MICROCOPY RESOLUTION TEST CHART
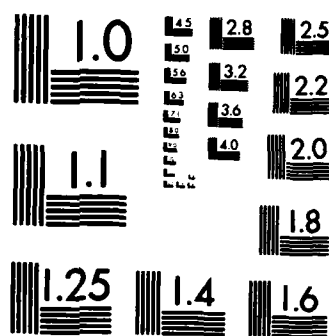
NATIONAL BUREAU OF STANDARDS-1963-A

AFGL-TR-84-0182

PRECISION AND ACCURACY OF INTERCONTINENTAL DISTANCE
DETERMINATIONS USING RADIO INTERFEROMETRY

T. A. Herring

Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

Scientific Report No. 1

July 1983

Approved for public release; distribution unlimited
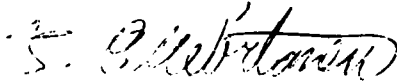
AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

CONTRACTOR REPORTS

This technical report has been reviewed and is approved for publication.

THEODORE E. WIRTANEN
Contract Manager

THOMAS P. ROONEY
Chief, Geodesy & Gravity Branch

FOR THE COMMANDER

DONALD H. ECKHARDT
Director
Earth Sciences Division

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFGL-TR-84-0182

6a. NAME OF PERFORMING ORGANIZATION Mass Institute of Technology	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Geophysics Laboratory (LWG)
6c. ADDRESS (City, State and ZIP Code) 77 Massachusetts Avenue Cambridge, MA 02139		7b. ADDRESS (City, State and ZIP Code) Hanscom AFB Massachusetts 01731

8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Geophysics Laboratory	8b. OFFICE SYMBOL (If applicable) LWG	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-82-K-0002

8c. ADDRESS (City, State and ZIP Code) Hanscom AFB, MA 01731 Monitor/Theodore E. Wirtanen	10. SOURCE OF FUNDING NOS.			

	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
	61102F	2309	2309G1	2309G1BD

11. TITLE (Include Security Classification) PRECISION AND ACCURACY OF INTERCONTINENTAL DISTANCE DETERMINATIONS USING RADIO INTERFEROMETRY

12. PERSONAL AUTHOR(S)
Herring, T.A.

13a. TYPE OF REPORT Scientific Rpt. No. 1	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) 83/7	15. PAGE COUNT 442

16. SUPPLEMENTARY NOTATION
Prepared in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the Massachusetts Institute of Technology

17	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Interferometry, Radio astronomy, Quasar

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See reverse side for abstract.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Theodore E. Wirtanen	22b. TELEPHONE NUMBER (Include Area Code) (617)861-3486	22c. OFFICE SYMBOL AFGL/LWG

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

ABSTRACT

Very-long-baseline interferometry (VLBI) is a technique which potentially can measure intercontinental distances with 1 cm precision. Already, measurements have been made which have nearly reached this goal, if we accept the standard deviations calculated from statistical arguments. However, the sources of errors in these measurements must be carefully analyzed and accounted for to ensure reliablity in the interpretation of results obtained from VLBI data analysis.

In this thesis, we concentrate on examining both the theory and the quality of the VLBI delay measurements. Using the redundant nature of the VLBI observations, we develop a number of quality tests. These tests are applied to data taken in VLBI experiments between July 1980 and January 1982. All of these tