namely

$$\omega_{\max} \approx \left(\frac{\Delta f}{f}\right)^{-1} \frac{1}{T} \quad (6-2)$$

From Figure 6-3 we see that even a rubidium standard is sufficiently stable to be used at X-band (~4cm) if $T \gtrsim 100$ seconds. Rubidium standards have in fact been used for a number of VLBI experiments operating at X-band. They have also been used at even much higher frequencies, namely near the water vapor line at 22 GHz, where some sources are extremely strong, allowing the integration time to be cut significantly. The measurements made using rubidium as a standard usually cannot be used directly to obtain high precision astrometric or geodetic results. However, it is sometimes possible to circumnavigate the long-term stability problems of the frequency standard by frequently observing a "reference" source. If such a reference source can be periodically observed at intervals short enough such that the random drift of the standard is small (with respect to some measurement goal) between observations, then the reference source itself may be considered a high-precision clock against which all measurements of time are referenced.

Often this procedure is difficult to implement over a large area of the sky because of the practical limitations of rapid antenna movement. Recently, however, VLBI experiments have been conducted by the author and others in which four antennas, forming two interferometers with nearly co-located baselines, have been used (References 44, 45). The pair of antennas forming each end of the baseline have a common frequency standard, so that both long-baseline interferometers are identically affected by frequency-standard instability. If the two interferometers are pointed toward two different sources in the sky, one interferometer may be considered the reference for the other. The observable of interest effectively becomes the difference between the single-interferometer observables, which has all effects of clock instabilities removed.

The hydrogen-maser standard is the only frequency standard which has long-term instabilities which are not large compared to the uncertainties in some of the other unknown parameters. Long-term stability approaching a part in 10^{14} implies stability of nearly 1 nanosecond over the course of 1 day, a typical duration for a VLBI experiment. The uncertainties in present estimates of the electrical path length of the atmosphere and ionosphere (at X-band) is a large fraction of a nanosecond. Thus, the use of a hydrogen-maser standard must always be considered desirable for

any high-accuracy long-term experiments since it reduces to a minimum the number of clock parameters which must be estimated. Experience has shown that the difference between two well-adjusted and properly-functioning hydrogen masers can usually be well-characterized over a one-day period by a constant offset and linear rate.

Certain types of experiments require the stability of the hydrogen maser. For example, experiments at 22 GHz (water vapor) which require integration times $\gtrsim 100$ secs to obtain sufficient SNR would not be successful without the use of a hydrogen maser (unless another simultaneous interferometer on a stronger source can provide a "clock" reference). Also several recent experiments by this author and others have been successful in "connecting" fringe phase from one observation to another taken some minutes later. In other words, successful attempts were made to extrapolate the phase delay from one observation of a source to the next without introducing any 2π ambiguities into the time development of the fringe phase. If, say, two 3-minute X-band observations of the same source are made 10 minutes apart, a minimum necessary condition to successfully "connect" fringe phases is that the phase uncertainties at X-band be no more than about one radian over ten minutes, corresponding to a stability of about 3×10^{-14} . Clearly, no frequency standard other than the hydrogen maser could be

used to attempt such an experiment. For the immediate future at least, hydrogen-maser frequency standards should be adequate for practically all VLBI experiments which require stability of a few nanoseconds over periods up to several days.

6.8-3 Frequency-Standards Used for October 1969 Experiment

The frequency-stability requirements for the October 1969 experiment were set primarily by the requirements of the astrometric and geodetic portions of the experiment. During the planning of the experiment it was believed that all systematic effects could be modelled to the level of a few nanoseconds over a day. This consideration included the assumption that the performance of the reference standards could be modelled adequately to the level of a few tenths of a nanosecond per day. Only the hydrogen maser can begin to meet this requirement. Masers were therefore used at all three sites during the experiment. The maser used at NRAO was a Varian Model H-10, one of the first such units ever to be produced commercially. Haystack and OVRO were supplied with newly designed and constructed units from the NASA Goddard Space Flight Center. The primary output of all units is a 5 MHz signal. In addition, the units at Haystack and OVRO each had provisions for generating two 1 pps signals, each of which could be arbitrarily offset

with calibrated thumbwheel switches, a feature particularly useful for monitoring critical timing points within the VLBI system. Each maser was also equipped with a set of emergency batteries which would, in the case of a power failure, keep the maser operating for a period of up to several hours.

6.9 THE OVRO ANTENNA

The antenna used at the OVRO site was one of a pair of equatorially-mounted primary-focus 90-ft-diameter dishes normally used as a short-baseline interferometer. The antennas are mounted on tracks and can be moved to a number of locations. The location of the antenna for the October 1969 experiment was chosen to be as close as possible to the antenna control building where the hydrogen-maser frequency standard was housed. The surface of the dish is divided into "inner" and "outer" areas. The inner 45-ft diameter is composed of a solid metal reflecting surface. The remaining outer area is of punched aluminum, with ~ 1/8" punched holes with a comparable separation. The efficiency of the dish is estimated to be ~40% at L-band (11 cm) and ~20% at X-band (3.8 cm). Source tracking is by means of a simple sidereal drive. Initial source acquisition is essentially manual. Pointing accuracy at L-band is sufficient to point "blind", but blind-pointing at X-band can not be considered reliable. Although pointing correction curves were generated to aid in pointing the antenna, they are often useless during periods of high winds at the site, of which there were several during the October

1969 experiment. Of the problems encountered during actual data-taking, antenna pointing at X-band was one of the most severe.

All front-end receiver equipment was housed in an aluminum box with ~3 cubic feet of interior volume. The receiver box was mounted on a ring which could be remotely rotated through 360° to adjust for proper linear polarization alignment with the other participating antennas. In addition, the receiver box could be extended or retracted several inches toward the center of the dish to adjust focussing of the feed horn.

Signals to and from the antenna passed through various long cables. The 5-MHz maser reference signal was sent through ~300' of buried Spiruline coaxial cable to the base of the antenna, and then through ~150' more of Spiruline to the antenna focus. The IF signal from the receiver to the control room followed a similar path. All control and monitor signals were sent through ~250' of heavy multiconductor cable strung over the ground from the control room to the antenna base, and then through various available twisted-pair and coaxial cables to the focus. Most heavy power supplies, including two high-voltage power supplies for the klystrons and a high-current thermoelectric-cooling-system controller were housed in the base of the antenna.

6.10 CONSTRUCTION, TESTING, AND INSTALLATION OF THE RECEIVER AT OVRO

6.10-1 Receiver Construction

All of the front-end receiver equipment was placed

within a specially constructed aluminum housing designed to be compatible with the size, shape and mounting restraints imposed by the antenna structure. A basic aluminum housing provided by OVRO was extensively modified for application to the October 1969 experiment. Large access panels were added for operational convenience. The entire structure was strengthened by the addition of welded reinforcing bars. A heavy-duty feed-mount support frame was constructed to hold the feed-horn assembly in the proper position.

A temperature-control system was installed to maintain a constant internal temperature. This was necessitated because of the large day-to-night temperature variations (~40°F) in the desert-like climate of the Owens Valley area. Four thermoelectric heat pumps were used for this application, each with a maximum capacity of 100 watts (heating or cooling). Internal dissipation within the receiver equipment was estimated to be ~300 watts, mostly due to the par-amp pump klystrons. Four fans were used to keep a constant flow of air circulating through the receiver equipment and the temperature-control air ducts along the side of the equipment box. Large finned radiators, in direct thermal contact with the thermoelectric heat pumps within the air ducts, served to control the temperature of the circulating air. Internal air temperature was sensed by a thermistor, whose "output" was used to control the magnitude and direction of current through the thermoelectric heat pumps. A special controller built by NRAO was used to drive the thermo-

electric units, with currents of up to 30 amperes (Reference 46). This controller uses silicon-controlled rectifiers to control the current to the heat pumps; careful filtering and screening are used to eliminate radio interference as much as possible.

A major problem in the construction of the receiver was the very limited space into which all of the receiver equipment had to be fitted. As mentioned earlier, only 3 ft³ of space was available for the installation of the complete X-band and L-band receivers and local-oscillator systems. Figure 6-4 shows a view of the completed receiver box as it would look installed on the antenna. In Figure 6-5 the outside shell and temperature-control system have been removed to show the construction of the interior equipment rack. Here the L-band feed horn is clearly visible. The concentric X-band feed is partially visible in Figure 6-4 behind a weatherproof fiberglass radome. Finally, Figure 6-6 shows a view of the completed receiver box with only the top cover removed. The four thermoelectric coolers, two on each side, are mounted on ducts into which radiating fins extend. Total weight of the finished box was nearly 300 lbs. Ground support equipment for the receiver box, excluding the hydrogen maser, was housed in standard equipment racks in both the antenna base and antenna control building.

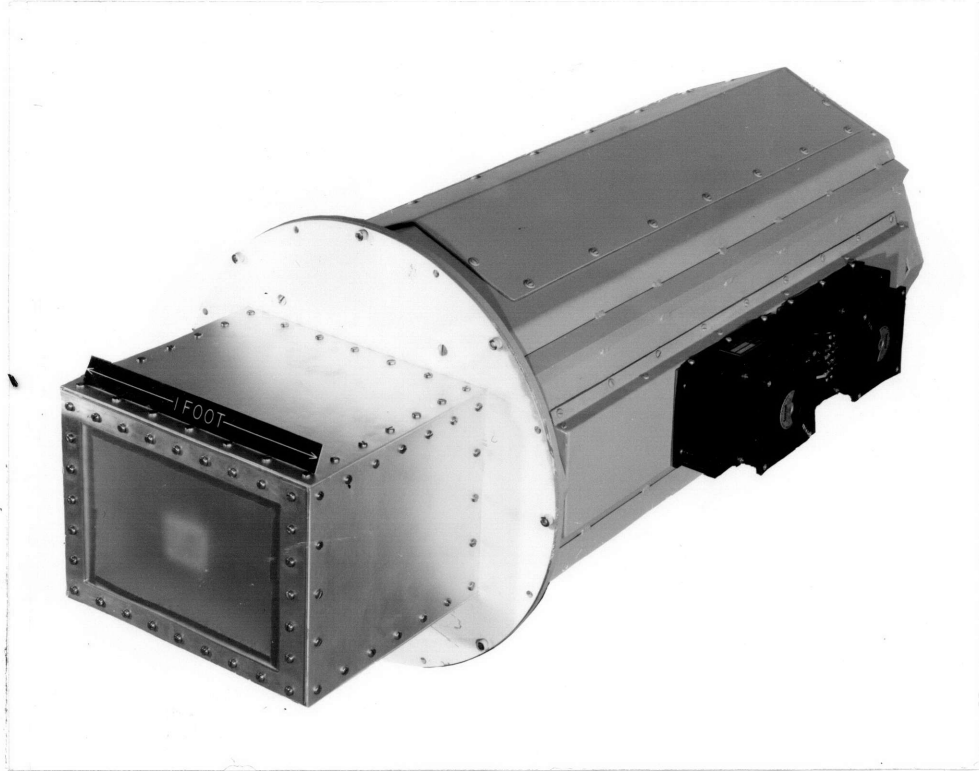


Figure 6-4: Completed Receiver Box

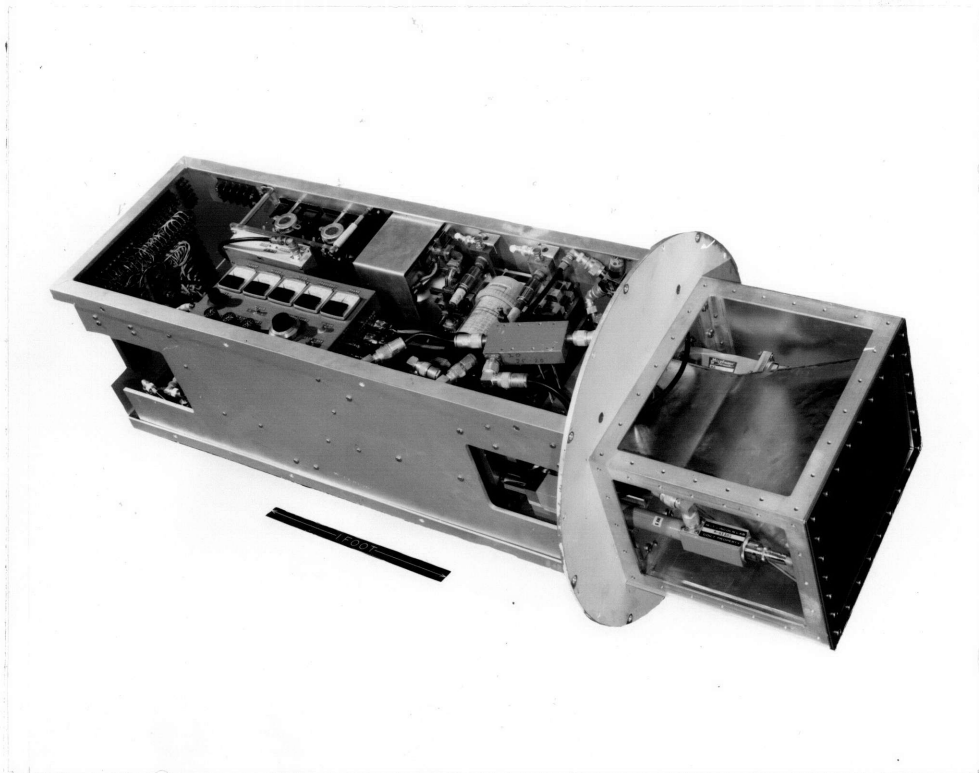


Figure 6-5: Interior Equipment Rack

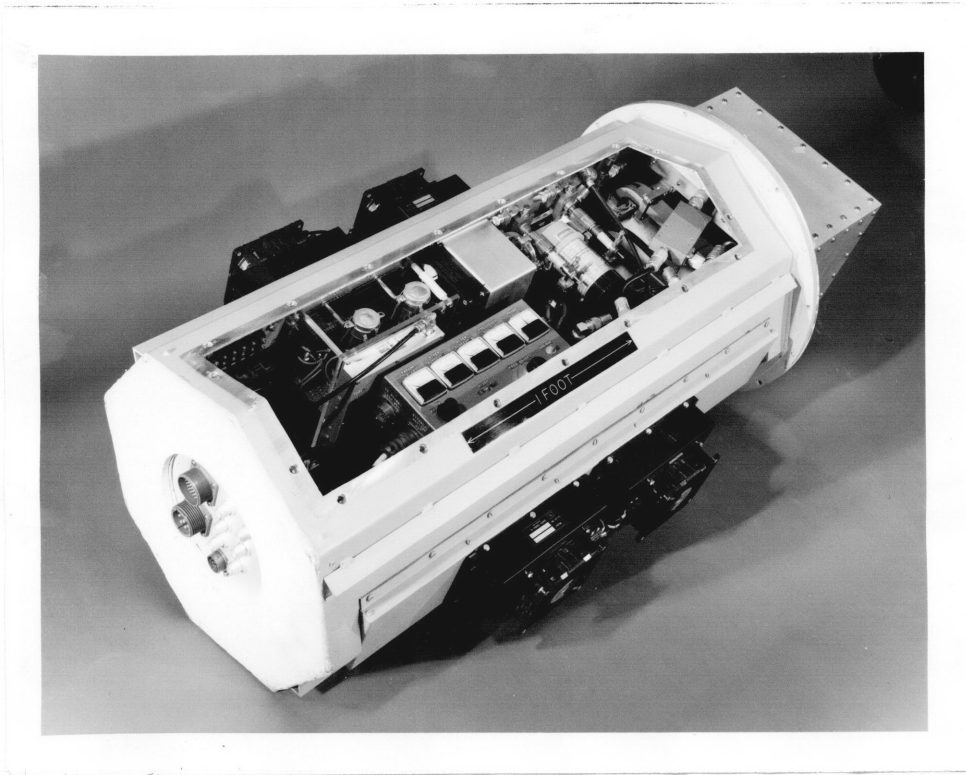


Figure 6-6

Receiver box with top cover removed

6.10-2 Receiver Testing and Installation

Unfortunately, the stringent demands of the experiment schedule did not allow nearly as complete a testing and checkout of the receiver system as would have been desirable. Only two days of time were available for complete system checkout before shipment to California. All systems seemed to be working well except for some difficulty in tuning both the X-band and L-band parametric amplifiers. Estimates of the total system temperatures were made and proved to be a good deal higher than expected. At that time there was little that could be done to fully investigate the problems. All further testing had to be done in California. This normally would not constitute a severe problem, except that much of the sophisticated test equipment which was available at Haystack was not available at OVRO. After arrival in California (a few days late as usual, following normal air freight procedures) the whole system was again tested "on the bench" as carefully as possible before the receiver box was actually mounted on the antenna. Again paramp problems were encountered. After a few hectic days of work the whole system was installed on the antenna and prodded into life. All systems appeared to be working properly except that system temperatures were much higher than desired. X-band system temperature was estimated to be $\sim 300^{\circ}\text{K}$ and L-band

temperature was nearer $\sim 800^{\circ}\text{K}$. Careful calibrations of system temperatures were difficult because of a fairly large uncertainty in the dish and feed system efficiency and a fairly large uncertainty in noise-tube temperatures. The reasons for the high system temperatures were never entirely explained and extensive post mortems were not possible because all equipment had to be removed from the antenna and effectively dismantled immediately upon completion of the experiment.

CHAPTER 7

CONDUCT, ANALYSIS AND RESULTS OF THE EXPERIMENT

7.1 PRELIMINARY SETUP AND TESTING

Toward the middle of September 1969 the necessary equipment was being gathered at the three sites for the upcoming experiment. The receiving equipment, which had all been constructed at Haystack, was shipped to NRAO and OVRO. Approximately 3000 magnetic tapes were obtained through the tape-reconditioning facility of the NASA Goddard Space Flight Center and were distributed to the three sites. A hydrogen maser, also from NASA/GSFC, was also shipped to each of the participating stations.

Time synchronization at Haystack and NRAO was accomplished by comparison of the VLBI clocks with WWV and Loran-C signals. WWV is used for a coarse synchronization to about the millisecond level and Loran-C for the fine synchronization to a few microseconds. At OVRO no Loran-C or equivalent signal was available, so that synchronization had to be done by other means. A rubidium clock, obtained for the duration of the experiment from the National Bureau of Standards at Boulder, Colorado, was synchronized to a primary cesium standard at NBS. The operating clock was then flown to the OVRO station and the hydrogen maser was synchronized to it.

At OVRO, where the author was located, several problems were encountered during the setup period. Particular problems

were encountered in tuning both the X-band and L-band parametric amplifiers and adjusting their gains and bandwidths to the desired specifications. System temperatures at both X-band and L-band were significantly higher than had been expected. When mounted on the antenna, the X-band system temperature was $\sim 300^{\circ}\text{K}$, but the tuning of the parametric amplifier was quite sensitive to mechanical vibrations and stresses. When tuned as best as possible the gain at the center of the band was approximately 20db with a 3db bandwidth of $\sim 30\text{MHz}$. Since the parametric amplifier was relied upon for sideband rejection as well, the actual X-band sideband rejection was $\sim 20\text{db}$ over most of the observed frequency band. At L-band the parametric amplifier performance was even poorer and the system temperature, mounted on the antenna, was $\sim 800^{\circ}\text{K}$. Gain was approximately 15db over a bandwidth of approximately 50 MHz. As with the X-band system, the L-band parametric amplifier was relied upon for sideband rejection, so that actual sideband rejection was about 15db.

Several other problems were encountered during the setup period. One problem was due to the very large day-to-night temperature changes, often $40\text{-}50^{\circ}\text{F}$, of the desert atmosphere at the OVRO site, taxing the front-end-box temperature control system to its limits. After several days of experimentation it was found that an internal temperature setting of about 80°F was about optimum for lessening, as much as possible, the load placed upon the temperature-control system.

Another problem, potentially more serious, was the ~300 lb weight of the box itself. This weight was several times more than had previously ever been mounted on the antenna and there was some concern about the mechanical strength of the systems which are used to focus (i.e., to move the feed toward or away from the surface of the dish) and to rotate the box. A decision was made to go ahead and try. Fortunately, nothing broke.

After the entire receiving and recording system was set up, test tape recordings of both the phase calibrator signal and broadband noise were made and sent back to Haystack for spectrum analysis. The results of these tests were satisfactory and the system was judged operationally ready to proceed even though the high system temperatures at both X-band and L-band were still of concern.

Similar setup and test operations took place at Haystack and NRAO under the direction of Hans Hinteregger at Haystack and Tom Clark and Curt Knight at NRAO. Numerous people were stationed at each site to help with the observations to be made during the coming days.

7.2 CONDUCT OF THE EXPERIMENT

7.2-1 Scheduling of Observations and Sources Observed

An attempt was made to schedule observations so that each primary source was sampled as often as possible each day. Of course, considerable attention was paid to 3C273B and 3C279 during the daytime, pursuant to the gravitational-bending measurement. Typically, a 3-minute observation was made every 10 to 15 minutes, according to a master schedule prepared in advance. Since simultaneous X-band and L-band observations were not possible, some periods of time were devoted to all X-band others to all L-band, and some periods to mixed X-band and L-band observations. Ideally, mixed X-band and L-band observations would have been scheduled for most of the experiment, but the 15-20 minute changeover time from X-band to L-band at Haystack would have made such a schedule inefficient.

Except for some brief periods of a few hours for maintenance and repair, observations were conducted on a 24-hour-a-day around-the-clock basis from 1 October to 15 October 1969. A basic set of eight sources, including 3C273B and 3C279, were chosen for concentrated observations. These sources are listed in Table 7-1 along with their a priori coordinates.

In addition some observations were also made on a number of weaker or more speculative sources from which, unfortunately, very few useful results were obtained, due partly at least to

Table 7-1 Primary Sources Observed and Their
A Priori Coordinates

Source	References for Position	α (HMS)	Position (1950.0)	δ (DMS)
3C84	8	3 16 29.57	41 19 52.1	
3C120	7	4 30 31.60	5 14 59.5	
4C39.25	47	9 23 55.34	39 15 24.4	
3C273B	48	12 26 33.246	2 19 43.38	
3C279	49	12 53 35.82	-5 31 07.6	
3C345	7	16 41 17.57	39 54 10.6	
VR042.22.01	47	22 00 39.30	42 02 08.5	
3C454.3	7	22 51 29.50	15 52 53.7	

receiver problems. Many of these sources simply could not be seen well enough (if at all) in total power to point the antennas properly, particularly at OVRO. With a more sensitive system many of these sources might have shown fringes or more consistent fringes. Sources such as CTAl02, 1127-14, 2134+00, VRO42.22.01, 3C371, OK290 and OJ287 have in fact shown good fringes in later experiments using more sensitive equipment, notably the extremely sensitive Haystack - Goldstone interferometer. The sources included in this latter category include 0735+17, 0736+01, 2134+00, 0438-43, 3C371, 3C380, CTAl02, 0742+10, 1127-14, NRA0530, OV080, 3C405, 3C418, OX161, OX074, 2145+06, 2127+04, 2203-18, 2345-16, 3C446, 2328+10, 0048-09, 0106+01, OC328, 0202+14, 0605-08, 0607-15, 1055+01, 3C309.1, NRA0140, 0420-01, 0834-20, OI363, OJ287, OK290, and 1148-00.

7.2-2 Observing Frequencies

The two basic observing frequencies were ~7840 MHz (X-band) and ~1660 MHz (L-band) (also see Section 5.2). Several different sets of switched-frequencies were used during various parts of the experiment. The exact sequences used, including single-band frequencies, are listed in Table 7-2; also indicated is whether the observations associated with a particular sequence are upper or lower sideband. Generally, all X-band observations were lower sideband and L-band observations were upper sideband.

Table 7-2 Frequency Sequences Used
During the Experiment*

	Channel Number						Sideband
	1	2	3	4	5	6	
1.	7833.1,	7832.1,	7829.1,	7827.1,	7809.1,	7797.1	L
2.	7833.1						L
3.	1616.9,	1617.9,	1620.9,	1622.9,	1640.9,	1652.9	U
4.	1616.9						U
5.	7809.1,	7832.1,	7829.1,	7827.1,	7833.1,	7797.1	L
6.	7809.1						L
7.	1640.9,	1617.9,	1620.9,	1622.9,	1616.9,	1652.9	U
8.	1640.9						U

*The frequencies listed here are the "total local-oscillator frequencies," that is, the RF frequency that is translated to 0 Hz at the video output.

Sequences 1-4 were used before 1300 UT on 9 Oct. Sequences 5-8 were used after 1300 UT on 9 Oct. Note that sequences 1 and 5 and sequences 3 and 7 are the same except for the order of the channels.

Although most of the observations were made using the switched-frequency technique to determine group delay, occasional single-band observations were made in order to aid in the determination of the a priori clock offsets in later data processing. Although in principle it is possible to determine the same information from frequency-switched observations, in practice it is very time consuming because of the significantly more complicated nature of the frequency-switched data processing.

7.2-3 Details of a Single Observation

Each single observation consisted of recording approximately 3 minutes of data on magnetic tape. The recording system, once armed, started the tape automatically at a precisely controlled instant relative to the master station clock. Also at this instant, the frequency-switching controller took over control of the synthesizer controlling the frequency-switched second local oscillator, switching at a rate of five times per second. During the first ~5 seconds of the observation, a high-level phase-calibration signal, as described in Section 6.2, was added into the first IF signal, appearing in the video at a frequency of 100 kHz in each frequency window. At the end of this 5 second period, the phase-calibration system was shut off and normal VLBI data were recorded. Total power information was recorded separately on a continuously-running paper-chart recorder.

An important detail that had to be adjusted for each observation was the direction of the polarization of the linearly-polarized feeds. The direction of polarization of the signal received at the Haystack antenna, which has an azimuth-elevation type of mount, changes as a function of azimuth and elevation. The NRAO and OVRO antennas, on the other hand, are of the polar-mount type and maintain a constant polarization angle on the sky. In order to keep the polarization angle of all antennas in the same direction (on the sky), the feeds at NRAO and OVRO were rotated, for each observation, to match the Haystack polarization angle. The feed at Haystack was of a non-rotatable type.

7.2-4 Operational Difficulties Encountered

Some operational difficulties were encountered at each of the three participating stations. Both Haystack and NRAO had been used previously for VLBI experiments and, as a result, there tended to be fewer problems there than at OVRO, which was being used for the first time.

At OVRO two primary and somewhat related problems were encountered. The higher than anticipated system temperatures at both X-band and L-band made pointing at all but the strongest sources very difficult. Due to the characteristics of the antenna pointing system, it was sometimes possible to point "blind" at an L-band source and acquire it successfully; at X-band, however, with a beamwidth of a few minutes of arc, the only way that a source could be reliably acquired was

to carefully monitor total power. Any source with a total flow of less than about 10 F.U. at X-band, which corresponds to an increase of about 0.4°K in antenna temperature, was practically undetectable since the gain fluctuation in the parametric amplifiers were on the order of 0.5°K over a period of a few seconds. The pointing difficulty was exaggerated whenever winds of more than a few mph, which were not uncommon, were buffeting the antenna. High winds forced the total shutdown of operations on several occasions.

Also at OVRO, some difficulty was experienced in keeping the X-band parametric amplifier properly in tune from one day to the next, apparently due to mechanical instabilities. Retuning was required several times during the course of the experiment. The only other further problem of note at OVRO was the loss of synchronization to the hydrogen maser when, approximately halfway through the experiment, a switch was accidentally changed. The VLBI equipment was immediately re-synchronized and remained synchronized for the remainder of the experiment.

At NRAO the major problem turned out to be several massive power failures at the site, resulting in loss of clock synchronization on several occasions. Following restoration of power, in each instance, all equipment had to be brought back up to operating status, usually accompanied by warm-up drifts and instabilities. Post data-analysis indicates that the hydrogen maser at NRAO had several clock synchronization

"breaks" during the course of the two week experiment.

At Haystack, operations proceeded smoothly for the most part. However, several days of bad weather, including rain on some occasions, caused the radome to be wet and raised the system temperature by $\sim 70-100^{\circ}\text{K}$.

7.3 DATA PROCESSING THROUGH VLBI 1 AND VLBI 2

7.3-1 Determination of the Coarse Clock Synchronizations

At the completion of the experiment all of the data were shipped back to the NASA Goddard Space Flight Center where the nearly 3000 magnetic tapes were organized and catalogued. The first stage of the data-reduction procedure was to take samples of single-band data from each baseline individually and do a systematic search for fringes by offsetting the a priori clock synchronization error in steps. Each pass through VLBI 1 and VLBI 2 covers an effective search range of about 6-7 microseconds. Typically, synchronization errors were less than ~15 microseconds, requiring only three or four passes through VLBI 1 and VLBI 2 to locate the proper synchronization. In principle, of course, it is necessary to examine only two baselines to determine the clock synchronization errors for all three baselines. In practice, trials are usually made on all three baselines to insure that the clock-synchronization determinations are proper. Sample data selected from throughout the experiment were processed in this manner to determine if any anomalous clock synchronization "breaks" occurred during the experiment, and to determine the periodic adjustment of clock synchronizations, if necessary, to compensate for differences in clock rates. The clock synchronization parameters determined by this procedure are very coarse usually accurate only to the order of a few tenths of a microsecond, and are intended only to become part of the

a priori model for the actual processing of the data.

7.3-2 Determination of Phase-Calibration Correction Phases

The second stage of the data processing requires determining the so-called "phase-calibrator correction phases" (see Equation 4-13). These are a set of constant phases which are to be added to the measured phase-calibration phases to take into account the initial arbitrariness of the oscillator phases of each of the frequency-switched channels. Once determined, this set of correction phases should remain constant for the duration of the experiment as long as the synthesizers controlling the switched-frequency oscillators are not disrupted. The procedure for determining these phases is relatively straightforward. Several 3-baseline samples of frequency-switched data are selected from each period of time for which the phase coherence of the switched-frequency oscillators is believed to have been maintained. These data are first processed through VLBI 1 and VLBI 2 with no correction phases, using the a priori clock synchronization parameters determined in the first stage of the processing. Then, using the data from the two baselines with the highest correlation amplitudes, the residual fringe phase for each frequency channel at zero residual delay is noted and then is applied as a "correction phase" for a reprocessing through VLBI 2. After several iterations of all the sample observations through VLBI 1 and VLBI 2 a set of correction phases is found which minimizes the fluctuations, from observation to observation,

of the individual-channel residual fringe phases around the mean residual fringe phase and causes the empirically-determined delay resolution functions to be nearly symmetric.¹ The correction phases for the third baseline are then chosen such that the sum of the correction phase around the three baselines for each frequency channel is zero. This insures that the local oscillator phases are handled consistently so that the sum of the group delays around the three baselines should be nearly zero (within the limits of the expected errors).

1

This procedure may seem to destroy knowledge about the clock synchronizations. However the actual clock synchronizations cannot be determined from this experiment to any higher accuracy than a single-band observation will allow. The reason is that only the relative phases of the switched-frequency local-oscillator signals can be determined by the phase-calibration method employed. Their phases, relative to the phase of the hydrogen-maser clocks, are not known and were not measured, and thus the exact relation between the actual clock synchronization and the apparent clock synchronization is not known.

Put another way, the set of "phase-calibrator correction phases" simply defines the relation of the switched-frequency oscillator phases to the readings on the station clock. In so doing, a small arbitrary constant is, in effect, added to the true clock readings at each site. This has no effect on any of the geodetic or astrometric measurements. The only consequence is to introduce a small unknown constant bias in the final clock synchronization solutions. A more sophisticated switched-frequency oscillator system, in which the relation of the oscillator phases to the readings on the station clocks is precisely known, would eliminate this small synchronization uncertainty and allow the true clock synchronization to be determined to the full accuracy of the switched-frequency technique.

7.3-3 Mass Processing of Data Through VLBI 1 and VLBI 2

Once the a priori clock synchronization estimates and phase-calibrator correction phases had been determined from samples of data taken throughout the experiment, all parameters were in hand for the mass processing of the entire set of data through VLBI 1 and VLBI 2. Both VLBI 1 and VLBI 2 could simultaneously process all three baselines of a three-station observation, typically requiring about 8-9 minutes of time using an IBM 360 Model 91.

The output of VLBI 1 was collected on a set of archival tapes to maintain a permanent record of the record-by-record correlations of the original data. Processing through VLBI 2 could then take place immediately or at any later time. The phase-calibrator correction phases could be entered as an input to VLBI 2, so that their values could be modified if necessary, without reprocessing through VLBI 2. The output of VLBI 2 was collected on punched cards, approximately six cards per baseline-observation. The data on these cards contained all of the information needed for final processing through VLBI 3 as well as some additional data for post-mortem analysis of individual observations.

7.3-4 Examples of Results of Individual Observations

In this section we shall examine the results from several typical individual observations. In so doing we shall be able to point up some of the strong and weak points of the experiment and also, to some extent, to compare the

experimental results with predictions.

The first observation which we shall examine is a switched-frequency X-band 3-station observation on the source 3C273B taken at 1545 UT on 2 October 1969. Figures 7-1, 7-2 and 7-3 show the results from each of the three baselines, respectively. The results from each individual baseline are displayed in three different ways in parts (a), (b) and (c) of each figure:

(a) Correlation amplitude vs. residual fringe rate.

The coarse delay-rate search (see Section 4.5-2) is conducted by computing the correlation amplitude for trial residual fringe rates spaced at $\sim 4.89\text{mHz}$. The plot shows 256 trial values, spanning a range of $\pm 0.625\text{Hz}$. The width of the signal peak should be approximately $1/T$, where T is the length of the observation. Any significant broadening of the signal peak or high sidelobes near the main peak may indicate poor oscillator stability, mispointing of one or both of the antennas during part of the observation (so that the actual time spent pointing at the source is $<T$), or unknown factors such as the atmosphere, ionosphere and the solar corona.

(b) Experimental delay resolution function vs. residual delay

The phase-calibrator and correction phases are applied to the residual phases before constructing the delay resolution function. The function is computed for 256 discrete residual

delays spanning ± 500 ns (i.e., one full delay ambiguity) around the best estimate of the delay determined in the initial signal search. Any asymmetry around the peak of the function is due to either noise or systematic bias in either the data, phase-calibrator, or correction phases. The final estimate of group delay is determined by interpolating to the peak of the delay resolution function. It is interesting to compare the experimental results with the theoretical delay resolution function as shown in Figure 4-1c.

(c) Correlation amplitude and residual phase vs. time for each of the six frequency channels. For this plot the final estimates of residual delay and delay rate have been applied as corrections so that both phase vs. time and phase vs. frequency-channel should be nearly constant and equal. Each point represents 2 seconds of actual integration time on a particular frequency-channel. Any systematic trends in phase vs. time will also be indicated by a broadening of the fringe rate spectra in the first plot, although deviations of ≈ 1 radian are necessary in order for the broadening to be evident. Deviations of the phase from one frequency-channel to another indicates that either the phase-calibrator system is not perfect, or that there are unaccounted-for dispersions in the equipment; phase deviations from frequency-channel to frequency-channel will also be evidenced by asymmetries about the peak of the delay resolution function.

Examination of Figures 7-1, 7-2 and 7-3 indicates that the fringe rate spectra are not noticeably broadened for any of the three baselines, nor do any high sidelobes exist, suggesting, at least, that there are no severe fringe-phase instabilities over periods of time from about ~ 0.2 to ~ 100 seconds. Large phase deviations with a period of $\lesssim 0.2$ seconds would tend more to reduce the amplitude of the peak than to broaden it because the averaging period before the computation of the fringe rate spectrum is ~ 0.2 seconds (i.e., one record). Large phase deviations with a period $\gtrsim 100$ seconds would tend to shift the signal in fringe rate more than broaden it or reduce its amplitude. The delay resolution functions for the NRAO-Haystack and Haystack-OVRO baselines are quite symmetric, while the NRAO-OVRO delay resolution has a slight asymmetry.

The residual phase vs. time plots, particularly those of data from the NRAO-Haystack and NRAO-OVRO baselines (Figures 7.1c and 7.2c) show systematic trends on the order of a few radians. Particularly evident is an apparent phase transient at the beginning of the NRAO-Haystack observation. Each frequency channel shows a transient of similar character. It is now believed that these transients were primarily due to the initiation of the switching of the frequency synthesizers at the beginning of the observation. The systematic behavior of the phase vs. time of the data from the NRAO-OVRO baseline is somewhat less pronounced due to the

much lower signal-to-noise ratio, but there does seem to be a fairly large component (~ 2 radians) of a systematic nature common to all six frequency channels. Such systematic behavior of the fringe phase was observed on many switched-frequency runs throughout the experiment, at both X-band and L-band.

Table 7-3 details the numerical results of the observations presented in Figures 7-1, 7-2 and 7-3. For each baseline the results are given for the actual measurement as well as for so-called "a priori" and "a posteriori" models. The "a priori" model gives the pre-experiment estimates of the observation parameters, including the a priori delay and delay rate and the originally expected correlation amplitude, signal-to-noise ratio (SNR), rms phase noise, observation length and the expected rms errors in the measured delay and delay rate. The "measured" parameters include the actual measured delay, delay rate, correlation amplitude, rms phase noise, and observation length. Based on the observed correlation amplitude and actual observation length, the "a posteriori" model gives the theoretical SNR, theoretical rms errors in the measured delay and delay rate, and the theoretical rms phase noise. Clearly, the correlation amplitudes of the actual data are below the original expectations, particularly on the baselines involving OVRO. Also, the "measured" rms phase noise is significantly higher than the "a posteriori" model would imply. This discrepancy is pri-

marily due to the large systematic trends remaining in the residual fringe phase over the duration of a 3-minute observation. If these systematic trends are removed, the measured rms phase noise should correspond closely with the theoretical value.

Figures 7-4, 7-5, 7-6 and Table 7-4 show the results of a typical 3-station single-band observation. The particular observation chosen is an L-band observation on 3C273B. Again, we note from Table 7-4 that the measured correlation amplitudes are much smaller than was a priori expected. The measured phase noise and mean noise levels agree well with the "a posteriori" model, which is based on the measured correlation amplitude and actual observation length. The phase noise calculations of the "a posteriori" model have been made using the large signal-to-noise ratio assumption implied in Equation (3-58) which is probably only barely adequate for the extremely low correlation amplitudes measured on the NRAO-OVRO and Haystack-OVRO baselines. In fact, the measured correlation amplitude of 0.0009 on the Haystack-OVRO baseline is very near the limits of detection (see Section 3.7) as is clearly evidenced by the surrounding noise level in the fringe rate spectra plot of Figure 7.6a.

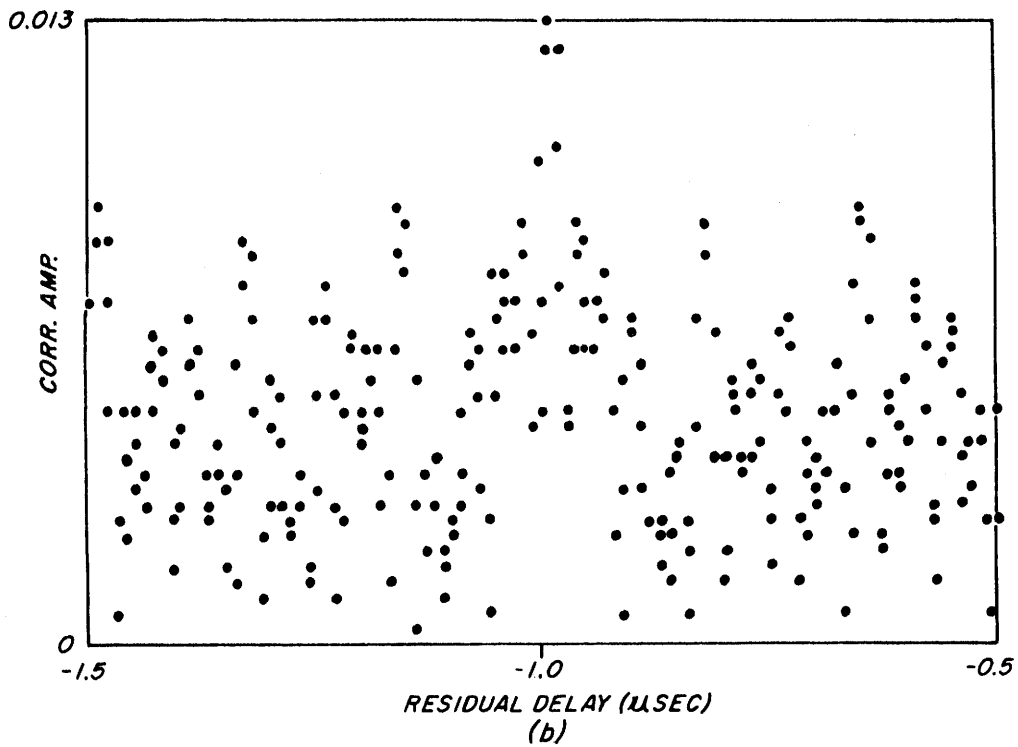
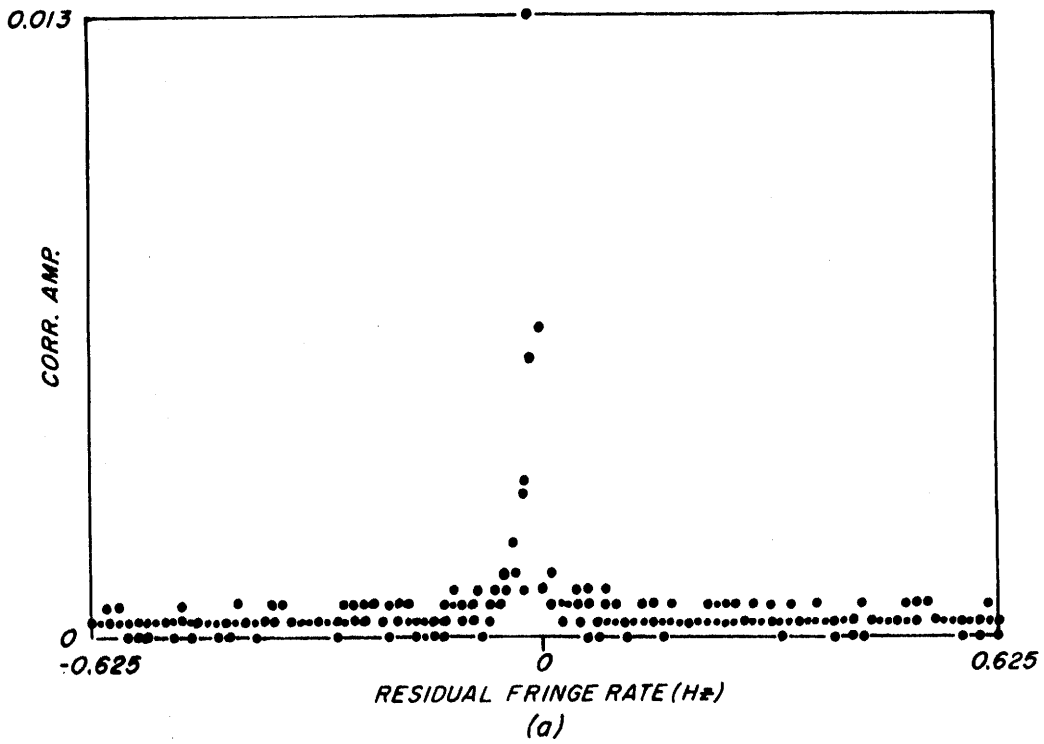


Figure 7-1

02 Oct 1969 1545UT 3C273B NRAO-Haystack X-band-switched
(a) Correlation amplitude vs. residual fringe rate.
(b) Delay resolution function amplitude vs. residual delay.

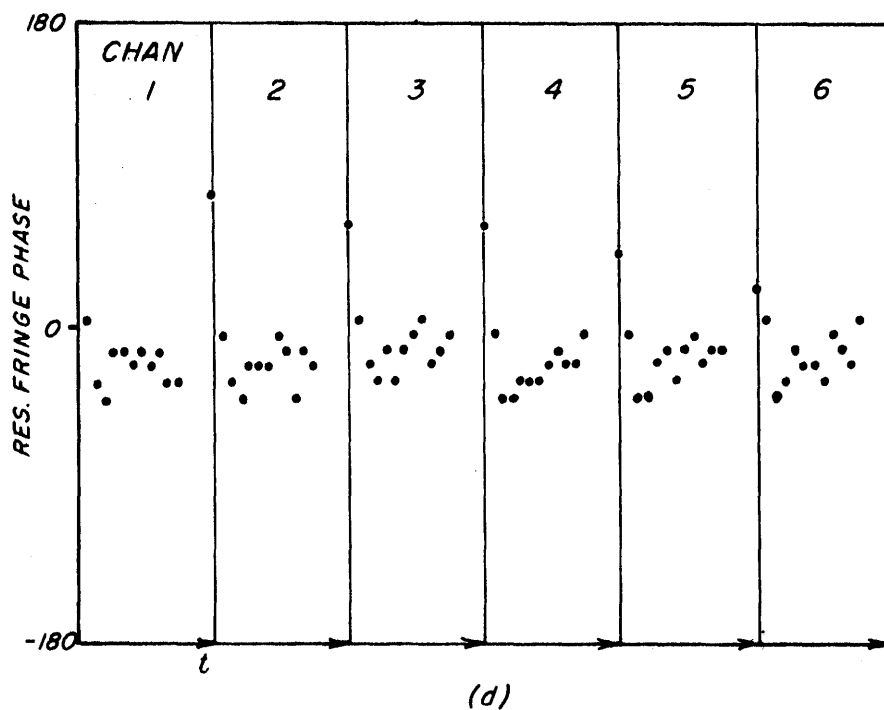
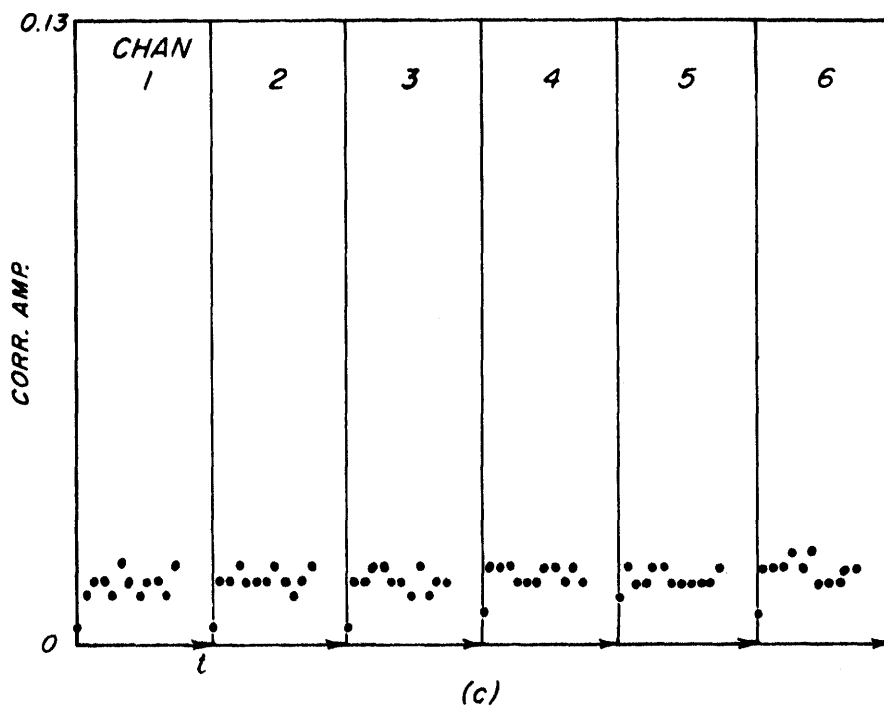


Figure 7-1 (contd)

(c) Individual-channel correlation amplitudes vs. time.

Each point represents 2 seconds of data on the corresponding channel.

(d) Individual-channel residual fringe phase vs. time.

Each point represents 2 seconds of real integration time.

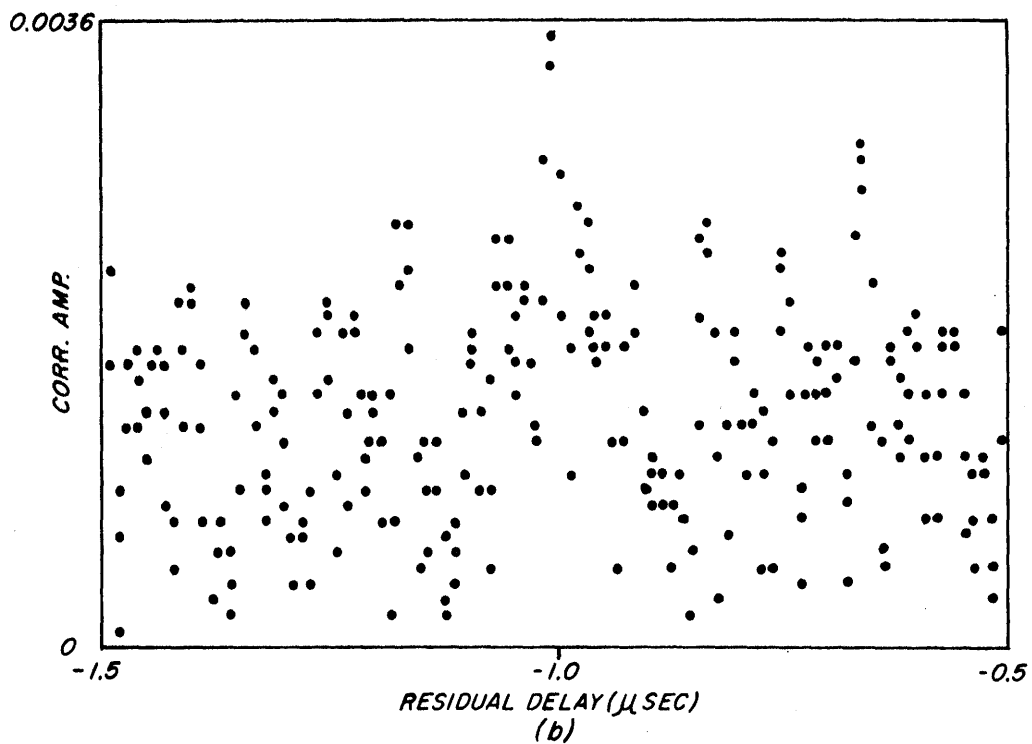
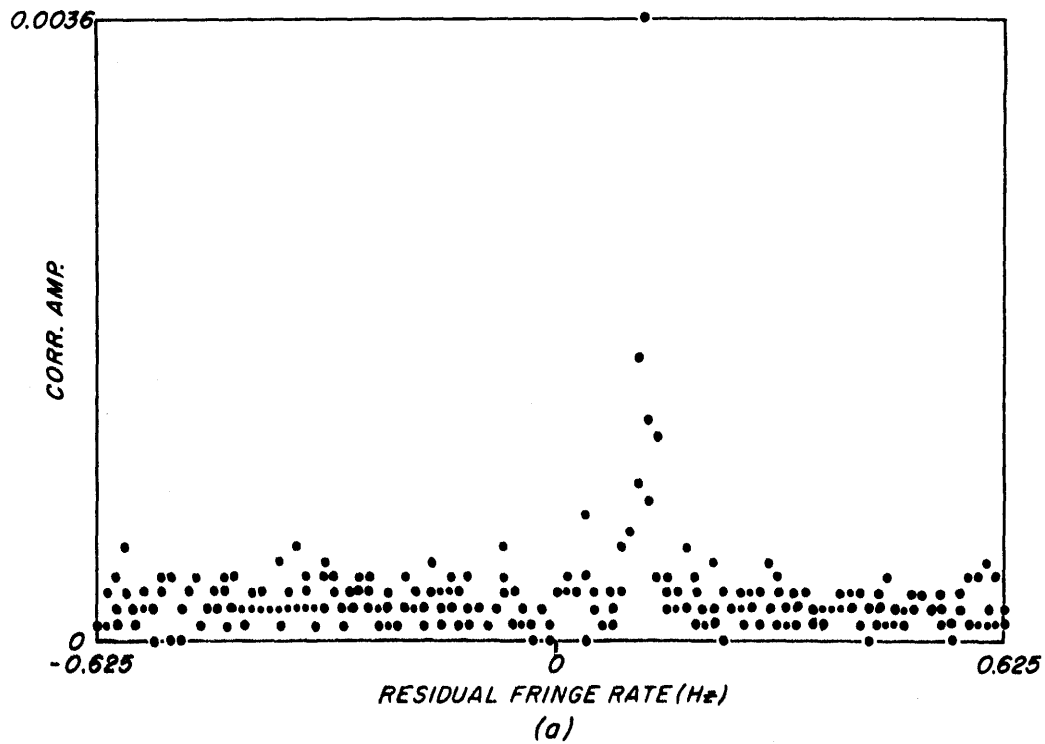


Figure 7-2

02Oct1969 1545UT 3C273B NRAO-OVRO X-band-switched
(a) Correlation amplitude vs. residual fringe rate.
(b) Delay resolution function amplitude vs. residual delay.

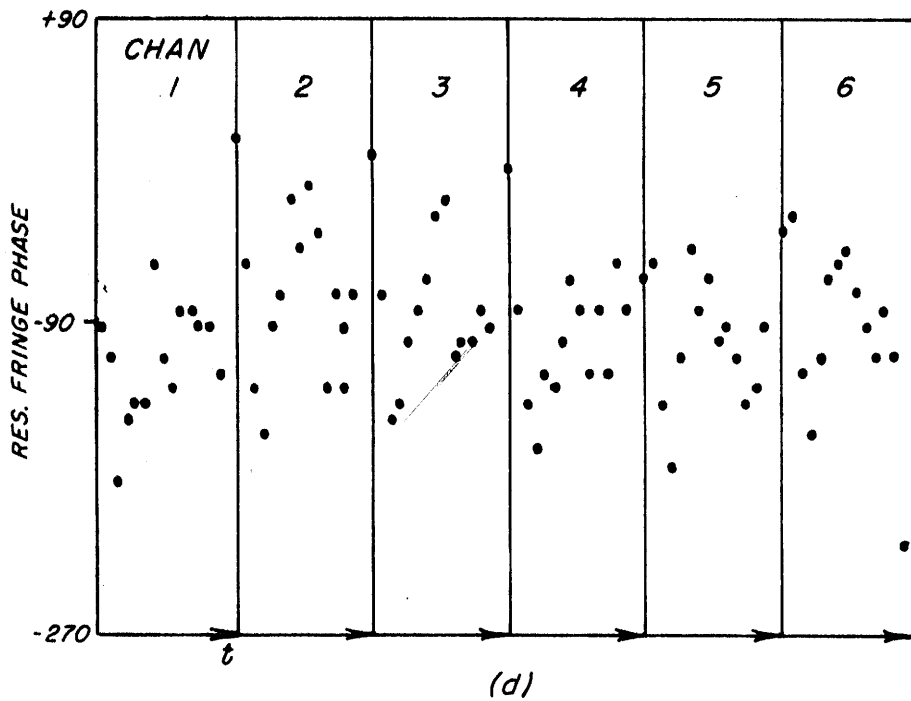
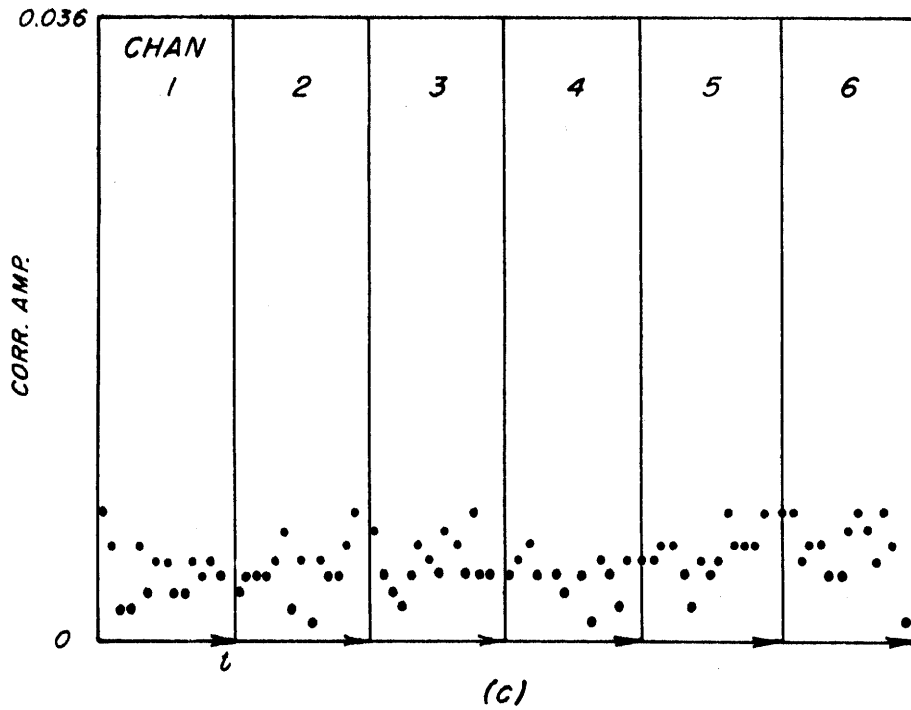


Figure 7-2 (contd.)

(c) Individual-channel correlation amplitudes vs. time.

Each point represents 2 seconds of data on the corresponding channel.

(d) Individual-channel residual fringe phase vs. time.

Each point represents 2 seconds of real integration time.

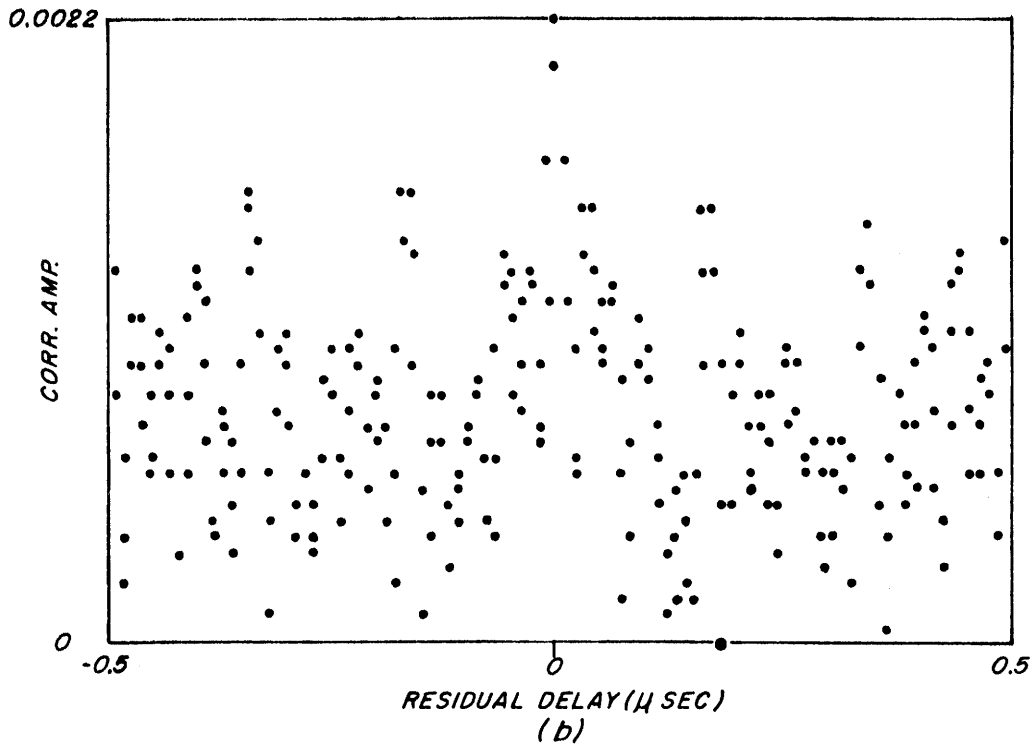
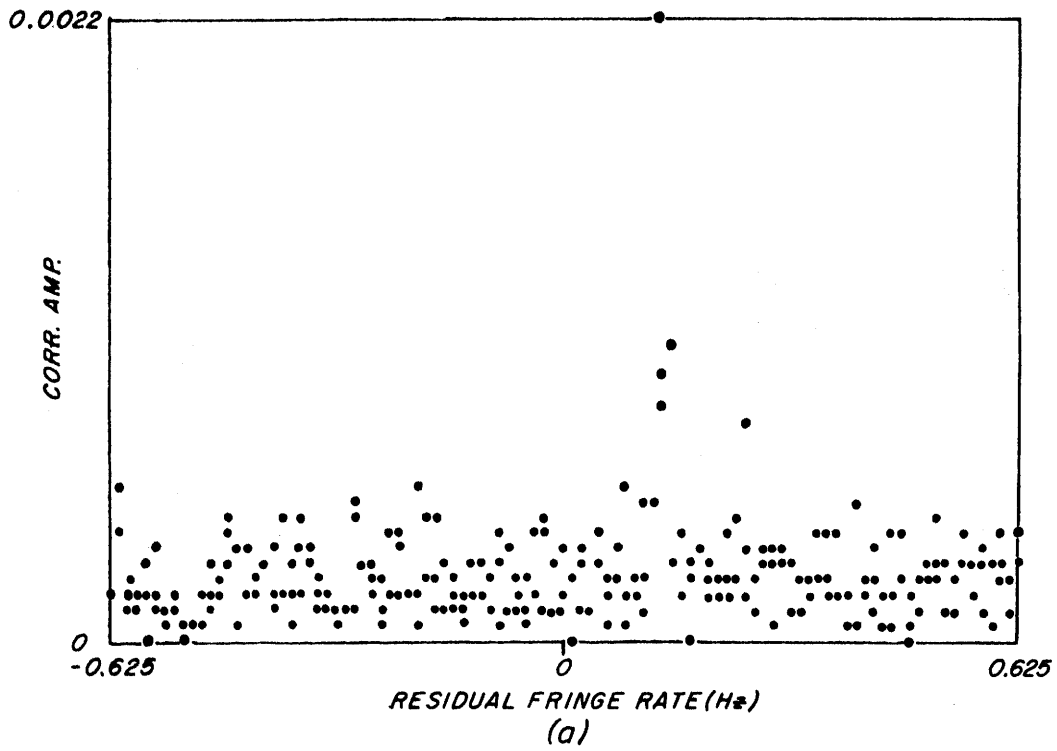


Figure 7-3

02 Oct 1969 1545UT 3C273B Haystack-OVRO X-band-switched

(a) Correlation amplitude vs. residual fringe rate.

(b) Delay resolution function amplitude vs. residual delay.

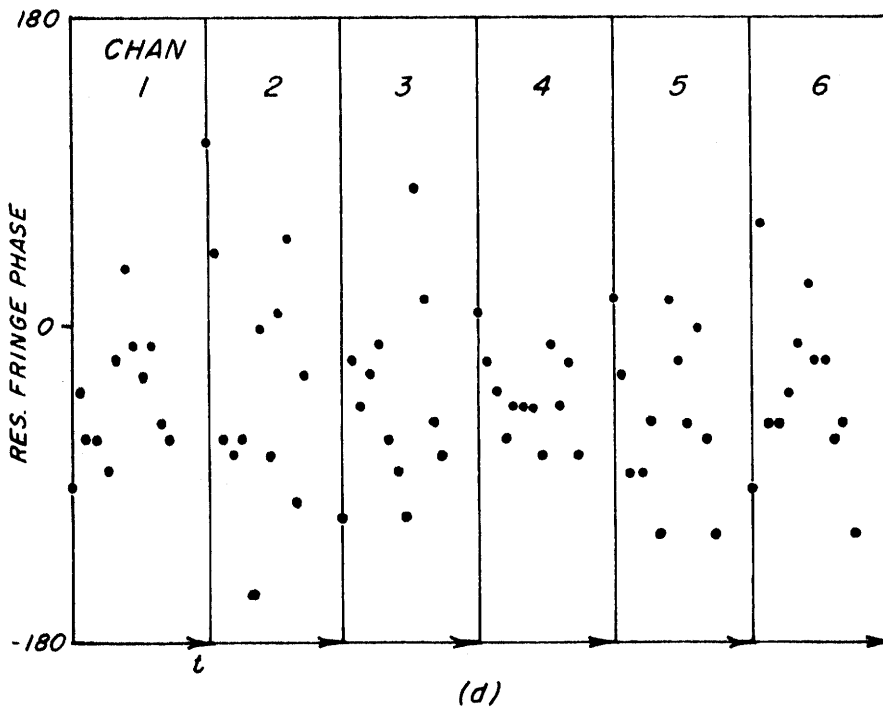
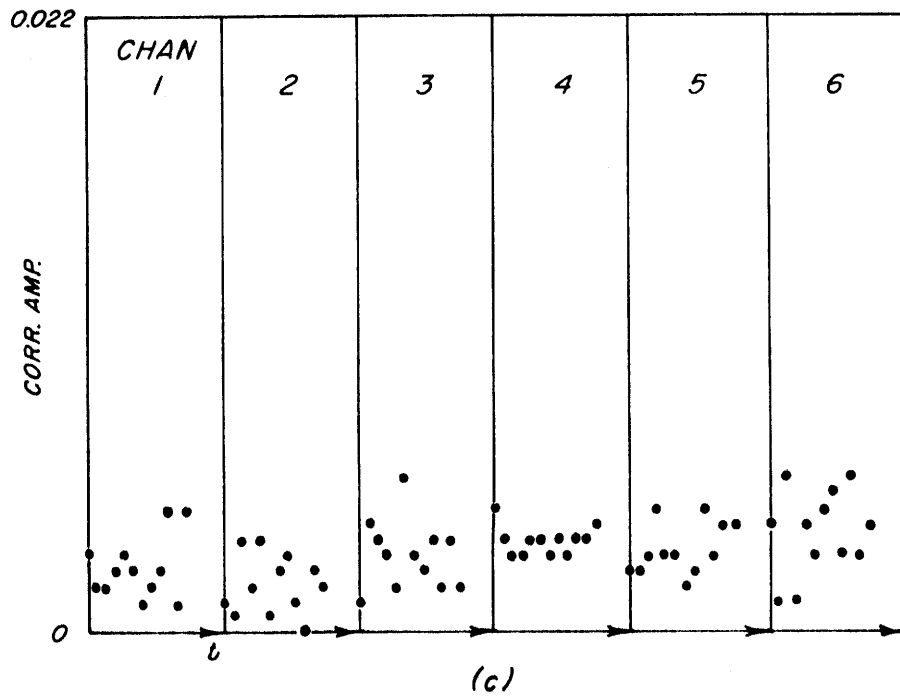


Figure 7-3 (contd.)

(c) Individual-channel correlation amplitudes vs. time.

Each point represents 2 seconds of data on the corresponding channel.

(d) Individual-channel residual fringe phase vs. time.

Each point represents 2 seconds of real integration time.

Table 7-3 Results of Switched-Frequency X-band Observation
on 3C273B at 1545 UT on 2 October 1969

	τ (μ sec)	σ_τ (nsec)	\hat{t} (nsec/ sec)	$\sigma_{\hat{t}}$ (psec/ sec)	Corr Amp	SNR	σ_ϕ^{**} (deg)	T (sec)
NRAO-Haystack								
A priori *	302.0051	0.11	184.8385	0.0041	-0.016	110	4.7	160
Measured	301.0135	--	184.8399	--	0.013	--	~15.	133
A posteriori ***	--	0.15	--	0.0065	0.013	83	5.8	133
NRAO-OVRO								
A priori	6710.6315	0.26	-643.6387	0.010	-0.007	45	10.7	160
Measured	6709.6266	--	-643.6555	--	0.0036	--	~40.	154
A posteriori	--	0.48	--	0.019	0.0036	25	20.8	154
Haystack-OVRO								
A priori	6408.6264	0.34	-828.4771	0.013	-0.005	35	15.0	160
Measured	6408.6200	--	-828.4950	---	0.0022	--	~50.	133
A posteriori	--	0.85	--	0.033	0.0022	14	34.0	133

* Based on parameters of Tables 5-1 and 7-1 with correlated flux estimated to be 15 f.u.

** RMS phase noise based on a 2-second integration period, which corresponds to the integration period of the points plotted in Figures 7-1b, 7-2b and 7-3b.

*** Based on measured correlation amplitude and integration time.

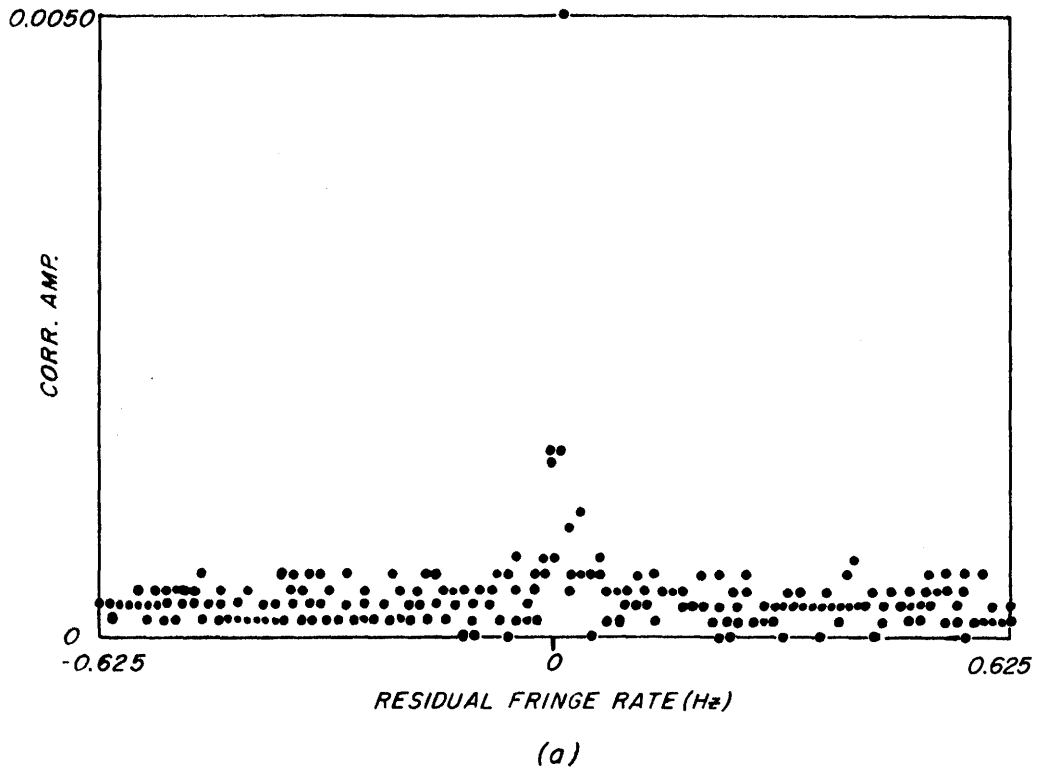


Figure 7-4

02 Oct 1969 1936 UT 3C273B NRAO-Haystack L-band-unswitched

(a) Correlation amplitude vs. residual fringe rate.

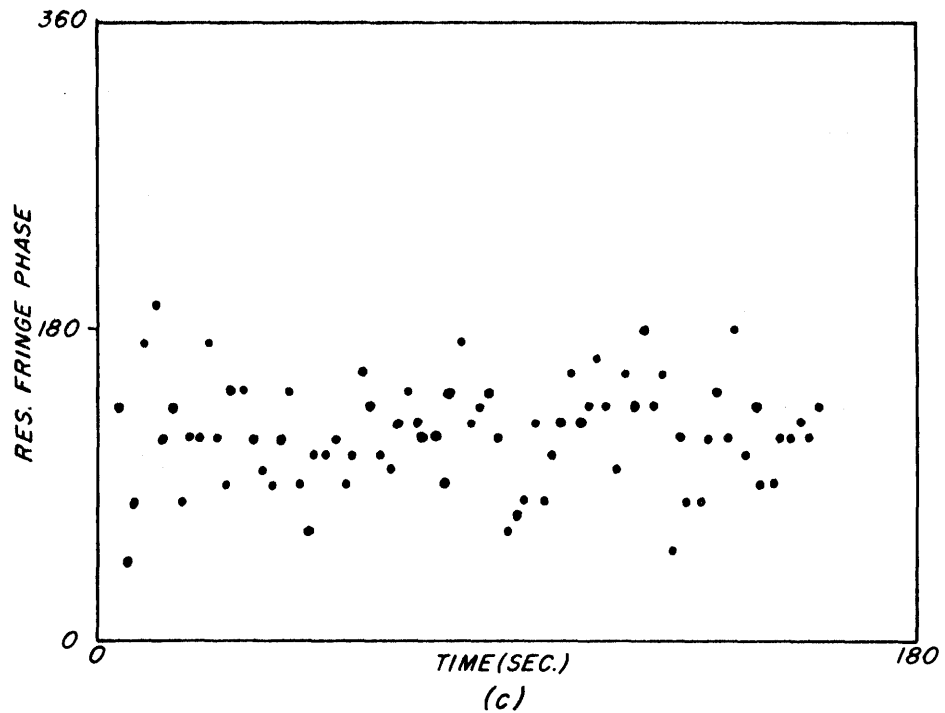
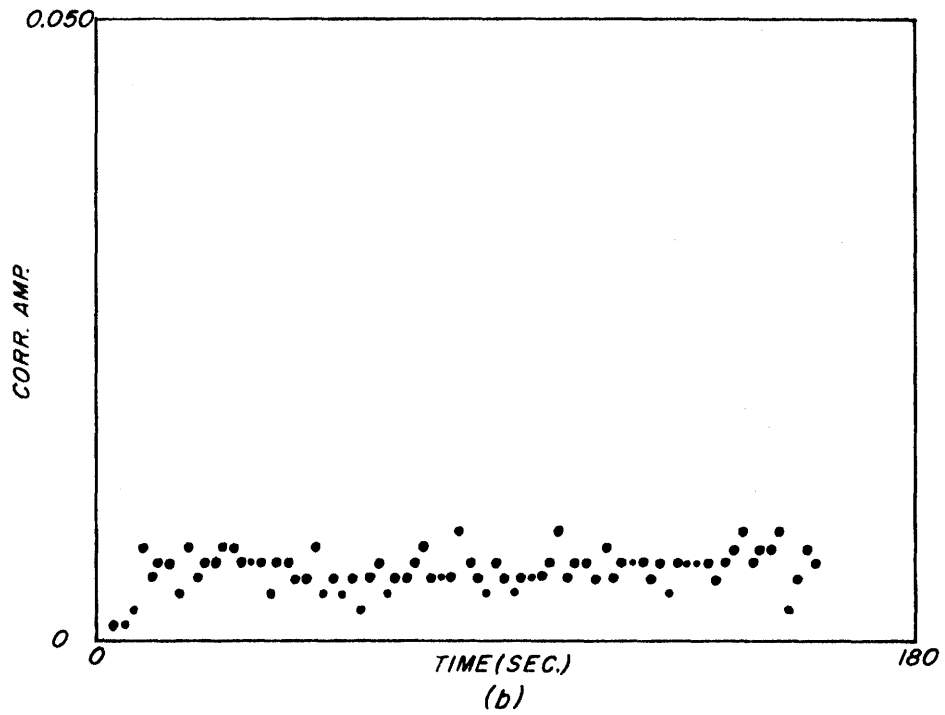


Figure 7-4 (contd.)

(b) Correlation amplitude vs. time.

Each point represents 2 seconds of data.

(c) Residual fringe phase vs. time.

Each point represents 2 seconds of data.

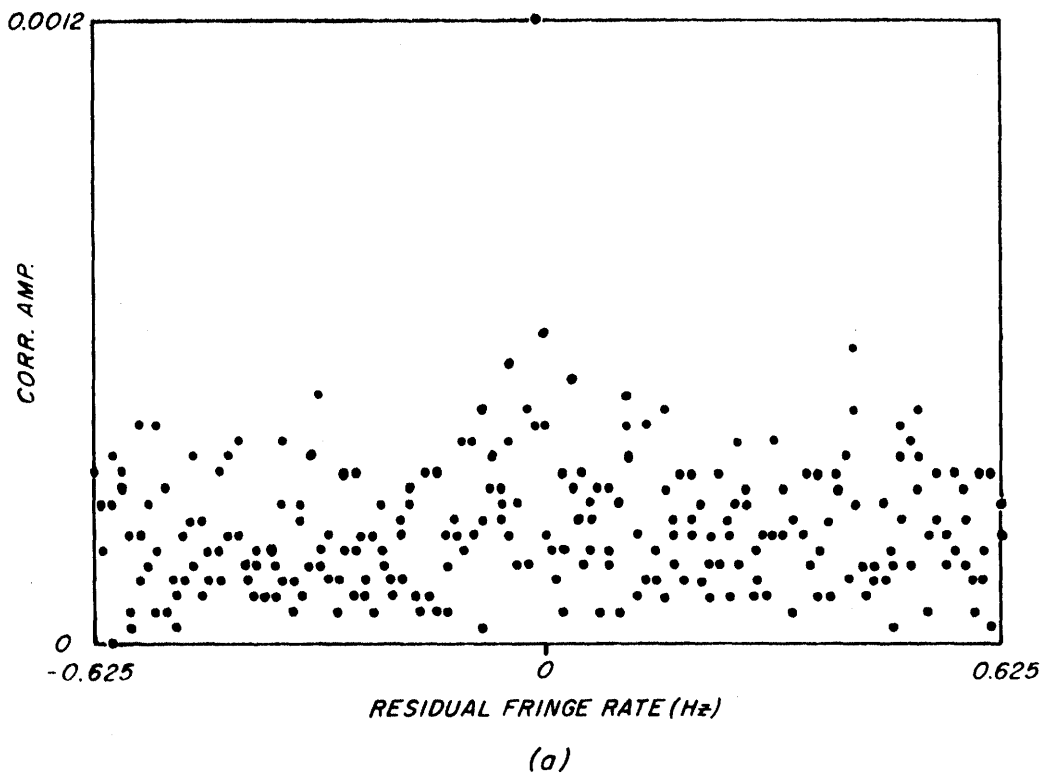


Figure 7-5

02 Oct 1969 1936 UT 3C273B NRAO-OVRO L-band-unswitched

(a) Correlation amplitude vs. residual fringe rate.

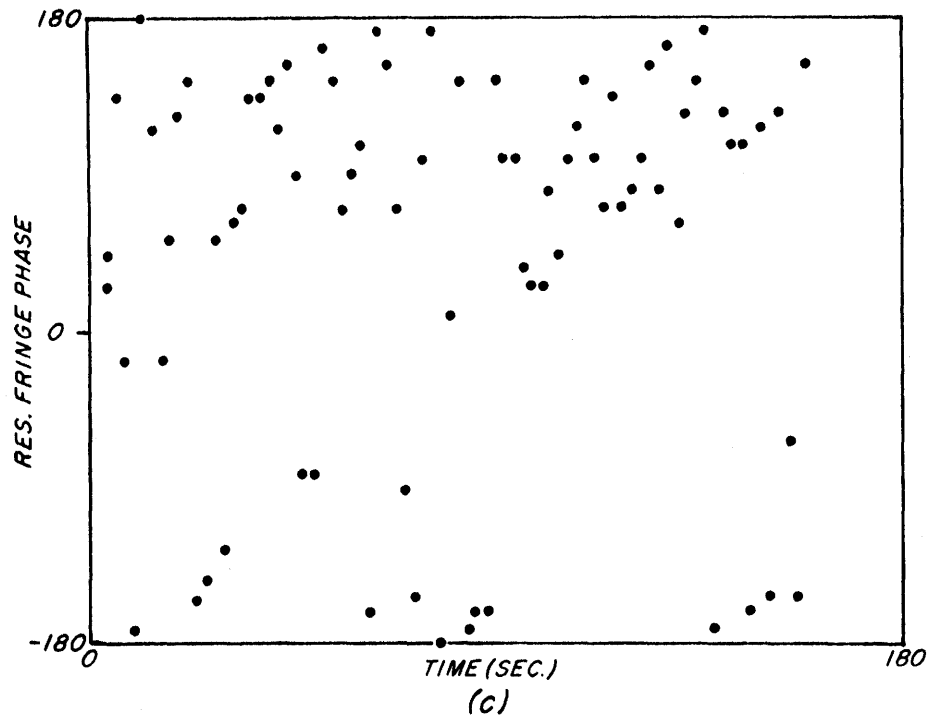
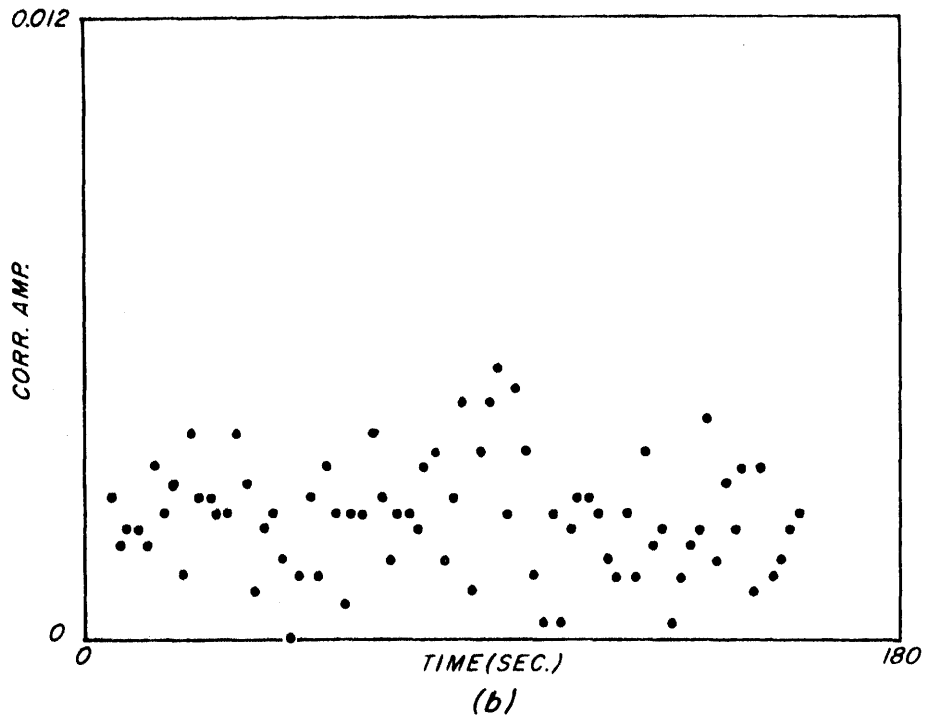


Figure 7-5 (contd)

(b) Correlation amplitude vs. time.

Each point represents 2 seconds of data.

(c) Residual fringe phase vs. time.

Each point represents 2 seconds of data.

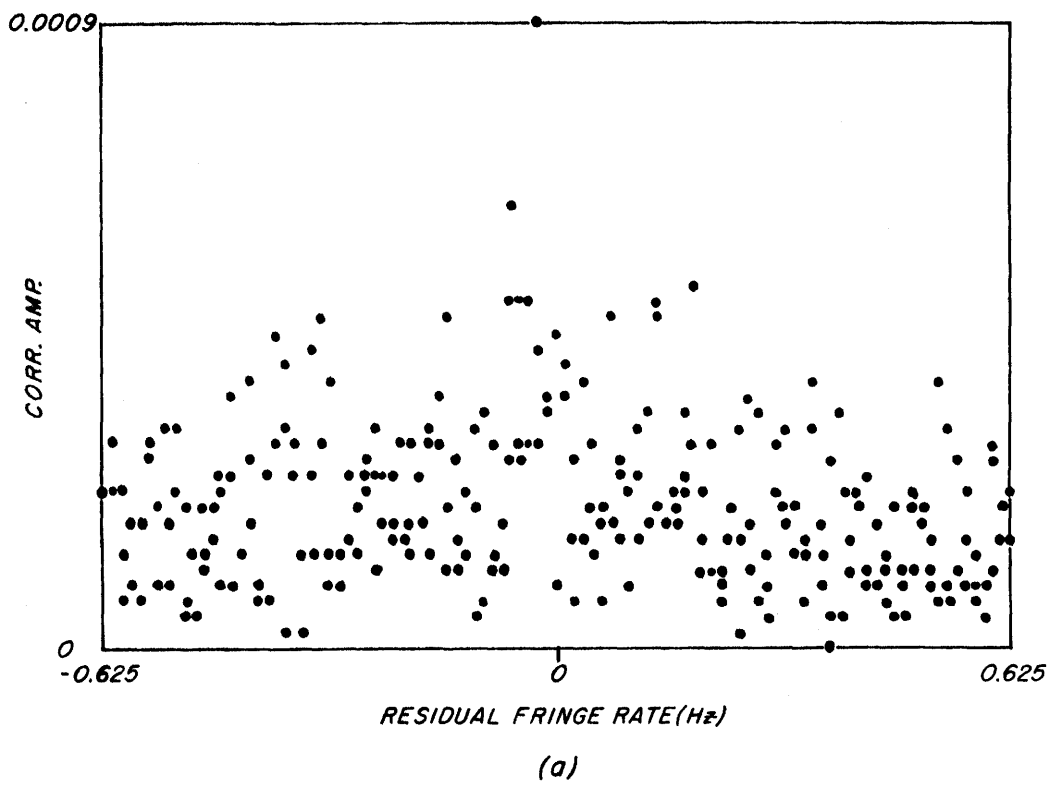


Figure 7-6

02 Oct 1969 1936 UT 3C273B Haystack-OVRO L-band-unswitched

(a) Correlation amplitude vs. residual fringe rate.

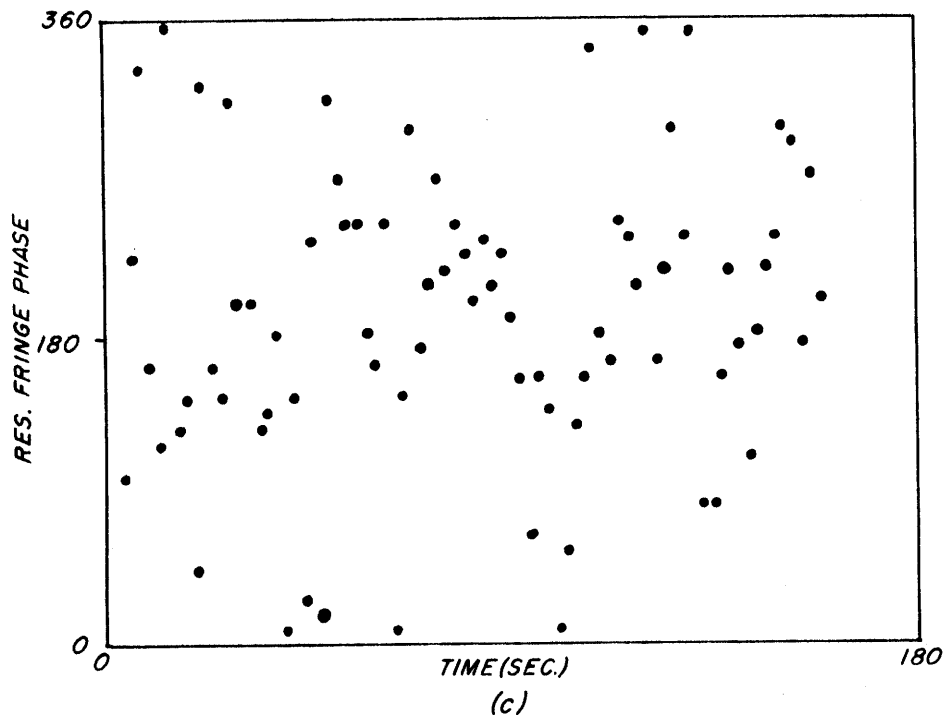
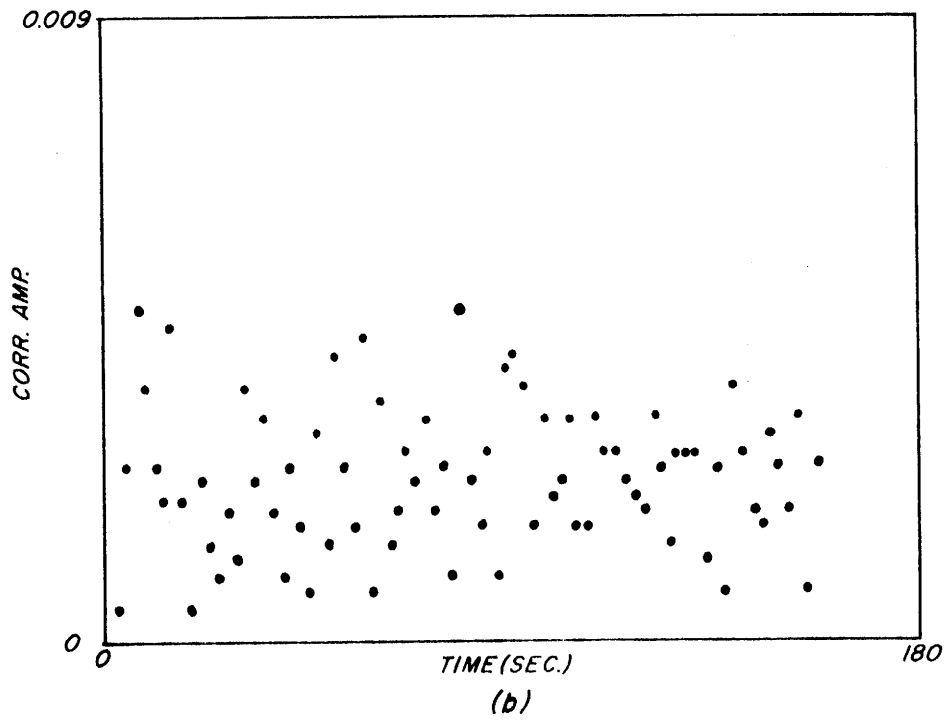


Figure 7-6 (contd.)

(b) Correlation amplitude vs. time.

Each point represents 2 seconds of data.

(c) Residual fringe phase vs. time.

Each point represents 2 seconds of data.

Table 7-4 Results of Single-Frequency L-band Observation
on 3C273B at 1936 UT on 2 October 1969

	τ (μ sec)	σ_τ (nsec)	\bar{t} (nsec/ sec)	σ_τ (psec/ sec)	Corr Amp	SNR	σ_ϕ (deg)	T (sec)
NRAO-Haystack								
A priori *	2282.073	0.10	76.1395	0.023	~0.015	100	5.0	160
Measured	2281.281	--	76.1443	--	0.005	--	~25.	154
A posteriori ***	--	0.29	--	0.065	0.005	35	15.0	154
NRAO-OVRO								
A priori	-3904.899	0.14	-755.5147	0.033	~0.010	70	7.5	160
Measured	-3905.780	--	-755.5242	--	0.0019	--	~50.	154
A posteriori	--	0.77	--	0.180	0.0019	13	40.0	154
Haystack-OVRO								
A priori	-6186.972	0.36	-831.6542	0.080	~0.004	28	18.7	160
Measured	-6187.136	--	-831.6688	--	0.0009	--	~90.	154
A posteriori	--	1.7	--	0.390	0.0009	6	83.0	154

* Based on parameters of Tables 5-1 and 7-1 with correlated flux estimated to be ~15 f.u.

** RMS phase noise based on a 2-second integration period, which corresponds to the integration period of the points plotted in Figures 7-4b, 7-5b and 7-6b.

*** Based on measured correlation amplitude and integration time.

7.3-5 Delay and Delay Rate Closure Tests

A powerful test of the internal consistency and quality of the data are the so-called "closure tests" in which the measured delays and delay rates are summed around the three baselines.² In order to understand exactly how the delays must be summed for these tests we must carefully define the meaning of "delay". Consider two stations, i and j , and let a marked wavefront reach station i at the instant that the reading on the station i clock is t_i . If the reading on the clock of station j is t_j at the instant this marked wavefront reaches station j , the delay τ_{ij} is defined as $\tau_{ij} \equiv t_j - t_i$.

Now consider a network of 3 stations, designated 1, 2 and 3. Suppose, for simplicity, that the clocks at all three stations are exactly synchronized (this assumption will not affect the closure result). Now let us examine a marked wavefront which arrives at station 1 at time $t = 0$. This wavefront will arrive at station 2 at $t = \tau_{12}$, at which time the delay from station 2 to station 3 will have increased by an amount $\tau_{12} \dot{\tau}_{23}$, where $\dot{\tau}_{23}$ is the delay rate of station 3 with respect to station 2 and where we have made the assumption (valid for this analysis) that $\dot{\tau}_{23}$ is constant over a time τ_{12} . Finally, the wavefront will arrive at station 3 at time $t = \tau_{23} + \tau_{12} \dot{\tau}_{23}$. But this must just be the delay between stations 1 and 3, τ_{13} , so that

²Not only is the "closure test" a good test of the data but also of the data reduction programs!

$$\tau_{13} = \tau_{12} + \tau_{23} + \tau_{12}\dot{\tau}_{23} \quad (7-1)$$

which is the so-called "delay closure" condition. We can now define a "delay closure error," $\Delta\tau_{\epsilon}$, given by

$$\Delta\tau_{\epsilon} \equiv \hat{\tau}_{12} + \hat{\tau}_{23} - \hat{\tau}_{13} + \hat{\tau}_{12}\hat{\dot{\tau}}_{23} \quad (7-2)$$

where $\hat{\tau}_{ij}$ and $\hat{\dot{\tau}}_{ij}$ are the estimated values of delay and delay rate, respectively.

By simple differentiation we obtain the corresponding "delay-rate closure error," $\Delta\dot{\tau}_{\epsilon}$, given by

$$\Delta\dot{\tau}_{\epsilon} \equiv \hat{\dot{\tau}}_{12} + \hat{\dot{\tau}}_{23} - \hat{\dot{\tau}}_{13} + \hat{\tau}_{12}\ddot{\tau}_{23} + \hat{\dot{\tau}}_{12}\hat{\tau}_{23} \quad (7-3)$$

where $\ddot{\tau}_{23}$ is the delay acceleration, usually computed a priori. The power of the closure tests of Equations (7-2) and (7-3) is that $\Delta\tau_{\epsilon}$ and $\Delta\dot{\tau}_{\epsilon}$ are completely model-independent since the closure condition of Equation (7-1) must be satisfied regardless of any vagarities of the real world!

We may now apply Equations (7-2) and (7-3) to compute the closure errors of the actual data. By far, the largest number of successful 3-baseline observations were taken on source 3C273B. The closure errors $\Delta\tau_{\epsilon}$ and $\Delta\dot{\tau}_{\epsilon}$ have been computed for these observations and the results are shown in Figure 7-7. The error bars for $\Delta\tau_{\epsilon}$ have been calculated as the root-mean-square of the σ_{τ} of the "a posteriori" model (see Tables 7-3 and 7-4), while the corresponding error bars for $\Delta\dot{\tau}_{\epsilon}$ are generally too small to be shown (see Tables 7-3 and 7-4 for typical values).

From Figure 7-7 we see that the delay closure errors, $\Delta\tau_{\epsilon}$, are scattered approximately in the manner one would expect for the given error bars. This indicates that the theoretical delay-estimation accuracies have probably been closely achieved in practice and that the actual data closely approximates the conditions assumed in the maximum-likelihood analysis for group delay.

The delay-rate closure errors, on the other hand, scatter much more widely than can be explained on the basis of their small individual error bars. Although not entirely explained, the probable reason for this wide scatter is that the assumptions of the delay-rate estimation algorithm have been violated in some manner. Recall from Section 4.5 that the delay rate estimation procedure fits only a constant residual delay rate to an observation and ignores any higher order residuals that may be present.³ If this assumption is violated, the theoretical values of the expected estimation errors will be badly underestimated. Based on the highly systematic character of many of the residual phases (see Section 7.3-4), the computed theoretical error bars are undoubtedly highly optimistic. The actual scatter of $\Delta\dot{\tau}_{\epsilon}$ is, in fact, a good indicator of the actual random errors in the delay-rate data.

³ Generally, for geodetic and astrometric purposes, it is not useful to attempt to fit higher order residuals to individual observations since there exists no provisions for modelling such short term (often random) in VLBI 3.

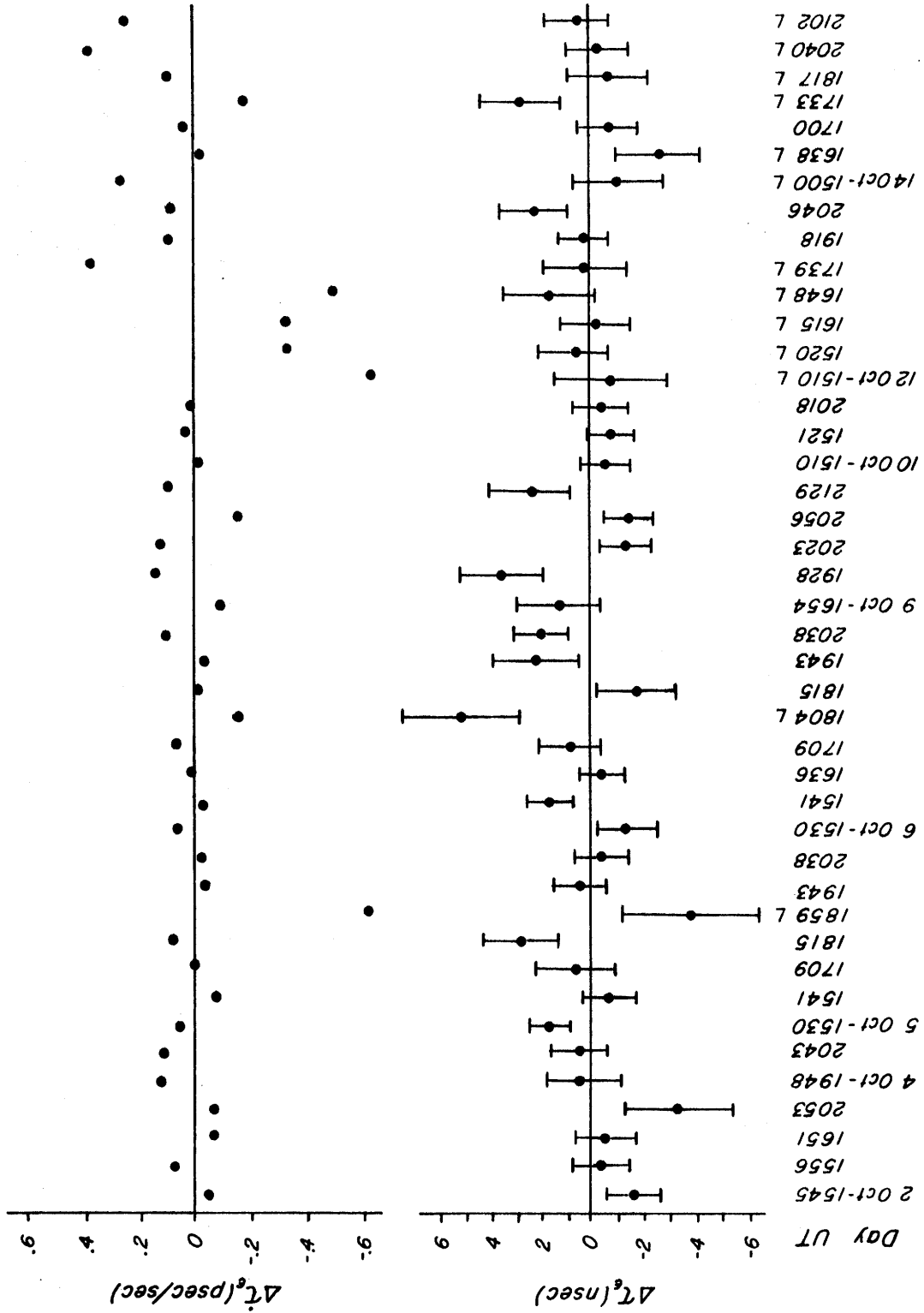


Figure 7-7

Delay and Delay-Rate Closure Errors for 3C273 Data. Error bars on ΔT_e are formal 1σ values based on observed correlation amplitudes. L-band observations are noted.

7.4 BASELINE AND SOURCE POSITION RESULTS

At the completion of processing through VLBI 1 and VLBI 2, the data were sorted and edited to remove all observations that were obviously bad. When this procedure was completed, approximated 500 observations remained, of which approximately 70% were from X-band. Each estimated delay and delay rate was assigned an error based on the measured correlation amplitude and length of the observation (see Tables 7-3 and 7-4 for examples). A correction for the ionosphere was also applied to the delay and delay rates, based on observations of Faraday rotation of radio signals transmitted from a satellite in synchronous orbit. The resulting corrections at L-band ranged up to ~3 nsec in delay and 1 psec/sec in delay rate and were roughly a factor of 25 smaller for X-band.

As we indicated earlier only about 500 out of the total of 3000 observations made were usable for the final processing through VLBI 3. Table 7-5 lists the number of usable delay and delay rate observations obtained for each source on each baseline. Note that the largest fraction of these observations are on the Haystack-NRAO baseline and that 3C273B observations form a large part of the total data set on the baselines involving OVRO. Also

	<u>Haystack-NRAO</u>		<u>OVRO-Haystack</u>		<u>NRAO-OVRO</u>	
	Delay	Rate	Delay	Rate	Delay	Rate
3C84	38	38	10	10	9	8
3C120	26	26	9	11	7	4
4C39.25	15	22	2	2	14	14
3C273B	74	75	59	68	87	98
3C279	44	32	1	2	0	6
3C345	32	32	3	3	19	17
VRO42.22.01	10	10	1	2	1	2
3C454.3	40	40	12	12	16	15

Table 7-5

Number of successful delay and delay-rate observations on each source and baseline.

note that generally, except for observations on 3C279, more observations yielded good information about delay rate than delay. This is due to the fact that delay-rate data is generally useful to a lower signal-to-noise ratio than switched-frequency delay data. The observations on 3C279, on the other hand, suffered severe fringe phase corruption due to phase scintillation caused by the solar corona, rendering some delay rate (i.e. fringe rate) data useless while the group delay information, which depends only on the difference phases between the frequency windows, remained good.

The data can be analyzed through VLBI 3 in a variety of ways to judge its accuracy and internal consistency. Source positions may, for example, be fixed at a priori values determined by other means and the baselines independently solved for. Or, conversely, the baselines may be fixed and the source positions solved for. The data from each baseline may be examined individually to solve for baseline and source coordinates, or, alternatively, the L-band and X-band data may be used in separate solutions. Many such computer experiments were performed on the data in order to try to uncover and/or understand systematic errors that appeared. On the basis of the results of these

many computer experiments the real accuracy of the data may be estimated and the final "best" estimates of the solution parameters are obtained.

7.4-1 Baseline Solutions

Figure 7-8 shows the baseline component solutions for examples of nine different processings of the data through VLBI 3. Table 7-6 lists the data and constraints that were applied in each of these solutions. In each case where the source positions were solution parameters, the right-ascension origin was defined by fixing the right-ascension of 3C273B at the value given by Hazard (Reference 48). The error bars shown in Figure 7-8 are based on the rms post-fit residuals. Clearly, the solutions for the Haystack-NRAO baseline appear to be the most internally consistent as the solutions for all three components fall within a range of ~1 meter. The scatter of the solutions on the two baselines involving OVRO are much larger, on the order of several meters. The error bars, however, are also significantly larger, indicating primarily the relative lack of data on those two baselines. Note particularly the very large error bars on the Haystack-OVRO and NRAO-OVRO baseline solutions using L-band data alone; this primarily reflects the fact that very little useful L-band data was obtained on those two baselines.

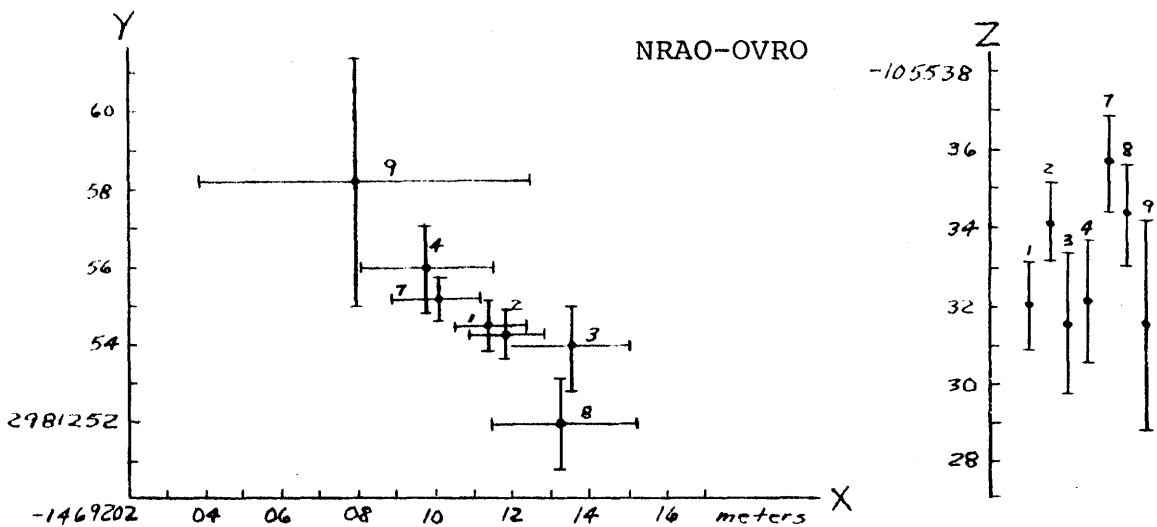
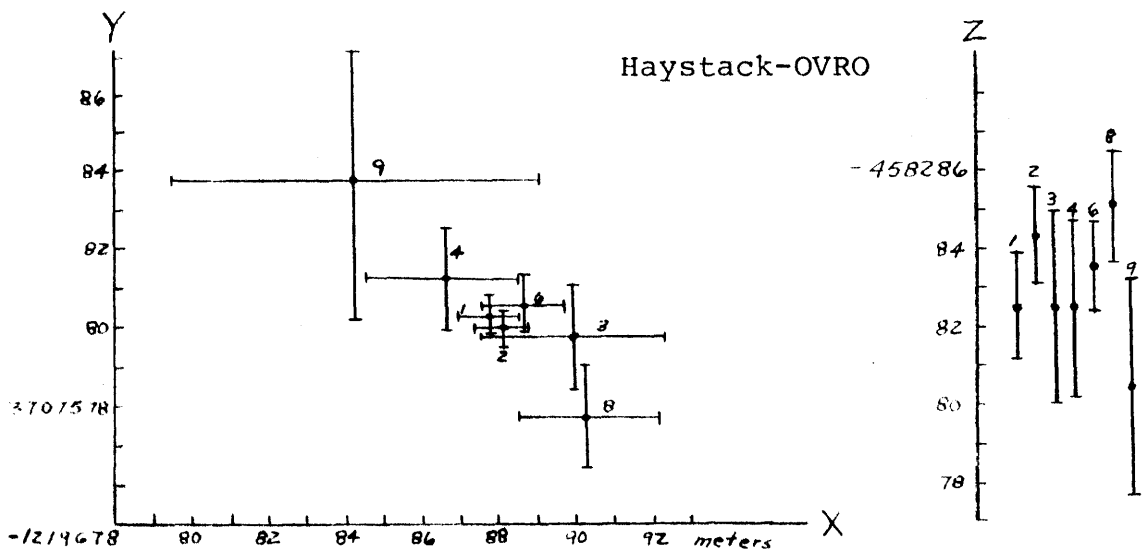
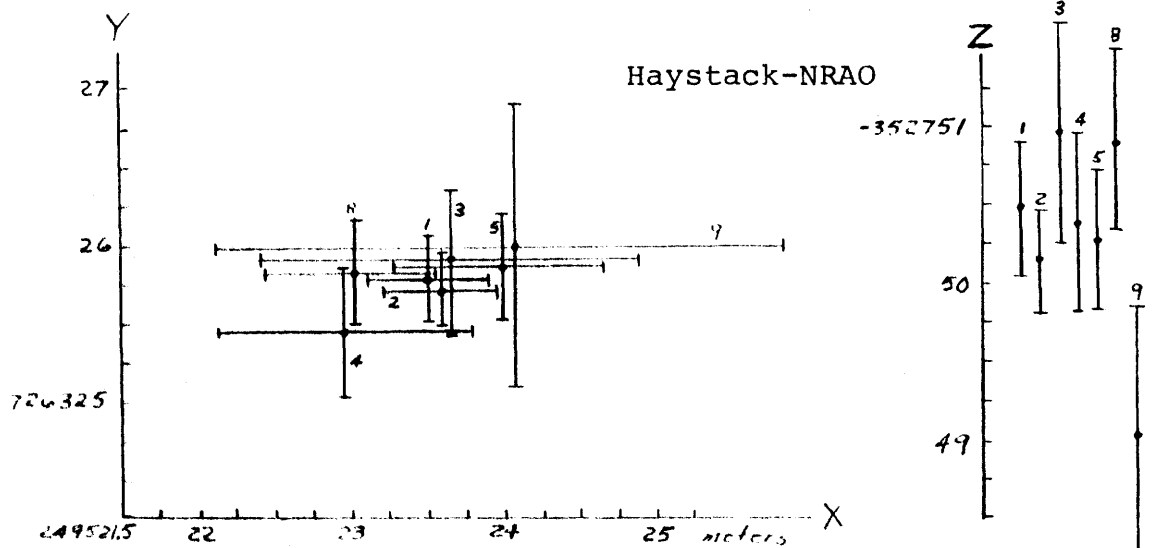


Figure 7-8
Baseline Solution Results

1	Solved*	All baselines [‡] ; delay and delay rate	88
2	Fixed**	All baselines; delay and delay rate	73
3	Solved	All baselines; delay only	88
4	Fixed	All baselines; delay only	73
5	Fixed	Haystack-NRAO only; delay and delay rate	34
6	Fixed	Haystack-OVRO only; delay and delay rate	43
7	Fixed	NRAO-OVRO only; delay and delay rate	34
8	Fixed	All baselines; X-band only; delay and delay rate	73
9	Fixed	All baselines; L-band only; delay and delay rate	43

Table 7-6

Processing parameters for results shown in Figure 7-8.

* Hazard's 3C273B right-ascension (Reference 48) used to fix right-ascension origin.

**The source positions were fixed at the best available coordinates, which, at the time of the processing of the October 1969 data through VLBI 3, were those determined from a 1971-1972 series of switched-frequency experiments performed by our group between Haystack and NASA 210' antenna in California, and are very close to the positions given in Reference .

‡The sum of the solution baseline vectors is automatically constrained to be zero.

†Includes total number of solution parameters, including clock parameters and, in some cases, the zenith atmospheric delay at each site.

It is interesting to examine the baseline lengths corresponding to the solutions of Figure 7-8. These are listed in Table 7-7 along with their formal errors based on post-fit residuals. For the Haystack-NRAO baseline the solutions all fall within a range of ~ 0.9 meters, which is comparable to the range of the solutions of the baseline components shown in Figure 7-8. On the two baselines involving OVRO, however, the range of baseline lengths is ~ 3 meters, which is about a factor of three smaller than the range of the solutions of the corresponding baseline components. This indicates that the orientation of these two baselines is less well-determined than their lengths. From Figure 7-8 we see that the orientation uncertainty appears to be largely confined to the equatorial plane (i.e., the x and y components), which is directly related to an uncertainty in the right-ascension origin. In fact, if we were to draw a line of constant-baseline-length on the x-y plots of Figure 7-8 (assuming a fixed z-component), the x and y baseline component solutions for the baselines involving OVRO would scatter fairly closely about this line.

7.4-2 Baseline Closure Test

An important test of the internal consistency

<u>Solution #</u>	<u>Haystack-NRAO</u>	<u>Haystack-OVRO</u>	<u>NRAO-OVRO</u>
1	845129.6±0.3 **	3929861.7±0.5	3325296.6±0.4
2	845129.4±0.3	3929861.8±0.4	3325296.8±0.4
3	845129.9±0.7	3929862.1±1.6	3325297.2±1.3
4	845129.0±0.5	3929862.3±1.2	3325297.1±1.0
5	845129.8±0.2	---	---
6	---	3929862.1±0.6	---
7	---	---	3325296.7±0.4
8	845129.7±0.3	3929860.3±0.7	3325295.1±0.6
9	845129.1±	3929863.8±2.8	3325298.5±2.2

Table 7-7
Baseline-Length Solutions (meters) *

* For all baseline solutions: At Haystack, the reference point is the intersection of the azimuth and elevation axes of the antenna. At NRAO, the reference point is the intersection of the polar axis with a line perpendicular to both the polar and declination axes. At OVRO, the reference point is the intersection of the polar and declination axes.

** Errors are based on rms post-fit residuals. The "true" 1σ errors are estimated to be 3-4 times larger (see Section 7.4-3).

of the data is the "baseline closure test", in which each of three baselines is solved for independently; then, if the data are internally consistent, the sum of the resulting baseline solution vectors should be zero to within the expected error. Solutions 5, 6, and 7 of Figure 7-8 and Table 7-6 are three such independent baseline solutions. For each of these solutions the source positions were fixed at the best a priori coordinates and only clock parameters and baseline components solved for (because of the relatively few data points on the baselines involving OVRO, simultaneous solutions of baseline components and source positions yielded large error bars and little useful information). The solution baseline components and their sums around the three baselines are given in Table 7-8. We see that the "closure errors" are somewhat larger than the formal errors would imply, but are nevertheless approximately consistent with what we might expect on the basis of the scatter of the solutions shown in Figure 7-8. On the basis of these closure tests and the scatter of the solutions in Figure 7-8, we can judge that the true ac-

curacy of the solution baseline components is on the order of several meters, somewhat poorer than originally expected but roughly consistent with the observed scatter in the baseline solution of Figure 7-8.

Table 7-8 Results of Baseline Closure Test for
Independently-Determined Baseline Solutions*

	X (meters)	Y (meters)	Z (meters)
Haystack-NRAO	249524.0 ± 0.7	726325.9 ± 0.4	-352750.3 ± 0.5
Haystack-OVRO	-1219688.5 ± 0.8	3707580.3 ± 0.8	-458283.8 ± 1.0
NRAO-OVRO	-1469210.0 ± 1.2	2981255.1 ± 0.6	-105535.9 ± 1.2
CLOSURE ERROR**	2.5m	0.7m	-2.4m

* Source positions fixed at current best estimates.

** Computed as (Haystack-NRAO) + (NRAO-OVRO) - (Haystack-OVRO).

7.4-3 Character of the Post-Fit Residuals

An examination of the post-fit delay and delay-rate residuals may yield some clue as to the reasons that the apparent accuracy of the baseline solution is somewhat poorer than would be theoretically expected. Figures 7-9 and 7-10 show some typical examples of post-fit delay and delay-rate residuals. The particular residuals shown are from baselines Haystack-NRAO and Haystack-OVRO of Solution 1 of Table 7-8, and include only a fraction of the data actually used in Solution 1. Strong non-random trends are evident in both the delay and delay-rate residuals, the trends being quite different even from day to day on the same source. The trends in the delay residuals, being on the order of a few nanoseconds, are consistent with the scatter of a few meters in the baseline solutions (although, of course, the correspondence is not exact; see Section 2.3-1). The trends in the delay-rate residuals, being on the order of $\sim 1-2$ picosec/sec, are likewise roughly consistent with the observed scatter in the baseline solutions.

The most-likely explanations for the non-random trends in the delay and delay-rate residuals were drifts and instabilities in the experiment instrumentation.

The delay-rate trends were probably caused by drifts in both the first and

second local oscillators relative to the hydrogen-maser frequency standard. The trends in the delay residuals, on the other hand, may have been caused either by drifts between the relative phases of the switched-frequency local-oscillators or by changes in the frequency-dispersive characteristics of any of the RF or IF amplifiers or filters. Despite many efforts to find an alternative explanation which might be amenable to modelling in VLBI 3, the observed systematic trends in the delay and delay-rate residuals remain mostly unexplained. And, since most of the instrumentation was dismantled soon after the conclusion of the experiment, extensive tests to determine the exact causes were not possible. We might add, however, that data taken in more recent experiments, using much-improved equipment, still show some unexplained systematic trends in the residual delays and delay-rates, although they are reduced by more than an order of magnitude from those observed in the October 1969 data (Reference 50). This lends support to the conclusion that the systematic trends in the residuals of the October 1969 data stem mostly from instrumental unknowns rather than from deficiencies in the models used by VLBI 3.

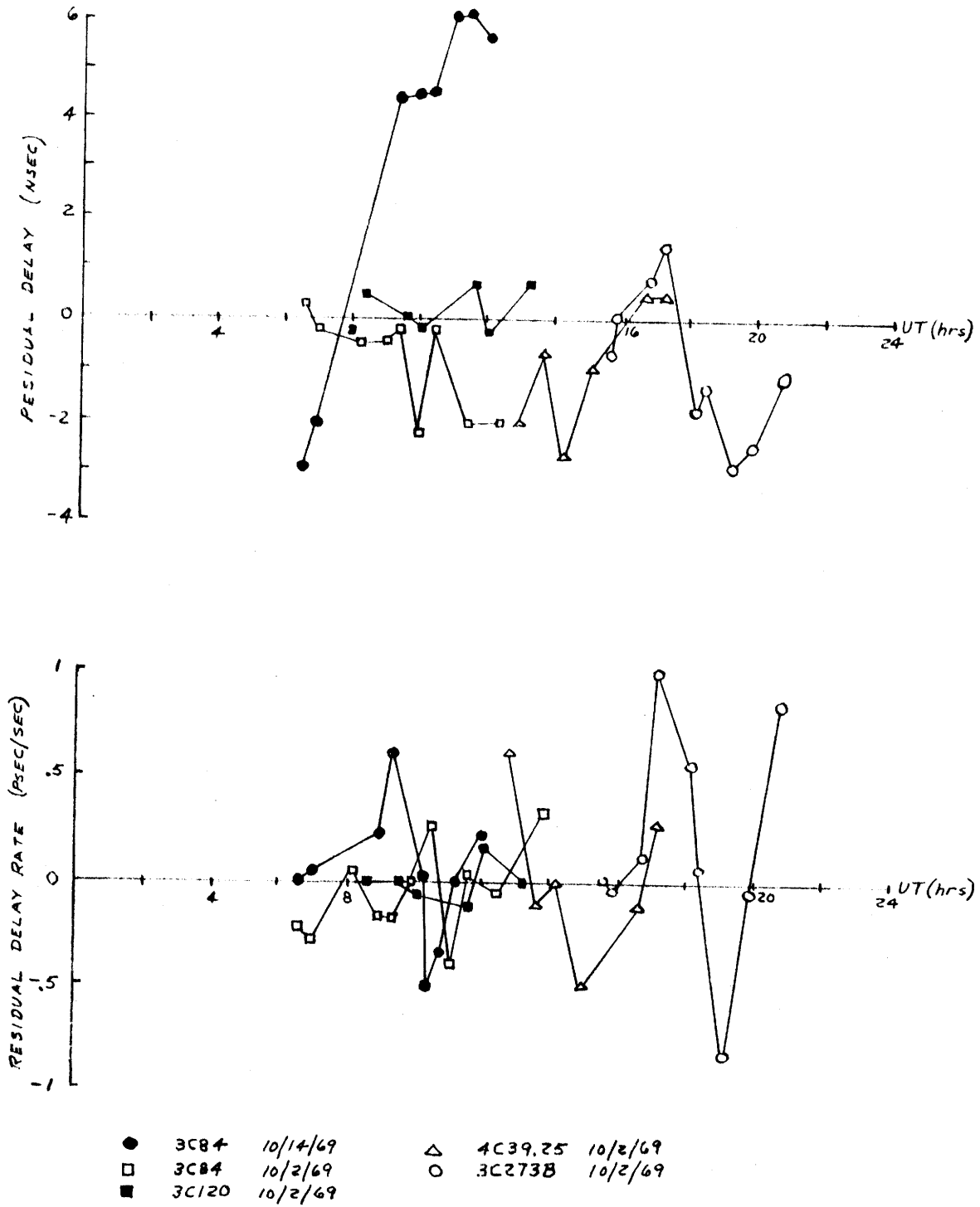


Figure 7-9
Typical Delay and Delay-Rate Residuals
from the Haystack-NRAO baseline

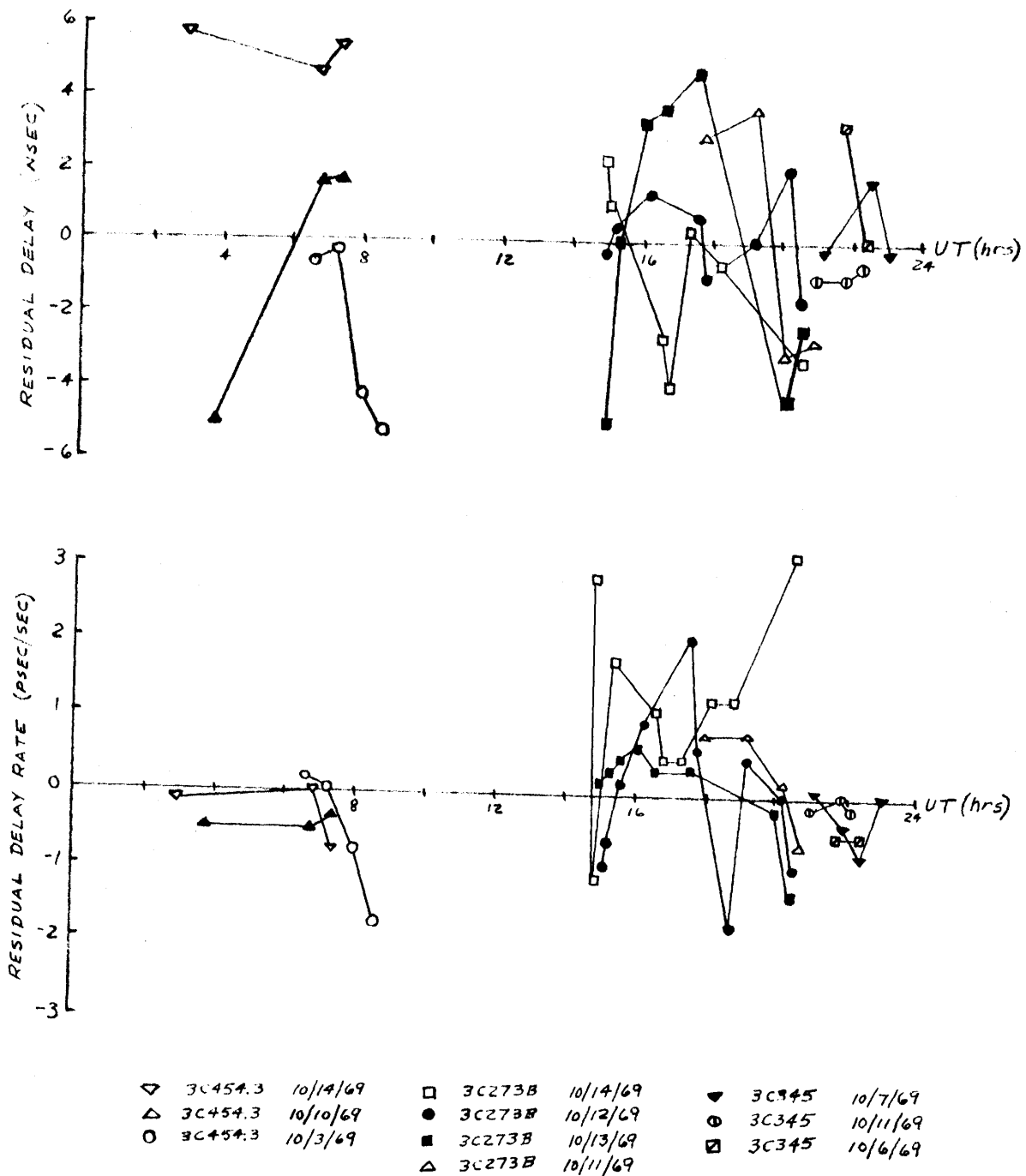


Figure 7-10
 Typical Delay and Delay-Rate Residuals
 from the Haystack-OVRO Baseline

7.4-4 Comparison of VLBI and Survey Baseline Results

A special high-precision ground survey was conducted in 1972 under the direction of the First Geodetic Survey Squadron, Warren Air Force Base, Wyoming, to measure the positions of the Haystack, NRAO, and OVRO antennas (Reference 51). The accuracy of this survey, although difficult to judge, is believed to be good at least to the ten-meter level in all components. Due to the fact that there may be some difference between the right-ascension origins of the VLBI and survey results, it may be expected that a slight rotation in the equatorial plane might be necessary in order to make a meaningful comparison. Table 7-9 shows the results of such a comparison. The VLBI results are from Solution 1 of Table 7-6 and have been rotated in the equatorial (i.e. X-Y plane) by 0.9 arc-seconds in the direction of increasing right-ascension in order to approximately minimize the least-squares error between the VLBI and survey results. The agreement is to within 6.4 meters in all components and to within 3.7 meters in length. In view of the fact that the estimated accuracy of the VLBI results is a few meters, these disagreements are not particularly surprising. And, since these particular

baselines have not, at least to date, been redetermined with improved instrumentation, it is difficult to draw any meaningful conclusions about the comparative accuracy of the VLBI and survey results. The only other VLBI determination of any of these baselines was the Haystack-NRAO baseline by Hinteregger in January 1969 (see Section 1.3 and Reference 14). His results are in disagreement with the VLBI results presented here by as much as ~5 meters, again indicating that the accuracy of the VLBI measurements up to and including October 1969 is on the order of a few meters.

An improved VLBI determination of the Haystack-NRAO baseline, however, should be soon available at which time it should be possible to make a more meaningful judgment of the results presented here.

Table 7-9 Comparison of VLBI and Survey Baseline Results*

	X (meters)	Y (meters)	Z (meters)	Length
Haystack-NRAO				
Survey	249516.0	726328.9	-352749.7	845129.7
VLBI	249519.0 ± 0.4	726326.5 ± 0.3	-352750.5 ± 0.7	845129.6 ± 0.3
S-V	-3.0	+2.4	+0.8	+0.1
Haystack-OVRO				
Survey	-1219713.1	3707575.2	-458288.1	3929865.4
VLBI	-1219709.8 ± 0.7	3707571.8 ± 0.7	-458282.5 ± 1.7	3929861.7 ± 0.5
S-V	-3.3	+3.4	-5.6	+3.7
NRAO-OVRO				
Survey	-1020864.9	3162957.2	-105538.4	3325297.2 ± 0.
VLBI	-1020866.5 ± 0.8	3162955.6 ± 0.6	-105532.0 ± 1.5	3325296.6 ± 0.4
S-V	+1.6	+1.6	-6.4	+0.6

* VLBI results are from Solution 1 of Table 7-6; the right-ascension origin has been rotated by +0.9 arc-seconds from the results given in Figure 7-8 in order to make a more meaningful comparison with the survey results (see text).

** VLBI baseline errors are based on rms post-fit residuals. The "true" 1σ errors are estimated to be 3-4 times larger due to unexplained systematic errors.

7.4-5 Source Position Results

From the data of the October 1969 experiment, the positions of the eight sources of Table 7-1 were determined, except for the right-ascension of 3273B, which was fixed at the value given by Hazard (Reference 48) to provide a right-ascension origin³. The results are given in Table 7-10 along with the comparable results from other data and experimenters using both VLBI and other techniques. The quoted error bars for the October 1969 solutions are based on the post-fit residuals.

Unlike the baseline solutions, it is possible to draw some conclusions about the accuracy of the source position results since more recent switched-frequency VLBI experiments, most notably those using the Haystack-Goldstone baseline (Reference 50), have considerably increased the accuracy of the position solutions. In all cases, except for the right-ascension of 3C84 and the declination of 3C120, the error bars of the October 1969 solutions overlap the solution with the smallest quoted error. The discrepancy in the 3C84 right-ascension is small, and the October 1969 solution for the declination of 3C120 falls amidst the rather large scatter of the variously-determined declination for that source. It is interesting to note that

³ It would also have been possible to have incorporated an a priori covariance matrix to constrain the right-ascension corrections, but the small improvement perhaps attainable in relating our reference to FK4 did not seem worth the effort.

the October 1969 solution for this coordinate is the only October 1969 solution coordinate which is in poor agreement with the other switched-frequency VLBI coordinates (Reference 50).

The large error bars on the October 1969 declination solutions of 3C273B and 3C279 are due to the very low declinations of these sources, which causes the measurements of delay and delay rate to be relatively insensitive to small changes in declination (see Chapter 2). The error bars on 3C279 are particularly large because of the relative scarcity of measurements on that source (see Table 7-5).

As a whole, the source position solutions for the October 1969 experiment are in quite close agreement with the best positions now available and, at the time the 1969 measurements were made, represented a significant improvement in source position measurements.

Table 7-10 Comparisons of Source Positions

Source	α (1950.0)	δ (1950.0)	Technique*	Ref.		
3C84	03 16	29.546 \pm 0.007†	41 19	51.79 \pm 0.1	VLBI	Oct 69 ^e
		29.539 \pm 0.01		51.75 \pm 0.2	VLBI	50
		29.568 \pm 0.007		51.84 \pm 0.1	VLBI	52
		29.58 \pm 0.04		52.1 \pm 0.4	SBI	53
		29.548 \pm 0.014		52.19 \pm 0.12	Photo	54
3C120	04 30	31.580 \pm 0.01	05 14	59.87 \pm 0.3	VLBI	Oct 69
		31.586 \pm 0.005		59.2 \pm 0.1	VLBI	50
		31.605 \pm 0.007		60.19 \pm 0.3	VLBI	52
		31.57 \pm 0.03		60.1 \pm 0.4	SBI	53
		31.605 \pm 0.024		60.10 \pm 0.3	Photo	55
		31.60 \pm 0.015		58.9 \pm 0.3	Photo	49
4C39.25	09 23	55.302 \pm 0.007	39 15	23.81 \pm 0.1	VLBI	Oct 69
		55.296 \pm 0.004		23.73 \pm 0.04	VLBI	50
		55.33 \pm 0.04		23.5 \pm 0.4	SBI	53
		55.29 \pm 0.01		24.3 \pm 0.2	SBI	56

Table 7-10 Comparisons of Source Positions (cont'd)

Source	α (1950.0)	δ (1950.0)	Technique*	Ref.
3C273B	12 26 33.246**	02 19 43.42 \pm 0.6	VLBI	Oct 69
	33.246**	43.2 \pm 0.1	VLBI	50
	33.242 \pm 0.01	42.6 \pm 1.1	VLBI	52
	33.246 \pm 0.01	43.38 \pm 0.1	Photo	48
3C279	12 53 35.811 \pm 0.02	-05 31 07.27 \pm 1.0	VLBI	Oct 69
	35.831 \pm 0.004	08.0 \pm 0.1	VLBI	50
	35.825 \pm 0.01	07.65 \pm 0.45	VLBI	52
	35.82 \pm 0.02	07.4 \pm 0.6	SBI	56
	35.67 \pm 0.10	05.7 \pm 1.9	SBI	57
	35.82 \pm 0.02	07.65 \pm 0.25	Photo	49
3C345	16 41 17.642 \pm 0.01	39 54 10.99 \pm 0.2	VLBI	Oct 69
	17.634 \pm 0.004	11.00 \pm 0.07	VLBI	50
	17.613 \pm 0.01	11.15 \pm 0.15	VLBI	52
	17.64 \pm 0.04	10.6 \pm 0.4	SBI	53
	17.61 \pm 0.01	11.3 \pm 0.2	SBI	56
	17.606 \pm 0.013	10.72 \pm 0.12	Photo	54
	17.56 \pm 0.03	10.7 \pm 0.3	Photo	49

Table 7-10 Comparisons of Source Positions (cont'd)

Source	α (1950.0)	δ (1950.0)	Technique*	Ref.		
VRO 42.22.01	22 00	39.399 \pm 0.02	42 02	08.18 \pm 0.2	VLBI	Oct 69
		39.394 \pm 0.007		08.3 \pm 0.1	VLBI	50
		39.365 \pm 0.01		08.53 \pm 0.15	VLBI	52
		39.31 \pm 0.04		09.0 \pm 0.4	SBI	53
		39.37 \pm 0.01		08.8 \pm 0.2	SBI	56
3C454.3	22 51	29.526 \pm 0.01	15 52	54.21 \pm 0.2	VLBI	Oct 69
		29.530 \pm 0.009		54.24 \pm 0.03	VLBI	50
		29.524 \pm 0.01		54.36 \pm 0.2	VLBI	52
		29.54 \pm 0.03		54.9 \pm 0.4	SBI	53
		29.6 \pm 0.10		56.3 \pm 3.1	SBI	57
		29.533 \pm 0.011		54.98 \pm 0.15	Photo	54
		29.485 \pm 0.035		54.45 \pm 0.4	Photo	49

-302-

* SBI = Short-Baseline Interferometry

** This value was used to define the right ascension origin.

† Quoted errors for October 1969 results are based on rms post-fit residuals.

@ The positions given are from Solution 1 of Table 7-6.

7.5 CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

The switched-frequency VLBI technique has been demonstrated as a means of improving the quality of VLBI measurements for purposes of geodesy and astrometry, although much work remains to be done to develop the technique to its full potential. The problems in the October 1969 experiment appear to have been predominantly instrumental and will undoubtedly be solved as state-of-the-art technology advances. Already much improvement has been made in instrumentation and at least a tenfold increase in measurement accuracy has been demonstrated in recent experiments conducted by our group between Haystack and the NASA 210' Goldstone antenna. The greatest remaining stumbling block now appears no longer to be instrumentation stability, but rather calibration of the excess electrical path lengths contributed by the ionosphere and the neutral atmosphere. The calibration of the ionosphere can, in principle, be solved by making simultaneous measurements at two widely-spaced sets of frequencies, thus taking advantage of the known frequency-dispersive characteristics of an ionized plasma. The calibration of the neutral atmosphere is more difficult. The dry component of the neutral atmosphere may perhaps

be estimated to the few-centimeter level by relatively straightforward modelling, but estimation of the wet component (due primarily to water vapor in the lower troposphere) appears to be substantially more difficult. Present uncertainties in the wet component are estimated to be as much as 30-40 cm. One hopeful approach to the solution of this problem is the use of passive radiometers observing near the ~22 GHz natural water-vapor resonance frequency, although as yet this technique is unproven.

In the next few years it is perhaps reasonable to expect that the overall uncertainties in switched-frequency VLBI measurements will approach the few-centimeter level. When this accuracy is achieved it should be possible to determine the positions of a set of reference sources to a hundredth of an arc-second or better while at the same time determining baseline coordinates to the few-centimeter level. Once an accurate set of reference-source positions has been compiled, the estimation of additional baselines becomes vastly simplified, and the number of measurements needed to determine accurate baseline components is drastically reduced.

An area of continuing improvement is VLBI recording systems. Recent improvements have increased the recording

bandwidth to several megahertz, while state-of-the-art improvements promise to increase this by still another order of magnitude in the near future. Such wide-bandwidth recording systems will allow accurate VLBI measurements to be made with much smaller and less expensive antennas than are required today. As costs are reduced it should be possible to increase the number of usable antennas substantially, while at the same time providing a much increased coverage of geodetic and astrometric measurements.

NOTATION FOR APPENDICES

In Appendices A, B, C, and D we compute the mean variance of the estimates of the "true" cross-correlation function based on both analog and digital data techniques. Appendices A, B, and C deal with the analog case, the main results being given by Equations (B. 17) and (B. 21). Appendix D deals with the digital case, the main results being summarized by Equations (D. 17), (D. 18), (D. 22), and (D. 23). In order to lessen confusion between similar quantities in the analog and digital cases, we shall adopt a slightly different set of notation in Appendices A, B, C, and D than has been used in the text. The following quantities will be used frequently:

- $x(t), y(t)$ Two stationary, ergodic, band-limited, zero-mean random time functions with Gaussian statistics and a correlation coefficient $\rho(\tau)$, as defined by Equation (B. 1).
- $R_{xx}(\tau), R_{yy}(\tau)$ The true unnormalized auto-correlation function of $x(t)$ and $y(t)$, respectively, as defined by Equation (A. 7).
- $R_{xy}(\tau)$ The true unnormalized cross-correlation function of $x(t)$ and $y(t)$ as defined by Equation (A. 2).

$\rho(\tau)$	The true normalized cross-correlation function, as defined by Equation (B. 1)
$\hat{R}_{xx}(\tau), \hat{R}_{yy}(\tau), \hat{R}_{xy}(\tau), \hat{\rho}(\tau)$	Estimates of $R_{xx}(\tau), R_{yy}(\tau), R_{xy}(\tau)$, and $\rho(\tau)$, respectively, based upon analog samples of $x(t)$ and $y(t)$
$x'(t), y'(t)$	Infinitely-clipped versions of $x(t)$ and $y(t)$ $x'(t) = \begin{cases} +1 & \text{for } x(t) \geq 0 \\ -1 & \text{for } x(t) < 0 \end{cases}$ $y'(t) = \begin{cases} +1 & \text{for } y(t) \geq 0 \\ -1 & \text{for } y(t) < 0 \end{cases}$
$\rho'(\tau)$	The true normalized cross-correlation function of $x'(t)$ and $y'(t)$
$\hat{\rho}'(\tau)$	Estimate of $\rho'(\tau)$ based on samples of $x'(t)$ and $y'(t)$, as defined by Equation (D. 2)
$\hat{\rho}_c(\tau)$	Estimate of $\rho(\tau)$ based on $\hat{\rho}'(\tau)$, as defined by Equation (D. 3)

APPENDIX A

COVARIANCE ANALYSIS OF THE ANALOG CROSS-CORRELATION ESTIMATE

Define the following estimate of the unnormalized cross-correlation function

$$\hat{R}_{xy}(n\Delta\tau) \equiv \frac{1}{K} \sum_{k=1}^K x(k\Delta t)y(k\Delta t + n\Delta\tau) \quad (\text{A.1})$$

where $x(t)$ and $y(t)$ are band-limited, stationary zero-mean jointly Gaussian random variables with true cross-correlation

$$\begin{aligned} R_{xy}(n\Delta\tau) &\equiv \langle x(t)y(t + n\Delta\tau) \rangle \\ &= \lim_{K \rightarrow \infty} \frac{1}{2K} \sum_{k=-K}^K x(k\Delta t)y(k\Delta t + n\Delta\tau) \end{aligned} \quad (\text{A.2})$$

and where t is large enough for the successive samples of $x(t)$ and of $y(t)$ to be independent. We wish to examine the mean and variance of $\hat{R}_{xy}(n\Delta\tau)$. The mean is easily found to be

$$\begin{aligned} \langle \hat{R}_{xy}(n\Delta\tau) \rangle &= \frac{1}{K} \sum_{k=1}^K \langle x(k\Delta t)y(k\Delta t + n\Delta\tau) \rangle \\ &= R_{xy}(n\Delta\tau) \end{aligned} \quad (\text{A.3})$$

The covariance of $\hat{R}_{xy}(n\Delta\tau)$ is, by definition,

$$\begin{aligned} \sigma_{\hat{R}_{xy}}^2(n,m) &\equiv \langle [\hat{R}_{xy}(n\Delta\tau) - R_{xy}(n\Delta\tau)] [\hat{R}_{xy}(m\Delta\tau) \\ &\quad - R_{xy}(m\Delta\tau)] \rangle \end{aligned}$$

$$\begin{aligned}
 &= \left\langle \hat{R}_{xy}(n\Delta\tau) \hat{R}_{xy}(m\Delta\tau) \right\rangle - R_{xy}(n\Delta\tau) R_{xy}(m\Delta\tau) \\
 &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \left\langle x(k\Delta t) y(k\Delta t + n\Delta\tau) x(g\Delta t) y(g\Delta t + m\Delta\tau) \right\rangle \\
 &\quad - R_{xy}(n\Delta\tau) R_{xy}(m\Delta\tau) \tag{A.4}
 \end{aligned}$$

Using the expansion of the expectation of the product of four zero-mean Gaussian variables [see Equation (3-41)] we have

$$\begin{aligned}
 &\left\langle x(k\Delta t) y(k\Delta t + n\Delta\tau) x(g\Delta t) y(g\Delta t + m\Delta\tau) \right\rangle \\
 &= \left\langle x(k\Delta t) y(k\Delta t + n\Delta\tau) \right\rangle \left\langle x(g\Delta t) y(g\Delta t + m\Delta\tau) \right\rangle \\
 &+ \left\langle x(k\Delta t) x(g\Delta t) \right\rangle \left\langle y(k\Delta t + n\Delta\tau) y(g\Delta t + m\Delta\tau) \right\rangle \\
 &+ \left\langle x(k\Delta t) y(g\Delta t + m\Delta\tau) \right\rangle \left\langle x(g\Delta t) y(k\Delta t + n\Delta\tau) \right\rangle \tag{A.5}
 \end{aligned}$$

The first term of the right-hand side of Equation (A.5) is just $R_{xy}(n\Delta\tau) R_{xy}(m\Delta\tau)$, so that $\sigma_{\hat{R}_{xy}}^2$ of Equation (A.4) becomes

$$\begin{aligned}
 \sigma_{\hat{R}_{xy}}^2(n, m) &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \{ R_{xx}[(g-k)\Delta t] R_{yy}[(g-k)\Delta t \\
 &\quad + (m-n)\Delta\tau] + R_{xy}[(g-k)\Delta t + m\Delta\tau] R_{xy}[(k-g)\Delta t \\
 &\quad + n\Delta\tau] \} \tag{A.6}
 \end{aligned}$$

where we have used the property of stationarity of $x(t)$ and $y(t)$ to retain the same limits on the summations and where R_{xx} and R_{yy} are the auto-correlation functions of $x(t)$ and $y(t)$ defined by

$$\begin{aligned}
 R_{xx}(n\Delta\tau) &\equiv \langle x(t)x(t+n\Delta\tau) \rangle \\
 R_{yy}(n\Delta\tau) &\equiv \langle y(t)y(t+n\Delta\tau) \rangle
 \end{aligned}
 \tag{A.7}$$

Now, it is easy to show that

$$\sum_{k=1}^K \sum_{g=1}^K f_1(q-k) f_2(k-g) = \sum_{i=-K}^K (K-|i|) f_1(i) f_2(-i)
 \tag{A.8}$$

where f_1 and f_2 are any two functions, so that Equation (A.6) can be rewritten

$$\begin{aligned}
 \sigma_{\hat{R}_{xy}}^2(n,m) &= \frac{1}{K^2} \sum_{i=-K}^K (K-|i|) \{ R_{xx}(i\Delta t) R_{yy}[i\Delta t \\
 &\quad + (m-n)\Delta\tau] + R_{xy}(i\Delta t + m\Delta\tau) R_{xy}(-i\Delta t \\
 &\quad + n\Delta\tau) \}
 \end{aligned}
 \tag{A.9}$$

For the practical case of interest, $x(t)$ and $y(t)$ are bandlimited and $K \gg 1$ so that

$$R_{xx}(i\Delta t) = R_{yy}(i\Delta t) = R_{xy}(i\Delta t) \approx 0
 \tag{A.10}$$

for $i > i_{\max}$ where $i_{\max} \ll K$. Then, under the further assumption that $m\Delta\tau \ll i_{\max}\Delta t$ and $n\Delta\tau \ll i_{\max}\Delta t$, Equation (A.9) may be simplified to

$$\sigma_{\hat{R}_{xy}}^2(n,m) \approx \frac{1}{K} \sum_{i=-K}^K \{R_{xx}(i\Delta t)R_{yy}[i\Delta t + (m-n)\Delta\tau] + R_{xy}(i\Delta t + m\Delta\tau)R_{xy}(-i\Delta t + n\Delta\tau)\} \quad (\text{A.11})$$

Thus we have found the (unnormalized) cross-correlation covariance in terms of the auto-correlation and cross-correlation functions of $x(t)$ and $y(t)$.

It is instructive to examine Equation (A.11) for the limiting case where Δt is large enough so that

$$R_{xy}(i\Delta t) = R_{xx}(i\Delta t) = R_{yy}(i\Delta t) = 0 \quad \text{for } i \neq 0 \quad (\text{A.12})$$

This condition implies that the successive products $x(i\Delta t)y(i\Delta t + n\Delta\tau)$, $x(i\Delta t)x(i\Delta t + n\Delta\tau)$, and $y(i\Delta t)y(i\Delta t + n\Delta\tau)$ are linearly independent, e.g., that

$$x(i\Delta t)y(i\Delta t+n\Delta\tau) \cdot x(j\Delta t)y(j\Delta t+n\Delta\tau) = 0 \quad (\text{A.13})$$

for $i \neq j$

with similar expressions for the other products. The condition of Equation (A.13) is approximately valid for most continuum VLBI work. Under this condition, Equation (A.11), with $n = m$, reduces to

$$\sigma_{\hat{R}_{xy}}^2(n) \approx \frac{1}{K} [R_{xx}(0)R_{yy}(0) + R_{xy}^2(n\Delta\tau)] \quad (\text{A.14})$$

We note that if we set $x(t) = y(t)$, then $\hat{R}_{xy}(n\Delta\tau)$ becomes an auto-correlation and

$$R_{xx}(n\Delta t) = R_{yy}(n\Delta t) = R_{xy}(n\Delta t) \quad (\text{A.15})$$

Noting that the auto-correlation is necessarily an even function of $n\Delta t$, we can write the covariance of the auto-correlation estimate $\hat{R}_{xx}(n\Delta t)$ by inspection

$$\begin{aligned} \sigma_{\hat{R}_{xx}}^2(n, m) \approx & \frac{1}{K} \sum_{i=-K}^K \{ R_{xx}(i\Delta t) R_{xx}[i\Delta t + (m-n)\Delta t] \\ & + R_{xx}(i\Delta t + m\Delta t) R_{xx}(i\Delta t - n\Delta t) \} \end{aligned} \quad (\text{A.16})$$

For the condition of Equation (A.13) this reduces further to

$$\sigma_{\hat{R}_{xx}}^2(n) \approx \frac{1}{K} [R_{xx}^2(0) + R_{xx}^2(n\Delta t)] \quad (\text{A.17})$$

which is consistent with the results of Weinreb (Reference 31) for the variance of the auto-correlation estimate.

APPENDIX B

COVARIANCE ANALYSIS OF THE NORMALIZED ANALOG CROSS-CORRELATION ESTIMATE

Define the true normalized cross-correlation function to be

$$\rho(n\Delta\tau) \equiv \frac{R_{xy}(n\Delta\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}} \quad (\text{B.1})$$

where R_{xy} , R_{xx} , R_{yy} have been defined in Equations (A.2) and (A.7). Define an estimate of $\rho(n\Delta\tau)$, designated $\hat{\rho}(n\Delta\tau)$, to be

$$\hat{\rho}(n\Delta\tau) \equiv \frac{\hat{R}_{xy}(n\Delta\tau)}{\sqrt{\hat{R}_{xx}(0)\hat{R}_{yy}(0)}} \quad (\text{B.2})$$

where \hat{R}_{xy} is defined in Equation (A.1) and \hat{R}_{xx} , \hat{R}_{yy} are auto-correlation function estimates defined by

$$\hat{R}_{xx}(n\Delta\tau) \equiv \frac{1}{K} \sum_{k=1}^K x(k\Delta t)x(k\Delta t+n\Delta\tau) \quad (\text{B.3})$$

$$\hat{R}_{yy}(n\Delta\tau) \equiv \frac{1}{K} \sum_{k=1}^K y(k\Delta t)y(k\Delta t+n\Delta\tau) \quad (\text{B.4})$$

In this appendix we wish to examine the mean and variance of $\hat{\rho}(n\Delta\tau)$. To do so, we first define the fractional errors

$$\epsilon_{xy}(n) \equiv \frac{\hat{R}_{xy}(n\Delta\tau) - R_{xy}(n\Delta\tau)}{R_{xy}(0)} \quad (\text{B.5})$$

$$\epsilon_{xx}(n) \equiv \frac{\hat{R}_{xx}(n\Delta\tau) - R_{xx}(n\Delta\tau)}{R_{xx}(0)} \quad (\text{B.6})$$

$$\epsilon_{yy}(n) \equiv \frac{\hat{R}_{yy}(n\Delta\tau) - R_{yy}(n\Delta\tau)}{R_{yy}(0)} \quad (\text{B.7})$$

Clearly, from Equations (B.3), (B.4), and (A.3)

$$\begin{aligned} \langle \hat{R}_{xy}(n\Delta\tau) \rangle &= R_{xy}(n\Delta\tau) \\ \langle \hat{R}_{xx}(n\Delta\tau) \rangle &= R_{xx}(n\Delta\tau) \\ \langle \hat{R}_{yy}(n\Delta\tau) \rangle &= R_{yy}(n\Delta\tau) \end{aligned} \quad (\text{B.8})$$

so that

$$\langle \epsilon_{xy}(n) \rangle = \langle \epsilon_{xx}(n) \rangle = \langle \epsilon_{yy}(n) \rangle = 0 \quad (\text{B.9})$$

From Equations (B.5), (B.6), (B.7) and (A.4) we have the covariances

$$\begin{aligned} \langle \epsilon_{xy}(n) \epsilon_{xy}(m) \rangle &= \frac{\sigma_{\hat{R}_{xy}}^2(n,m)}{R_{xy}^2(0)} \\ \langle \epsilon_{xx}(n) \epsilon_{xx}(m) \rangle &= \frac{\sigma_{\hat{R}_{xx}}^2(n,m)}{R_{xx}^2(0)} \\ \langle \epsilon_{yy}(n) \epsilon_{yy}(m) \rangle &= \frac{\sigma_{\hat{R}_{yy}}^2(n,m)}{R_{yy}^2(0)} \end{aligned} \quad (\text{B.10})$$

where the covariances of \hat{R}_{xy} , \hat{R}_{xx} , and \hat{R}_{yy} are obtainable from Equations (A.11) and (A.16).

Using Equations (B.5), (B.6), (B.7), we can rewrite $\hat{\rho}(n\Delta\tau)$ of Equation (B.2) as

$$\hat{\rho}(n\Delta\tau) = \frac{R_{xy}(n\Delta\tau) + R_{xy}(0)\epsilon_{xy}(n)}{[R_{xx}(0) + R_{xx}(0)\epsilon_{xx}(0)]^{1/2} [R_{yy}(0) + R_{yy}(0)\epsilon_{yy}(0)]^{1/2}} \quad (\text{B.11})$$

Since ϵ_{xy} , ϵ_{xx} , ϵ_{yy} are all $\ll 1$, $\hat{\rho}(n\Delta\tau)$ can be expanded around $\epsilon_{xy} = \epsilon_{xx} = \epsilon_{yy} = 0$ to yield

$$\begin{aligned} \hat{\rho}(n\Delta\tau) &\approx \frac{1}{\sqrt{R_{xx}(0)R_{yy}(0)}} \{ [R_{xy}(n\Delta\tau) + R_{xy}(0)\epsilon_{xy}(n)] \\ &\cdot [1 - \frac{1}{2}\epsilon_{xx}(0)] [1 - \frac{1}{2}\epsilon_{yy}(0)] \} \\ &\approx [\rho(n\Delta\tau) + \rho(0)\epsilon_{xy}(n)] [1 - \frac{1}{2}\epsilon_{xx}(0)] [1 - \frac{1}{2}\epsilon_{yy}(0)] \\ &\approx \rho(n\Delta\tau) [1 - \frac{1}{2}\epsilon_{xx}(0) - \frac{1}{2}\epsilon_{yy}(0)] + \rho(0)\epsilon_{xy}(n) \end{aligned} \quad (\text{B.12})$$

where we have neglected all terms of order ϵ^2 . Using Equation (B.12) in the definition of the covariance of $\hat{\rho}(n\Delta\tau)$ we then have

$$\begin{aligned} \sigma_{\hat{\rho}}^2(n,m) &\equiv \langle [\hat{\rho}(n\Delta\tau) - \rho(n\Delta\tau)] [\hat{\rho}(m\Delta\tau) - \rho(m\Delta\tau)] \rangle \\ &= \langle [\rho(0)\epsilon_{xy}(n) - \frac{1}{2}\rho(n\Delta\tau)(\epsilon_{xx}(0) + \epsilon_{yy}(0))] \cdot \\ &\quad [\rho(0)\epsilon_{xy}(m) - \frac{1}{2}\rho(m\Delta\tau)(\epsilon_{xx}(0) + \epsilon_{yy}(0))] \rangle \quad (\text{B.13}) \end{aligned}$$

$$\begin{aligned}
 &= \rho^2(0) \left\langle \epsilon_{xy}(n) \epsilon_{xy}(m) \right\rangle - \frac{1}{2} \rho(0) \rho(m\Delta\tau) \left\langle \epsilon_{xx}(0) \epsilon_{xy}(n) \right. \\
 &\quad \left. + \epsilon_{yy}(0) \epsilon_{xy}(n) \right\rangle - \frac{1}{2} \rho(0) \rho(n\Delta\tau) \left\langle \epsilon_{xx}(0) \epsilon_{xy}(m) + \epsilon_{yy}(0) \epsilon_{xy}(m) \right\rangle \\
 &\quad + \frac{1}{4} \rho(n\Delta\tau) \rho(m\Delta\tau) \left\langle \epsilon_{xx}^2(0) + 2\epsilon_{xx}(0) \epsilon_{yy}(0) + \epsilon_{yy}^2(0) \right\rangle \tag{B.14}
 \end{aligned}$$

Equations (B.10), (A.11), and (C.7) can now be used to evaluate the last expression of Equation (B.14) in order to explicitly state $\sigma_{\hat{\rho}}^2(n,m)$ in terms of the cross-correlation and auto-correlation functions of $x(t)$ and $y(t)$. The resulting expression is rather complicated. For the practical cases of interest, however, we may make the assumption of linearly independent samples (as discussed below Equation (A.12)), and set $m = n$ to get simply the variance $\sigma_{\hat{\rho}}^2(n)$; using Equations (B.10), (A.14), (A.17), (C.9), and (C.17) we have

$$\begin{aligned}
 \left\langle \epsilon_{xy}^2(n) \right\rangle &= \frac{1}{KR_{xy}^2(0)} [R_{xx}(0)R_{yy}(0) + R_{xy}^2(n\Delta\tau)] \\
 \left\langle \epsilon_{xx}^2(0) \right\rangle &= \left\langle \epsilon_{yy}^2(0) \right\rangle = \frac{2}{K} \\
 \left\langle \epsilon_{xx}(0) \epsilon_{xy}(n) \right\rangle &= \left\langle \epsilon_{yy}(0) \epsilon_{xy}(n) \right\rangle = \frac{2}{K} \frac{R_{xy}(n\Delta\tau)}{R_{xy}(0)} \\
 \left\langle \epsilon_{xx}(0) \epsilon_{yy}(0) \right\rangle &= \frac{2}{K} \frac{R_{xy}^2(0)}{R_{xx}(0)R_{yy}(0)}
 \end{aligned} \tag{B.15}$$

Substituting these into Equation (B.14), with $m = n$, we have the result

$$\begin{aligned} \sigma_{\hat{\rho}}^2(n) = & \frac{\rho^2(0)}{K} \left\{ \frac{R_{xx}(0)R_{yy}(0)}{R_{xy}^2(0)} + \frac{R_{xy}^2(n\Delta\tau)}{R_{xy}^2(0)} \right\} \\ & - \frac{4\rho(0)\rho(n\Delta\tau)}{K} \frac{R_{xy}(n\Delta\tau)}{R_{xy}(0)} \\ & + \frac{\rho^2(n\Delta\tau)}{K} \left\{ 1 + \frac{R_{xy}^2(0)}{R_{xx}(0)R_{yy}(0)} \right\} \end{aligned} \quad (\text{B.16})$$

Using the definition of $\rho(n\Delta\tau)$ in Equation (B.1), the above immediately reduces to

$$\sigma_{\hat{\rho}}^2(n) = \frac{1}{K} \{1 - \rho^2(n\Delta\tau) [2 - \rho^2(0)]\} \quad (\text{B.17})$$

which is the result we are seeking.

It is illuminating to examine Equation (B.17) for cases of both large and small correlation. For the case of auto-correlation (i.e. $\rho(0) = 1$) Equation (B.17) reduces to

$$\sigma_{\hat{\rho}}^2(n) = \frac{1}{K} [1 - \rho^2(n\Delta\tau)] \quad (\text{B.18})$$

which is identical to the result given by Weinreb (Reference 31) for auto-correlation. For the case of small correlation (i.e., $\rho(0) \ll 1$), Equation (B.17) reduces to

$$\sigma_{\hat{\rho}}^2(n) \approx \frac{1}{K} [1 - 2\rho^2(n\Delta\tau)] \quad (\text{B.19})$$

Finally, for the case of no correlation, Equation (B.17) reduces simply to

$$\sigma_{\hat{\rho}}^2(n) = \frac{1}{K} \quad (\text{B.20})$$

The mean of $\hat{\rho}(n\Delta\tau)$ follows directly from Equation (B.12) and (B.9),

$$\langle \hat{\rho}(n\Delta\tau) \rangle = \rho(n\Delta\tau) \quad (\text{B.21})$$

showing that the estimate $\hat{\rho}(n\Delta\tau)$, at least to the order of ϵ , is an unbiased estimate.

APPENDIX C

CROSS-COVARIANCE ANALYSIS OF THE ANALOG CROSS-CORRELATION

AND

AUTO-CORRELATION ESTIMATES

In Appendix B we derived the covariance of the normalized analog cross-correlation estimate. During the course of that derivation it was necessary to know the cross-covariances of the estimate "errors" $\epsilon_{xy}(n)$, $\epsilon_{xx}(n)$, $\epsilon_{yy}(n)$. These quantities are defined in Equations (B.5), (B.6) and (B.7). In this appendix we shall derive the necessary cross-covariances and, as a side result, the cross-covariance of the analog cross-correlation and auto-correlation estimates. From Equations (B.5) and (B.6) we can write directly

$$\begin{aligned} \langle \epsilon_{xx}(n) \epsilon_{xy}(m) \rangle &= \frac{1}{R_{xx}(0)R_{xy}(0)} \langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{xy}(m\Delta\tau) \\ &\quad - R_{xx}(n\Delta\tau)R_{xy}(m\Delta\tau) \rangle \end{aligned} \quad (C.1)$$

where $\hat{R}_{xy}(n\Delta\tau)$, $R_{xy}(n\Delta\tau)$, $R_{xx}(n\Delta\tau)$, $\hat{R}_{xx}(n\Delta\tau)$ have been defined by Equations (A.1), (A.2), (A.7), (B.3), respectively. From Equations (A.1) and (A.7) we can evaluate the cross-covariance

$$\begin{aligned} \langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{xy}(m\Delta\tau) \rangle &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \langle x(k\Delta t)x(k\Delta t+n\Delta\tau)x(g\Delta t)y(g\Delta t \\ &\quad +m\Delta\tau) \rangle \end{aligned} \quad (C.2)$$

which, using the expansion of the expectation of the product of four zero-mean Gaussian variables [See Equation (3-41)], becomes

$$\begin{aligned}
 \left\langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{xy}(m\Delta\tau) \right\rangle &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \left\{ \left\langle x(k\Delta t) x(k\Delta t \right. \right. \\
 &\quad \left. \left. + n\Delta\tau) \right\rangle \left\langle x(g\Delta t) y(g\Delta t \right. \right. \\
 &\quad \left. \left. + m\Delta\tau) \right\rangle + \left\langle x(k\Delta t) x(g\Delta t) \right\rangle \right. \\
 &\quad \cdot \left\langle x(k\Delta t + n\Delta\tau) y(g\Delta t + m\Delta\tau) \right\rangle \\
 &\quad \left. + \left\langle x(k\Delta t) y(g\Delta t + m\Delta\tau) \right\rangle \left\langle x(k\Delta t \right. \right. \\
 &\quad \left. \left. + n\Delta\tau) x(g\Delta t) \right\rangle \right\} \\
 &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \{ R_{xx}(n\Delta\tau) R_{xy}(m\Delta\tau) \\
 &\quad + R_{xx}[(g-k)\Delta t] R_{xy}[(g-k)\Delta t \\
 &\quad + (m-n)\Delta\tau] + R_{xy}[(g-k)\Delta t \\
 &\quad + m\Delta\tau] R_{xx}[(g-k)\Delta t - n\Delta\tau] \}
 \end{aligned} \tag{C.3}$$

Noting the first term of the last expression is independent of k and g , and using the property of Equation (A.8), Equation (C.3) becomes

$$\begin{aligned}
 \langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{xy}(m\Delta\tau) \rangle &= R_{xx}(n\Delta\tau) R_{xy}(m\Delta\tau) \\
 &+ \frac{1}{K^2} \sum_{i=-K}^K (K-|i|) \{ R_{xx}(i\Delta t) R_{xy}(i\Delta t \\
 &+ (m-n)\Delta\tau) + R_{xy}(i\Delta t+m\Delta\tau) R_{xx}(i\Delta t-n\Delta\tau) \}
 \end{aligned} \tag{C.4}$$

Proceeding as in Equation (A.10), we assume

$$R_{xx}(i\Delta t) = R_{xy}(i\Delta t) \approx 0 \tag{C.5}$$

for $i > i_{\max}$ where $i_{\max} \ll K$. Then, under the further assumption that $n\Delta\tau \ll i_{\max}\Delta t$, Equation (C.4) may be written

$$\begin{aligned}
 \langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{xy}(m\Delta\tau) \rangle &= R_{xx}(n\Delta\tau) R_{xy}(m\Delta\tau) \\
 &+ \frac{1}{K} \sum_{i=-K}^K \{ R_{xx}(i\Delta t) R_{xy}[i\Delta t \\
 &+ (m-n)\Delta\tau] + R_{xy}(i\Delta t+m\Delta\tau) R_{xx}(i\Delta t \\
 &-n\Delta\tau) \}
 \end{aligned} \tag{C.6}$$

Using Equation (C.6) in Equation (C.1) yields

$$\begin{aligned}
 \langle \epsilon_{xx}(n) \epsilon_{xy}(m) \rangle &= \frac{1}{R_{xx}(0) R_{xy}(0)} \cdot \frac{1}{K} \sum_{i=-K}^K \{ R_{xx}(i\Delta t) R_{xy}[i\Delta t \\
 &+ (m-n)\Delta\tau] + R_{xy}(i\Delta t+m\Delta\tau) R_{xx}(i\Delta t \\
 &-n\Delta\tau) \}
 \end{aligned} \tag{C.7}$$

For the practical case of interest we assume Δt is

large enough so that the conditions of Equation (A.12) apply and (C.7) becomes

$$\begin{aligned} \langle \epsilon_{xx}(n) \epsilon_{xy}(m) \rangle \approx & \frac{1}{KR_{xx}(0)R_{xy}(0)} \{R_{xx}(0)R_{xy}[(m-n)\Delta\tau] \\ & + R_{xy}(m\Delta\tau)R_{xx}(-n\Delta\tau)\} \end{aligned} \quad (C.8)$$

This result, and the equivalent result for $\langle \epsilon_{yy}(n) \epsilon_{xy}(m) \rangle$, are required in order to evaluate the expression for the covariance of the normalized cross-correlation estimate as given in Equation (B.12). In particular, the value of $\langle \epsilon_{xx}(0) \epsilon_{xy}(m) \rangle$ is required, which follows immediately from Equation (C.8),

$$\langle \epsilon_{xx}(0) \epsilon_{xy}(m) \rangle = \frac{2}{K} \frac{R_{xy}(m\Delta\tau)}{R_{xy}(0)} \quad (C.9)$$

The expression for $\langle \epsilon_{yy}(0) \epsilon_{xy}(m) \rangle$ is identical.

Finally, we must evaluate the quantity

$\langle \epsilon_{xx}(n) \epsilon_{yy}(m) \rangle$, which follows in a manner exactly analogous to the above. From Equations (B.6) and (B.7) we can write directly

$$\begin{aligned} \langle \epsilon_{xx}(n) \epsilon_{yy}(m) \rangle = & \frac{1}{R_{xx}(0)R_{yy}(0)} \{ \langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{yy}(m\Delta\tau) \rangle \\ & - R_{xx}(n\Delta\tau)R_{yy}(m\Delta\tau) \} \end{aligned} \quad (C.10)$$

From the definitions of \hat{R}_{xx} and \hat{R}_{yy} in Equations (B.3) and (B.4) we can write

$$\begin{aligned} \left\langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{yy}(m\Delta\tau) \right\rangle &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \left\langle x(k\Delta t)x(k\Delta t \right. \\ &\quad \left. +n\Delta\tau)y(g\Delta t)y(g\Delta t+m\Delta\tau) \right\rangle \end{aligned} \quad (C.11)$$

which, by the expansion of the product of Gaussian variables, becomes

$$\begin{aligned} \left\langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{yy}(m\Delta\tau) \right\rangle &= \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \left\{ \left\langle x(k\Delta t)x(k\Delta t \right. \right. \\ &\quad \left. \left. + n\Delta\tau) \right\rangle \left\langle y(g\Delta t)y(g\Delta t \right. \right. \\ &\quad \left. \left. + m\Delta\tau) \right\rangle + \left\langle x(k\Delta t)y(g\Delta t) \right\rangle \right. \\ &\quad \left. \left\langle x(k\Delta t+n\Delta\tau)y(g\Delta t+m\Delta\tau) \right\rangle \right. \\ &\quad \left. + \left\langle x(k\Delta t)y(g\Delta t+m\Delta\tau) \right\rangle \left\langle x(k\Delta t \right. \right. \\ &\quad \left. \left. + n\Delta\tau)y(g\Delta t) \right\rangle \right\} \\ &= R_{xx}(n\Delta\tau)R_{yy}(m\Delta\tau) \\ &+ \frac{1}{K^2} \sum_{k=1}^K \sum_{g=1}^K \left\{ R_{xy}[(g-k)\Delta t]R_{xy}[(g-k)\Delta t+(m-n)\Delta\tau] \right. \\ &\quad \left. + R_{xy}[(g-k)\Delta t+m\Delta\tau]R_{xy}[(g-k)\Delta t-n\Delta\tau] \right\} \end{aligned} \quad (C.12)$$

Applying the property of Equation (A.8), Equation (C.12) becomes

$$\left\langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{yy}(m\Delta\tau) \right\rangle = R_{xx}(n\Delta\tau)R_{yy}(m\Delta\tau)$$

$$\begin{aligned}
 & + \frac{1}{K^2} \sum_{i=-K}^K (K-|i|) \{ R_{xy}(i\Delta t) R_{xy}[(i\Delta t+(m-n)\Delta\tau] \\
 & + R_{xy}(i\Delta t+m\Delta\tau) R_{xy}(i\Delta t-n\Delta\tau) \} \quad (C.13)
 \end{aligned}$$

Proceeding as near Equation (C.15), Equation (C.13) further reduces, in the practical case, to

$$\begin{aligned}
 \langle \hat{R}_{xx}(n\Delta\tau) \hat{R}_{yy}(m\Delta\tau) \rangle & = R_{xx}(n\Delta\tau) R_{yy}(m\Delta\tau) \\
 & + \frac{1}{K} \sum_{i=-K}^K \{ R_{xy}(i\Delta t) R_{xy}[i\Delta t+(m-n)\Delta\tau] \\
 & + R_{xy}(i\Delta t+m\Delta\tau) R_{xy}(i\Delta t-n\Delta\tau) \} \quad (C.14)
 \end{aligned}$$

Substituting this into Equation (C.10) we have

$$\begin{aligned}
 \langle \epsilon_{xx}(n) \epsilon_{yy}(m) \rangle & = \frac{1}{R_{xx}(0) R_{yy}(0)} \cdot \frac{1}{K} \cdot \\
 & \sum_{i=-K}^K \{ R_{xy}(i\Delta t) R_{xy}[i\Delta t+(m-n)\Delta\tau] + R_{xy}(i\Delta t+m\Delta\tau) R_{xy}(i\Delta t \\
 & -n\Delta\tau) \} \quad (C.15)
 \end{aligned}$$

Assuming that Δt is large enough so that the conditions of Equation (A.12) are met, we obtain

$$\begin{aligned}
 \langle \epsilon_{xx}(n) \epsilon_{yy}(m) \rangle & = \frac{1}{KR_{xx}(0) R_{yy}(0)} \{ R_{xy}(0) R_{xy}[(m-n)\Delta\tau] \\
 & + R_{xy}(m\Delta\tau) R_{xy}(-n\Delta\tau) \} \quad (C.16)
 \end{aligned}$$

The value of $\langle \epsilon_{xx}(0)\epsilon_{yy}(0) \rangle$ is needed in order to evaluate the expressions in Equation (B.14). From Equation (C.16),

$$\langle \epsilon_{xx}(0)\epsilon_{yy}(0) \rangle = \frac{2}{K} \frac{R_{xy}^2(0)}{R_{xx}(0)R_{yy}(0)} \quad (C.17)$$

The result of Equation (C.17) is used in Appendix B to assist in the computation of the covariance of the normalized analog cross-correlation estimate.

APPENDIX D

MEAN AND VARIANCE OF THE NORMALIZED CROSS-CORRELATION ESTIMATE
FROM AN INFINITELY-CLIPPED SIGNAL

The method used here to compute the mean and variance of the normalized cross-correlation estimate is similar to that used by Weinreb (Reference 31) to compute similar quantities for the normalized auto-correlation estimate.

In Section 3.5 we derived the so-called "Van Vleck clipping correction" which indicates the relationship between the "true" cross-correlation function $\rho(\tau)$ and the one-bit-sample (so-called "clipped") cross-correlation function $\rho'(\tau)$ [see Equation (3-105)].

$$\rho(\tau) = \sin\left[\frac{\pi}{2}\rho'(\tau)\right] \quad (\text{D. 1})$$

In general only a finite number of samples of the clipped waveforms $x'(t)$ and $y'(t)$ are available from which an estimate of $\rho'(\tau)$ may be made. We shall define this estimate, designated $\hat{\rho}'(\tau)$, to be

$$\hat{\rho}'(\tau) \equiv \frac{1}{K} \sum_{m=1}^K x'(m\Delta t)y'(m\Delta t+\tau) \quad (\text{D. 2})$$

where Δt is the sampling interval chosen to make the samples independent (see Appendix A for more details). We now wish to estimate $\rho(\tau)$, the true cross-correlation function, based on $\hat{\rho}'(\tau)$. Designate this estimate to

be $\hat{\rho}_c(\tau)$. One possible choice of an estimator follows from the relation of Equation (D.1), from which we are led to define $\hat{\rho}_c(\tau)$ as

$$\hat{\rho}_c(\tau) \equiv \sin\left[\frac{\pi}{2}\hat{\rho}'(\tau)\right] \quad (\text{D.3})$$

We now wish to examine the statistics of $\hat{\rho}_c(\tau)$ in detail. Since Equation (D.3) is a non-linear estimator (and hence may be a biased estimator) we must determine the mean as well as the variance. Clearly, from Equation (D.2), the mean of $\hat{\rho}'(\tau)$ is just $\rho'(\tau)$; the mean of $\hat{\rho}_c(\tau)$ is not so easily found, however, The mean and mean-square (and hence the variance) of $\hat{\rho}_c(\tau)$ can be expressed in terms of $p[\hat{\rho}'(\tau)]$, the probability density function of $\hat{\rho}'(\tau)$, (dropping the explicit dependence on τ for convenience), as

$$\langle \hat{\rho}_c \rangle = \int_{-\infty}^{\infty} \sin\left(\frac{\pi}{2}\hat{\rho}'\right) p(\hat{\rho}') d\hat{\rho}' \quad (\text{D.4})$$

$$\langle \hat{\rho}_c^2 \rangle = \int_{-\infty}^{\infty} \sin^2\left(\frac{\pi}{2}\hat{\rho}'\right) p(\hat{\rho}') d\hat{\rho}' \quad (\text{D.5})$$

In order to evaluate $p(\hat{\rho}')$ we may use its characteristic function $M_{\hat{\rho}'}(\nu)$, which is the Fourier transform of $p(\hat{\rho}')$ [see, for example, Reference 28]

$$p(\hat{\rho}') = \frac{1}{2\pi} \int_{-\infty}^{\infty} M_{\hat{\rho}'}(\nu) e^{-i\hat{\rho}'\nu} d\nu \quad (\text{D.6})$$

Substituting this into Equation (D.4) we have

$$\begin{aligned}
 \langle \hat{\rho}_c \rangle &= \int_{-\infty}^{\infty} \frac{e^{\frac{i\pi\hat{\rho}'}{2}} - e^{-\frac{i\pi\hat{\rho}'}{2}}}{2i} \int_{-\infty}^{\infty} M_{\hat{\rho}'}(v) e^{-i\hat{\rho}'v} dv d\hat{\rho}' \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M_{\hat{\rho}'}(v) \left[\frac{e^{i\hat{\rho}'(\frac{\pi}{2} - v)} - e^{-i\hat{\rho}'(\frac{\pi}{2} + v)}}{2i} \right] dv d\hat{\rho}' \\
 &= \frac{1}{2i} \int_{-\infty}^{\infty} M_{\hat{\rho}'}(v) [\delta(v - \frac{\pi}{2}) - \delta(v + \frac{\pi}{2})] dv \quad (D.7)
 \end{aligned}$$

where we have used the property

$$\int_{-\infty}^{\infty} e^{i\alpha(\beta - \beta')} d\alpha = 2\pi\delta(\beta - \beta')$$

Therefore

$$\langle \rho_c \rangle = \frac{1}{2i} [M_{\hat{\rho}'}(\frac{\pi}{2}) - M_{\hat{\rho}'}(-\frac{\pi}{2})] \quad (D.8)$$

since

$$\int_{-\infty}^{\infty} f(x)\delta(ax)dx = \frac{1}{a}f(0)$$

Similarly, it is easy to show

$$\langle \hat{\rho}'^2 \rangle = \frac{1}{2}M_{\hat{\rho}'}(0) - \frac{1}{4}M_{\hat{\rho}'}(\pi) - \frac{1}{4}M_{\hat{\rho}'}(-\pi) \quad (D.9)$$

We must now find $M_{\hat{\rho}'}(v)$. From Equation (D.2), we see that $\hat{\rho}'$ is just the sum of K independent random variables.

Therefore we make use of the property that the characteristic function of the sum of statistically independent random variables is just the product of the character-

istic functions of the individual terms (Reference 28) :

$$M_{\hat{\rho}}(v) = \prod_{m=1}^K M_m(v) \quad (D.10)$$

where $M_m(v)$ is the characteristic function for the term $\frac{1}{K}x'(m\Delta t)y'(m\Delta t+\tau)$. Each such term may assume either the value of $\frac{1}{K}$ with probability $2P_{++}$, or the value $-\frac{1}{K}$ with probability $2P_{+-}$ (where we have used the fact that $P_{++} = P_{--}$ and $P_{-+} = P_{+-}$, implicit from the probability density of Equation (3-90) , so that $M_m(v)$ becomes

$$M_m(v) = 2P_{++}e^{i\frac{v}{K}} + 2P_{+-}e^{-i\frac{v}{K}} \quad (D.11)$$

Using Equations (3-99), (3-100), and (3-105) of the text we have

$$P_{++} = \frac{1}{4}[1+\rho'] = \frac{1}{4} [1 + \frac{2}{\pi} \sin^{-1}\rho] \quad (D.12)$$

$$P_{+-} = \frac{1}{4}[1 - \rho'] = \frac{1}{4} [1 - \frac{2}{\pi} \sin^{-1}\rho] \quad (D.13)$$

Upon substitution of Equations (D.12) and (D.13) into Equation (D.11) and thence into Equation (D.10), we obtain an expression for $M_{\hat{\rho}}(v)$ which can be substituted into Equations (D.8) and (D.9) to yield

$$\begin{aligned} \langle \hat{\rho}_c(\tau) \rangle &= \frac{1}{2j} [(\cos \frac{\pi}{2K} + j\frac{2}{\pi} \sin^{-1}\rho(\tau) \sin \frac{\pi}{2K})^K \\ &\quad - (\cos \frac{\pi}{2K} - j\frac{2}{\pi} \sin^{-1}\rho(\tau) \sin \frac{\pi}{2K})^K] \quad (D.14) \end{aligned}$$

$$\begin{aligned} \langle \hat{\rho}_c^2(\tau) \rangle &= \frac{1}{2} - \frac{1}{4} [(\cos \frac{\pi}{K} + j \frac{2}{\pi} \sin^{-1} \rho(\tau) \sin \frac{\pi}{K})^K \\ &\quad + (\cos \frac{\pi}{K} - j \frac{2}{\pi} \sin^{-1} \rho(\tau) \sin \frac{\pi}{K})^K] \end{aligned} \quad (D.15)$$

These expressions, then, represent the mean and mean square of $\hat{\rho}_c(\tau)$, the estimate of $\rho(\tau)$ based on a finite number of infinitely-clipped samples of $x(t)$ and $y(t)$, in terms of the true cross-correlation $\rho(\tau)$. We note that no approximations whatever have been used in their derivation except that the samples of $x(t)$ and $y(t)$ are assumed to be independent.

For most cases of interest, K is a very large number compared to π , in which case the following approximations may be made

$$\left. \begin{aligned} \cos \frac{a}{K} &\sim 1 - \frac{a^2}{2K^2} \\ \sin \frac{a}{K} &\sim \frac{a}{K} \\ (1 + \frac{a}{K} + \frac{b}{K^2})^K &\sim e^a (1 - \frac{a^2}{2K} + \frac{b}{K}) \end{aligned} \right\} \quad (D.16)$$

The last approximation may be seen to be appropriate by taking the logarithm of both sides of the equation and then expanding the logarithm. Substituting these approximations into Equations (D.14) and (D.15) and using

$$\sigma_{\hat{\rho}_c}^2 = \langle \hat{\rho}_c^2(\tau) \rangle - \langle \hat{\rho}_c(\tau) \rangle^2 \quad \text{we readily find}$$

$$\langle \hat{\rho}_c(\tau) \rangle \approx \rho(\tau) \left[1 - \frac{\pi^2}{8K} (1 - \rho'^2(\tau)) \right] \quad (\text{D.17})$$

$$\sigma_{\hat{\rho}_c}^2 \approx \frac{\pi^2}{4K} [1 - \rho(\tau)^2] [1 - \rho'^2(\tau)] \quad (\text{D.18})$$

where $\rho'(\tau) = \frac{2}{\pi} \sin^{-1} \rho(\tau)$ as given in Equation (D.1). From Equation (D.17) we see that $\langle \hat{\rho}_c(\tau) \rangle$ is biased by the amount $\rho(\tau) \frac{\pi^2}{8K} \left[\left(\frac{2}{\pi} \sin^{-1} \rho(\tau) \right)^2 - 1 \right]$ from the desired value $\rho(\tau)$. For large K , this bias will be on the order of \sqrt{K} times smaller than $\sigma_{\hat{\rho}_c}$ and hence is usually neglected. In practice, the effective value of K depends largely on the exact algorithms used for data processing. For the VLBI 1 program (see Section 4.3), for example, the effective value of K is 16, while for some earlier algorithms it is much, much larger (see, for example, Moran, Reference 6).

Equations (D.14), (D.15), (D.17) and (D.18) have been numerically computed for several values of K . Figure D-1 shows the typical behavior of Equations (D.14) and (D.17) for several values of K . Note that, for $\rho \lesssim 0.1$, the quantity $\langle \hat{\rho}_c \rangle / \rho$ is nearly constant. Table D-1 lists, for many typical values of K , the values $\langle \hat{\rho}_c \rangle / \rho$ and $\sigma_{\hat{\rho}_c}$ for $\rho \lesssim 0.1$.

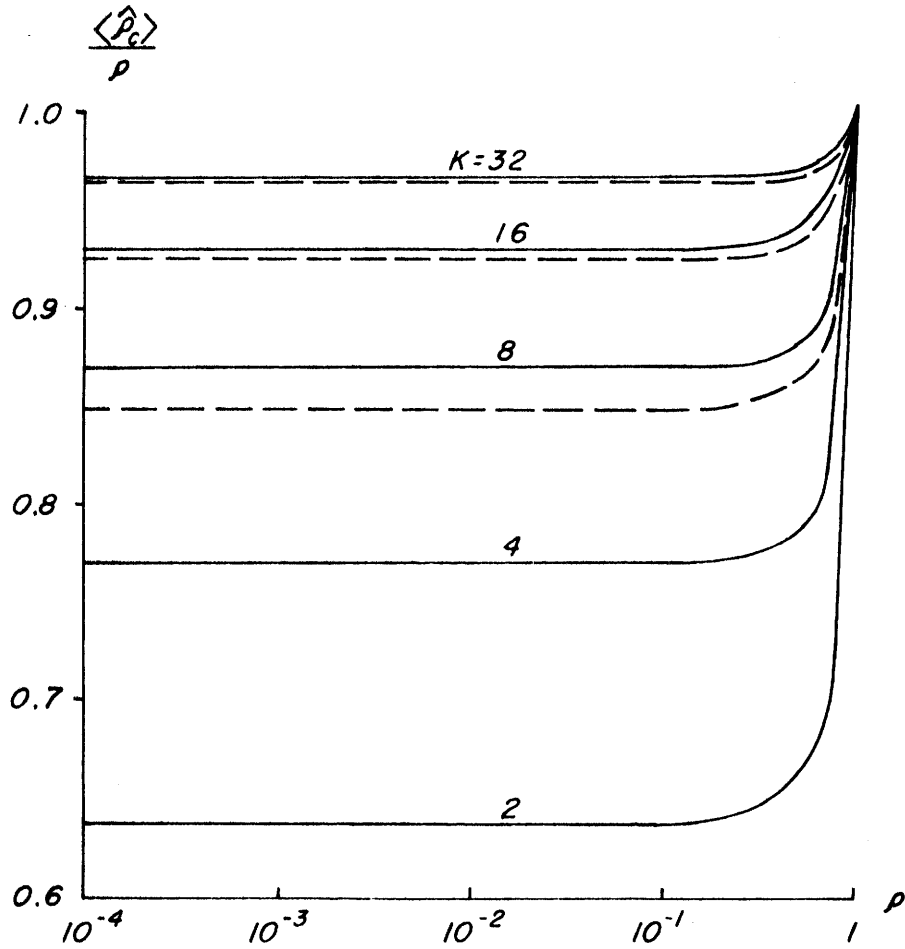


Figure D-1

Typical behavior of $\frac{\langle \hat{p}_c \rangle}{\rho}$ for several values of K as computed from Equation (D.14) [solid lines] and its approximation for large K Equation (D.17) [dashed lines].

K	$\frac{\langle \hat{\rho}_c \rangle}{\rho}$	$\sigma_{\hat{\rho}_c}$
2	0.637	0.707
4	0.769	0.612
8	0.867	0.484
16	0.929	0.365
32	0.963	0.267
64	0.981	0.193
128	0.990	0.137
256	0.995	0.098
512	0.998	0.069
1024	0.999	0.049

Table D-1

Table of $\langle \hat{\rho}_c \rangle / \rho$ and $\sigma_{\hat{\rho}_c}$ as computed using Equations (D.14) and (D.15). The values given are accurate to $\sim \pm 0.002$ over the range $0 < \rho \leq 0.1$. As ρ approaches 1, $\langle \hat{\rho}_c \rangle / \rho \rightarrow 1$ and $\sigma_{\hat{\rho}_c} \rightarrow 0$.

Figure D-1

Bias and Variance of Estimate $\hat{\rho}_c(\tau)$

An Alternate Method

An alternate, somewhat more straightforward approach may also be taken in arriving at the mean and variance of $\hat{\rho}_c(\tau)$. We may consider $\hat{\rho}'(\tau)$ of Equation (D.2) to be a one-dimensional random walk with K steps, each step being described by a random variable s with probabilities

$$\Pr[s = +1] = 2P_{++} \tag{D. 19}$$

$$\Pr[s = -1] = 2P_{-+}$$

which are just the probabilities of matching and non-matching bits, respectively, as discussed immediately before Equation (D.11) and where P_{++} and P_{-+} are related to ρ and ρ' by Equations (D.12) and (D.13). We may then write $\hat{\rho}'(\tau)$ in terms of s as

$$\hat{\rho}'(\tau) = \frac{1}{K} \sum_{m=1}^K s_m \tag{D.20}$$

Note that $\hat{\rho}'(\tau)$ can take on only K discrete values in the range $-1 \leq \hat{\rho}'(\tau) \leq 1$. The probability that $\hat{\rho}'(\tau) = \frac{q}{K}$ is the probability that $(K+q)/2$ bit pairs are matching and $(K-q)/2$ bit pairs are non-matching. But this is just the familiar binomial probability distribution for $(K+q)/2$ "successes" (matching bit pairs) and $(K-q)/2$ "failures" (non-matching bit pairs) in K trials.

(See, for example, Papoulis, Reference 28.) Hence

$$\Pr[\hat{\rho}'_c(\tau) = \frac{q}{K}] = \binom{K}{\frac{K+q}{2}} (2P_{++})^{\frac{K+q}{2}} (2P_{+-})^{\frac{K-q}{2}} \tag{D.21}$$

where q may take on the discrete values $-K, -K+2, \dots, 0, \dots, K-2, K$. Then the mean and mean square of $\hat{\rho}_c(\tau)$, from Equations (D.1) and (D.21), are

$$\begin{aligned} \langle \hat{\rho}_c(\tau) \rangle &= \sum_q \sin\left(\frac{\pi}{2} \cdot \frac{q}{K}\right) \Pr[\hat{\rho}'_c(\tau) = \frac{q}{K}] \\ &= \sum_q \sin\left(\frac{\pi}{2} \cdot \frac{q}{K}\right) \binom{K}{\frac{K+q}{2}} (2P_{++})^{\frac{K+q}{2}} (2P_{+-})^{\frac{K-q}{2}} \quad (D.22) \end{aligned}$$

$$\begin{aligned} \langle \hat{\rho}_c^2(\tau) \rangle &= \sum_q \sin^2\left(\frac{\pi}{2} \cdot \frac{q}{K}\right) \Pr[\hat{\rho}'_c(\tau) = \frac{q}{K}] \\ &= \sum_q \sin^2\left(\frac{\pi}{2} \cdot \frac{q}{K}\right) \binom{K}{\frac{K+q}{2}} (2P_{++})^{\frac{K+q}{2}} (2P_{+-})^{\frac{K-q}{2}} \quad (D.23) \end{aligned}$$

These expressions may be evaluated numerically to yield the same results as given in Table D-1. We note, incidentally, that the characteristic of the distribution of (D.21) is

$$\begin{aligned} M_{\hat{\rho}'_c}(v) &= \sum_q \Pr(\hat{\rho}'_c = \frac{q}{K}) e^{+iv(\frac{q}{K})} \\ &= \sum_q e^{i\frac{qv}{K}} \binom{K}{\frac{K+q}{2}} (2P_{++})^{\frac{K+q}{2}} (2P_{+-})^{\frac{K-q}{2}} \\ &= \left(2P_{++} e^{\frac{iv}{K}} + 2P_{+-} e^{\frac{-iv}{K}} \right)^K \quad (D.24) \end{aligned}$$

which is, as we should expect, identical to $M_{\hat{\rho}_c}(v)$ as given by Equations (D.10) and (D.11).

APPENDIX E

REVIEW OF LEAST-SQUARES FIT OF DATA TO A STRAIGHT LINE

Suppose that we have a set of data, designated Y which consists of N independent pairs of measurements (x_i, y_i) of an independent noiseless variable x and a dependent noisy variable y. We wish to fit the data with an equation of the form

$$\hat{y} = \hat{a} + \hat{b}x \quad (E.1)$$

where we must determine the coefficients \hat{a} and \hat{b} .

Assume that the data points have been derived from a parent distribution which determines the probability of making any particular observation, and that the actual relationship between y and x is given by

$$y(x) = a + bx \quad (E.2)$$

We cannot determine the coefficients a and b exactly from a finite number of observations, but we want to extract from the data the most-likely value for a and b, where we shall define "most-likely" below. Assume further that the parent distribution is Gaussian, so that for any given value of $x = x_i$ the probability P_i for making the observed measurement y_i is

$$P_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - Y(x_i)}{\sigma_i}\right]^2\right\} \quad (E.3)$$

The probability $P(Y)$ of making the observed set of measurements is then

$$P(Y) = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[\frac{Y_i - Y(x_i)}{\sigma_i} \right]^2 \right\} \quad (\text{E.4})$$

where we have allowed each individual measurement to have its own variance (i.e. weight), σ_i .

We now wish to find the "maximum likelihood" (or "most-likely") estimates of a and b, which are defined as the values of \hat{a} and \hat{b} which maximize the conditional probability

$$P(Y|\hat{a}, \hat{b}) = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[\frac{Y_i - \hat{Y}_i}{\sigma_i} \right]^2 \right\} \quad (\text{E.5})$$

which is equivalent to maximizing the logarithm

$$\ln P(Y|\hat{a}, \hat{b}) = \sum_i \ln \left[\frac{1}{\sigma_i \sqrt{2\pi}} \right] - \frac{1}{2} \sum_i \left[\frac{Y_i - \hat{Y}_i}{\sigma_i} \right]^2 \quad (\text{E.5})$$

where the summations are understood to run from $i=1$ to $i=N$. The first term is independent of a and b, so maximizing Equation (E.5) is equivalent to minimizing the quantity

$$\begin{aligned} \epsilon^2 &\equiv \sum_i \left[\frac{Y_i - \hat{Y}_i}{\sigma_i} \right]^2 \\ &= \sum_i \left[\frac{1}{\sigma_i} (Y_i - \hat{a} - \hat{b}x_i) \right]^2 \end{aligned} \quad (\text{E.6})$$

which is just the least-squares criterion. Therefore, for a Gaussianly distributed variable, we see that the maximum-likelihood estimates and least-squares estimates are identical.

In order to solve for \hat{a} and \hat{b} we take the partial derivatives of ϵ^2 and set them to zero

$$\frac{\partial \epsilon^2}{\partial \hat{a}} = -2 \sum_i \frac{1}{\sigma_i^2} (y_i - \hat{a} - \hat{b}x_i) = 0 \quad (\text{E.7})$$

$$\frac{\partial \epsilon^2}{\partial \hat{b}} = -2 \sum_i \left[\frac{x_i}{\sigma_i^2} (y_i - \hat{a} - \hat{b}x_i) \right] = 0 \quad (\text{E.8})$$

These equations can be rearranged to yield a pair of simultaneous equations (all sums are from $i=1$ to N)

$$\sum_i \frac{y_i}{\sigma_i^2} = \hat{a} \left(\sum_i \frac{1}{\sigma_i^2} \right) + \hat{b} \sum_i \frac{x_i}{\sigma_i^2} \quad (\text{E.9})$$

$$\sum_i \frac{x_i y_i}{\sigma_i^2} = \hat{a} \left(\sum_i \frac{x_i}{\sigma_i^2} \right) + \hat{b} \sum_i \frac{x_i^2}{\sigma_i^2} \quad (\text{E.10})$$

which have the solutions

$$\hat{a} = \frac{1}{\Delta} \left[\left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) \left(\sum_i \frac{y_i}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right) \left(\sum_i \frac{x_i y_i}{\sigma_i^2} \right) \right] \quad (\text{E.11})$$

$$\hat{b} = \frac{1}{\Delta} \left[\left(\sum_i \frac{1}{\sigma_i^2} \right) \left(\sum_i \frac{x_i y_i}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right) \left(\sum_i \frac{y_i}{\sigma_i^2} \right) \right] \quad (\text{E.12})$$

where

$$\Delta = \left(\sum_i \frac{1}{\sigma_i^2} \right) \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2 \quad (\text{E.13})$$

Estimation of Variances of \hat{a} and \hat{b}

Suppose that the quantity z_j is an explicit function of the j^{th} set of individual measurement u_j, v_j, \dots

$$z_j = f(u_j, v_j, \dots) \quad (\text{E.14})$$

Further assume that the most probable value for z_j is given by \bar{z} such that

$$\bar{z} = f(\bar{u}, \bar{v}, \dots) \quad (\text{E.15})$$

In the limit of an infinite number of sets of measurements we can write the variance of z as

$$\sigma_z^2 = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{j=1}^K (z_j - \bar{z})^2 \quad (\text{E.16})$$

For small deviations of z_j from \bar{z} we can express $z_j - \bar{z}$ in terms of the partial derivatives with respect to u and v

$$z_j - \bar{z} \approx (u_j - \bar{u}) \frac{\partial z}{\partial u} + (v_j - \bar{v}) \frac{\partial z}{\partial v} + \dots \quad (\text{E.17})$$

where the partial derivatives are understood to be evaluated at $u = u_j, v = v_j, \dots$. Substituting this result into Equation (E.16) we then have

$$\sigma_z^2 \approx \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{j=1}^K [(u_j - \bar{u}) \frac{\partial z}{\partial u} + (v_j - \bar{v}) \frac{\partial z}{\partial v} + \dots]^2$$

$$\begin{aligned} \approx \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{j=1}^K [(u_j - \bar{u})^2 \left(\frac{\partial z}{\partial u}\right)^2 + (v_j - \bar{v})^2 \left(\frac{\partial z}{\partial v}\right)^2 \\ + 2(u_j - \bar{u})(v_j - \bar{v}) \frac{\partial z}{\partial u} \frac{\partial z}{\partial v} + \dots] \end{aligned} \quad (\text{E.18})$$

If $\frac{\partial z}{\partial u}$, $\frac{\partial z}{\partial v}$, ... are to first order independent of j , then this Equation (E.18) may be rewritten as

$$\sigma_z^2 \approx \sigma_u^2 \left(\frac{\partial z}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial z}{\partial v}\right)^2 + 2\sigma_{uv}^2 \frac{\partial z}{\partial u} \frac{\partial z}{\partial v} + \dots \quad (\text{E.19})$$

If the measurements u_j, v_j, \dots are statistically independent, then $\sigma_{uv}^2 = 0$ and we are left with the result

$$\sigma_z^2 \approx \sigma_u^2 \left(\frac{\partial z}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial z}{\partial v}\right)^2 + \dots \quad (\text{E.20})$$

We can now apply this result to estimate the variances of \hat{a} and \hat{b} of Equations (E.11) and (E.12). The set of measurements $y_1, \dots, y_i, \dots, y_N$ correspond to u_j, v_j, \dots in Equation (E.14), so that the estimated variances of \hat{a} and \hat{b} are

$$\sigma_{\hat{a}}^2 \approx \sum_{k=1}^N [\sigma_k^2 \left(\frac{\partial \hat{a}}{\partial y_k}\right)^2] \quad (\text{E.21})$$

$$\sigma_{\hat{b}}^2 \approx \sum_{k=1}^N [\sigma_k^2 \left(\frac{\partial \hat{b}}{\partial y_k}\right)^2] \quad (\text{E.22})$$

From Equations (E.11) and (E.12)

$$\frac{\partial \hat{a}}{\partial y_k} = \frac{1}{\Delta} \left[\frac{1}{\sigma_k^2} \sum_i \frac{x_i^2}{\sigma_i^2} - \frac{x_k}{\sigma_k^2} \sum_i \frac{x_i}{\sigma_i^2} \right] \quad (\text{E.23})$$

$$\frac{\partial \hat{b}}{\partial y_k} = \frac{1}{\Delta} \left[\frac{x_k}{\sigma_k^2} \sum_i \frac{1}{\sigma_i^2} - \frac{1}{\sigma_k^2} \sum_i \frac{x_i}{\sigma_i^2} \right] \quad (\text{E.24})$$

Substituting Equation (E.23) into (E.21) we then have

$$\begin{aligned}
 \sigma_a^2 &\approx \sum_k \frac{1}{\sigma_k^2 \Delta^2} \left[\left(\sum_i \frac{x_i^2}{\sigma_i^2} \right)^2 - 2x_k \left(\sum_i \frac{x_i}{\sigma_i^2} \right) \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) + x_k^2 \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2 \right] \\
 &= \frac{1}{\Delta^2} \left[\left(\sum_i \frac{1}{\sigma_i^2} \right) \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right)^2 - 2 \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2 \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) + \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2 \right] \\
 &= \frac{1}{\Delta^2} \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) \left[\left(\sum_i \frac{1}{\sigma_i^2} \right) \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2 \right] \\
 &= \frac{1}{\Delta} \sum_i \frac{x_i^2}{\sigma_i^2} \tag{E.25}
 \end{aligned}$$

In a similar manner, substituting Equation (E.24) into Equation (E.22) yields

$$\sigma_b^2 \approx \frac{1}{\Delta} \sum_i \frac{1}{\sigma_i^2} \tag{E.26}$$

where Δ is given by Equation (E.13).

For the case where σ_i is independent of i , Equations (E.11), (E.12), (E.25), and (E.26) reduce to

$$\hat{a} = \frac{1}{\Delta} \left[\left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) \left(\sum_i y_i \right) - \left(\sum_i x_i \right) \left(\sum_i \frac{x_i y_i}{\sigma_i^2} \right) \right] \tag{E.27}$$

$$\hat{b} = \frac{1}{\Delta'} [N(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)] \quad (\text{E.28})$$

$$\sigma_{\hat{a}}^2 \approx \frac{\sigma^2}{\Delta'} \sum_i x_i^2 \quad (\text{E.29})$$

$$\sigma_{\hat{b}}^2 \approx N \frac{\sigma^2}{\Delta'} \quad (\text{E.30})$$

where $\Delta' = N(\sum_i x_i^2) - (\sum_i x_i)^2$. (E.31)

Δ' may be written in a more convenient form

$$\Delta' = N^2 \Delta x_{\text{rms}}^2 \quad (\text{E.32})$$

where

$$\Delta x_{\text{rms}}^2 \equiv \frac{1}{N} \sum_i (x_i - \bar{x})^2 \quad (\text{E.33})$$

where

$$\bar{x} \equiv \frac{1}{N} \sum_i x_i \quad (\text{E.34})$$

Using Equation (E.32) in Equation (E.30) we then

have

$$\sigma_{\hat{b}}^2 \approx \frac{\sigma^2}{N \Delta x_{\text{rms}}^2} \quad (\text{E.35})$$

which emphasizes the intuitive notion that the estimate of the slope of the fitted straight line becomes better for a broader range of the independent sample variable x_i .

APPENDIX F

DERIVATION OF THE ALGORITHMS USED BY VLBI 1 AND VLBI 2

In this appendix we examine the effect of mixing the radio-frequency signals with a local oscillator signal; then we will develop a practical algorithm, consistent with the maximum-likelihood (ML) estimation procedures of Section 3.2, for the processing of the recorded video signals to estimate the delay and delay rate. For simplicity we shall first analyze the case for a constant delay, τ , and then we will extend it to the case for a slowly changing delay (i.e. $|\dot{\tau}| \ll 1$).

Designate the radio-frequency signals arriving at station 1 and 2 as $x_1(t)$ and $x_2(t)$, respectively, with the corresponding radio-frequency spectra $X_1(\omega)$ and $X_2(\omega)$. Let $x_1(t)$ and $x_2(t)$ each be single-sideband (SSB) mixed with a local oscillator signal at frequency ω_0 . Designate the resulting video signals as $x_1^V(t)$ and $x_2^V(t)$ and their corresponding spectra as $X_1^V(\omega^V)$ and $X_2^V(\omega^V)$, respectively, where ω^V is the video frequency. After the mixing operation (assumed upper-sideband) the radio-frequency spectra are simply translated in frequency by ω_0 so that

$$X_1^V(\omega^V) = X_1(\omega) \tag{F.1}$$

$$X_2^V(\omega^V) = X_2(\omega) \tag{F.2}$$

From expression (3-12), the ML estimate of is found by computing

$$\max_{\tilde{\tau}} \left| \sum_{\omega} X_1(\omega) X_2^*(\omega) e^{-i\omega\tilde{\tau}} \right| \quad (\text{F.3})$$

which, using (F.1) and (F.2), is equivalent to

$$\max_{\tilde{\tau}} \left| \sum_{\omega^V} X_1^V(\omega^V) X_2^{V*}(\omega^V) e^{-i(\omega_0 + \omega^V)\tilde{\tau}} \right| \quad (\text{F.4})$$

Let us define the "counter-rotated video cross-spectrum" $S_{12}(\omega^V)$ as

$$S_{12}(\omega^V) \equiv X_1^V(\omega^V) \cdot X_2^{V*}(\omega^V) e^{-i\omega^V\tilde{\tau}} \cdot e^{-i\omega_0\tilde{\tau}} \quad (\text{F.5})$$

so that (F.4) becomes

$$\max_{\tilde{\tau}} \left| \sum_{\omega^V} S_{12}(\omega^V) \right| \quad (\text{F.6})$$

From the familiar Fourier transform relations

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{i\omega t} d\omega \quad (\text{F.7})$$

$$G(\omega) = \int_{-\infty}^{\infty} g(t) e^{-i\omega t} dt \quad (\text{F.8})$$

it is easy to demonstrate the relation

$$x_2^V(\tilde{\tau}-t) \leftrightarrow X_2^{V*}(\omega^V) e^{-i\omega^V\tau} \quad (\text{F.9})$$

so that using the property that multiplication in the frequency domain is equivalent to convolution in the time domain (see, for example, Papoulis, Reference 28, p. 159), we see that $S_{12}(\omega^V)$ of Equation (F.5) may be rewritten as

$$S_{12}(\omega^V) = \int_{-\infty}^{\infty} [x_1^V(\Delta\tau) \oplus x_2^V(\tau - \Delta\tau)] e^{-i\omega^V \Delta\tau} d\Delta\tau \cdot e^{-i\omega_o \tilde{\tau}} \quad (F.10)$$

where \oplus indicates convolution. Or, writing out the convolution operation explicitly, we have

$$S_{12}(\omega^V) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x_1^V(t) x_2^V(t + \tilde{\tau} - \Delta\tau) dt \right\} e^{-i\omega^V \Delta\tau} d\Delta\tau \cdot e^{-i\omega_o \tilde{\tau}} \quad (F.11)$$

where it is clear that the convolution, in this case is really just a "cross-correlation" of $x_1^V(t)$ with $x_2^V(t)$. If $x_1^V(t)$ and $x_2^V(t)$ are narrowband signals which have been sampled at the Nyquist rate with sampling period $\Delta\tau_s$, and if $\tilde{\tau}$ is within approximately $\pm\Delta\tau_s$ of the actual delay, τ , then the range of $\Delta\tau$ in (F.11) may be restricted to approximately $\pm 2\Delta\tau_s$ around $\tilde{\tau}$ since the cross-correlation function dies rapidly as $\tilde{\tau}$ moves away from τ . In practice, the cross-correlation is computed over a range $-3\Delta\tau_s \leq \Delta\tau \leq 3\Delta\tau_s$ so that $S_{12}(\omega^V)$ of (F.11) becomes a **summation**

$$S_{12}(\omega^V) \approx \sum_{m=-3}^{+3} R_{12}(\tilde{\tau}-m\Delta\tau_s) e^{-i\omega^V m\Delta\tau} \quad (\text{F.12})$$

where we have defined the "cross-correlation function"

$R_{12}(\tilde{\tau}-m\Delta\tau_s)$ as

$$R_{12}(\tilde{\tau}-m\Delta\tau_s) \equiv \sum_t x_1^V(t) x_2^V(t+\tilde{\tau}-m\Delta\tau_s) \quad (\text{F.13})$$

The summation over t in (F.13) is, by implication, performed over the discrete samples of $x_1^V(t)$ and $x_2^V(t)$.

For the Mark I recording system, where the recording bandwidth is 360 kHz and $\Delta\tau_s \approx 1.39 \mu\text{sec}$, $S_{12}(\omega^V)$ is normally computed for seven evenly-spaced frequencies $\omega^V = 45, 90, 135, 180, 225, 270, 315$ kHz.

In practice, a correction must usually be applied to $S_{12}(\omega^V)$ due to the fact that $x_1^V(t)$ and $x_2^V(t)$ can not be offset by exactly $\tilde{\tau}$, but are instead restricted to the nearest integral number of sample periods. If we designate the delay corresponding to the integral number of sample periods as $\tilde{\tau}'$, then the actual quantity computed in (F.13) will be $R_{12}(\tilde{\tau}'-m\Delta\tau_s)$ instead of the desired $R_{12}(\tilde{\tau}-m\Delta\tau_s)$. But clearly, from the form of the summation of Equation (F.12), this introduces a simple correction factor into the computation of $S_{12}(\omega^V)$, namely

$$S_{12}(\omega^V) \approx \sum_{m=-3}^{+3} R_{12}(\tilde{\tau}' + m\Delta\tau_s) e^{-i\omega^V m\Delta\tau} \cdot e^{-i\omega^V(\tilde{\tau} - \tilde{\tau}')} \quad (\text{F.14})$$

The last factor on the right-hand side of (F.14) is known as the "fractional-bit correction".

For the case where the delay rate, $\dot{\tau}$, is small but non-zero (i.e. $|\dot{\tau}| \ll 1$) we may approximate the spectrum of the signal received at station 2 as simply being Doppler-shifted by an amount $\omega_m \dot{\tau}$, where ω_m is the mid-band radio frequency. This approximation is valid over a time interval, ΔT , which is short enough that the change in delay over ΔT is small compared to the inverse bandwidth B , i.e., $\dot{\tau} \Delta T \ll 1/B$. The video signal at station 2 is Doppler-shifted by the same amount. Let $X_2^{V'}(\omega^V)$ be the spectrum of the video signal which would be received at station 2 if $\dot{\tau} = 0$; then the relation between the actual received video spectrum $X_2^V(\omega^V)$, and $X_2^{V'}(\omega^V)$ is simply

$$X_2^{V'}(\omega^V) \approx X_2^V(\omega^V + \omega_m \dot{\tau}) \quad (\text{F.15})$$

or, equivalently, the relation between the corresponding video signals $x_2^{V'}(t)$ and $x_2^V(t)$ is

$$x_2^{V'}(t) \approx x_2^V(t) e^{-i\omega_m \dot{\tau} t} \quad (\text{F.16})$$

But this is exactly equivalent to mixing $x_2^V(t)$ with an additional oscillator of frequency $\omega_m \dot{\tau}$ (the fringe rate) in order to bring the frequency spectra $X_1^V(\omega)$ and $X_2^V(\omega)$ into "alignment" before cross-correlation. Therefore, the inclusion of a delay rate may be accounted for by "numerically mixing" $x_2^V(t)$ with the frequency $\omega_m \dot{\tau}$ so that $R_{12}(\tilde{\tau}' - m\Delta\tau_s)$ becomes

$$R_{12}(\tilde{\tau}' - m\Delta\tau_s) \approx \sum_{t=0}^{\Delta\tau} \{ x_1^V(t) x_2^V(t + \tilde{\tau}' - m\Delta\tau_s) e^{-i\omega_m \tilde{\tau}(t + \tilde{\tau}' - m\Delta\tau_s)} \} \quad (F.17)$$

where it is understood that $\dot{\tau}\Delta\tau \ll 1/B$. In VLBI 1, the right-hand side of (F.17) is computed by first performing a direct cross-correlation of $x_1^V(t)$ and $x_2^V(t)$ over a short "sub-segment" of data of length Δt such that the fringe phase changes only a small amount over the interval Δt , i.e. $|\omega_m \dot{\tau}\Delta t| \ll 1$. The result of this sub-segment cross-correlation is then multiplied by $\exp\{-i\omega_m \dot{\tau}(t - \tilde{\tau}' - m\Delta\tau_s)\}$ and summed with the similar results from the other sub-segments within ΔT (see Section 4.3). The interval ΔT is chosen to be 0.2 seconds, the length of one record of data. The values of $\tilde{\tau}$ and $\dot{\tau}$ are updated for each record, and $R_{12}(\tilde{\tau}' - m\Delta\tau_s)$ and $S_{12}(\omega^V)$ are computed individually for each record. Then, as indicated by expression (3-12),

the values of $S_{12}(\omega^v)$ are summed over the observation and then over frequency, so that the ML estimation of delay and delay rate may be written as

$$\max_{\tilde{\tau}, \dot{\tilde{\tau}}} \left| \sum_{j=1}^7 \sum_{k=1}^K S_{12}^k(\omega_j^v) \right| \quad (\text{F.18})$$

where K is the total number of records in the observation and $S_{12}^k(\omega_j^v)$ is the "counter-rotated cross-video spectrum" for the k^{th} record as given by Equation (F.14).

REFERENCES

1. Carr, T. D., May, J., Olsson, C. N., Walls, G. F.,
IEEE NEREM Record, 7, 222 (1965).
2. May, J., Carr, T. D., Quart. J. Florida Acad. Sci.,
30, 1 (1967).
3. Broten, N. W., Legg, T. H., Locke, J. L., McLeish, C. W.,
Richards, R. S., Chisholm, R. M., Gush, H. P.,
Yen, J. L., Galt, J. A., Science, 156, 1592 (1967).
4. Bare, C. C., Clark, B. G., Kellermann, K. I., Cohen,
M. H., Jauncey, D. L., Science, 157, 189 (1967).
5. Moran, J. M., Crowther, P. P., Burke, B. F., Barrett,
A. H., Rogers, A. E. E., Ball, J. A., Carter,
J. C., Bare, C. C., Science, 157, 676 (1967).
6. Moran, J. M., Ph.D. Thesis, M.I.T., (1968).
7. Wade, C. M., Astrophysical Journal, 162, 391 (1970).
8. Smith, J. W., Nature Physical Science, 232, 150 (1971).
9. Wade, C. M., Gent, H., Adgie, R. L., Crowther, J. H.,
Nature, 228, 146 (1970).
10. Rogers, A. E. E., Radio Science, 5, 1239 (1970).
11. Swenson, G. W., Annual Review of Astronomy and
Astrophysics, 7, 353, (1969).
12. Shockley, W. R., B. S. Thesis, M.I.T. (1971).
13. Hinteregger, Ph.D. Thesis, M.I.T., (1972).
14. Hinteregger, H. F., Shapiro, I. I., Robertson, D. S.,
Knight, C. A., Ergas, R. A., Whitney, A. R.,
Rogers, A. E. E., Moran, J. M., Clark, T. A.,
and Burke, B. F., Science, 178, 396 (1972).
15. Preston, R. A., Ph.D. Thesis, M. I. T., (1972).
16. Preston, R. A., Ergas, R. A., Hinteregger, H. F.,
Knight, C. A., Robertson, D. S., Shapiro,
I. I., Whitney, A. R., Rogers, A. E. E., and
Clark, T. A., Science, 178, 407, (1972).

17. Rogers, A. E. E., "Spectral Line Interferometry and Interferometer Noise Analysis", M.I.T. Lincoln Laboratory Tech. Report 441, (1968).
18. Burke, B. F., Physics Today, 22, 54, (1969).
19. Kellermann, K. I., Scientific American, 226, 72, (1972).
20. Klemperer, W. K., Proc. IEEE, 60, 602, (1972).
21. Cohen, M. H., Annual Review of Astronomy and Astrophysics, 7, 619, (1969).
22. Cohen, M. H., Proc. IEEE, 61, 1192 (1973).
23. Williams, J. G., JPL Space Programs Summary, II, 37, (1970).
24. Ash, M. E., Determination of Earth Satellite Orbits, M.I.T. Lincoln Lab Tech. Note 1972-5, (1972).
25. Wozencraft, J. M., and Jacobs, I. M., Principles of Communication Engineering, Wiley and Sons, (1965).
26. Van Trees H. L., Detection, Estimation, and Modulation Theory, Part I, Wiley and Sons, (1968).
27. Davenport, W. B., and Root, W. L., An Introduction to the Theory of Random Signals and Noise, McGraw-Hill, (1958).
28. Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill, (1965).
29. Hildebrand, F. B., Advanced Calculus for Applications, Prentice-Hall, (1963).
30. Van Vleck, J. H., and Middleton, D., Proc. IEEE, 54, 2, (1966).
31. Weinreb, S., "A Digital Spectral Analysis Technique and its Application to Radio Astronomy", M.I.T. Tech. Report 412, (1963).
32. Staelin, D. H., The Detection and Measurement of Radio Astronomical Signals, Lecture Notes, (1966).

33. Thomas, J. B., Deep Space Network Progress Report, JPL Tech. Report 32-1526, VIII, (1972).
34. Dwight, H. B., Tables of Integrals and Other Mathematical Data, MacMillan, (1957).
35. Kraus, J. D., Radio Astronomy, McGraw-Hill, (1966).
36. Chao, C. C., "A Preliminary Estimation of Tropospheric Influence on the Range and Range Rate Data during the Closest Approach of the MM71 Mars Mission," JPL Tech. Memo. 391-129, (1970).
37. Evans, J. V., in Radar Astronomy, 99, McGraw-Hill, (1968).
38. Arzac, J., Ph.D. Thesis, Univ. of Paris, (1966).
39. Getsinger, W. J., and Kessler, A. H., "Computer-Design of Diode-Using Microwave Components, and a Computer-Dimensioned, X-band Parametric Amplifier," M.I.T. Lincoln Lab Internal Report, (1968).
40. Rogers, A. E. E., "Broad-band Passive 90° RC hybrid with low component sensitivity for use in the video range of frequencies," Proc. IEEE (Letters), 59, 1617, (1971).
41. Hewlett-Packard General Catalog, 1973.
42. Throne, D. H., Hewlett-Packard Journal, July 1968.
43. Weinreb, S., NRAO Internal Memo, May 1970.
44. Knight, C. A., Ph.D. Thesis, M.I.T., in preparation.
45. Hemenway, P. D., Ph.D. thesis, U. of Virginia (1974).
46. Payne, J., "Front-End Box Temperature Controller," NRAO Electronics Division Internal Report No. 81 (1968).
47. Cohen, M. H., Shaffer, D. B., Astronomical Journal, 76, 91 (1971).
48. Hazard, C., Sutton, J., Argue, A. N., Kenworthy, C. M., Morrison, L. V., and Murray, C. A., Nature Physical Science, 233, 89 (1971).
49. Kristian, J., Sandage, A., Astrophysical Journal, 162, 391 (1970).

50. Rogers, A. E. E., Counselman, C. C., Hinteregger, H. F., Knight, C. A., Robertson, D. S., Shapiro, I. I., Whitney, A. R., Clark, T. A., Astrophysical Journal, 186, 801 (1973).
51. USAF, private communication.
52. Cohen, M. H., Ap. Letters, 12, 81 (1972).
53. Adgie R. L., Crowther, J. H., and Gent, H., M.N.R.A.S., 159, 233 (1972).
54. Argue, A. N., and Kenworthy, C. M., M.N.R.A.S., 160, 197 (1972)
55. Couper, H. A., Ap. Letters, 10, 121 (1972).
56. Brosche, P., Wade, C. M., and Hjellming, R. M., Astrophysical Journal, in press (1973).
57. Hunstead, R. W., M.N.R.A.S., 157, 367 (1972).

BIOGRAPHICAL NOTE

Alan Robert Whitney was born in Cheyenne, Wyoming, on September 11, 1944. He attended public schools in Cheyenne and entered M.I.T. in September 1962. In 1964 he joined the M.I.T. Electrical Engineering Cooperative program in conjunction with Bell Telephone Laboratories and was awarded the B.S. and M.S. degrees in Electrical Engineering in June 1967. Upon entering graduate school at M.I.T., he was awarded a National Science Foundation fellowship for full-time graduate study leading to an advanced degree. He has engaged in part-time work with Bell Telephone Laboratories and M.I.T.'s Lincoln Laboratory, working in the areas of digital computers, frequency-stable lasers, and radio propagation through rain. He is a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu.

Among Mr. Whitney's major outside interests are music, hiking, canoeing, skiing and snowshoeing.



Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.5668 Fax: 617.253.1690
Email: docs@mit.edu
<http://libraries.mit.edu/docs>

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Some pages in the original document contain pictures, graphics, or text that is illegible.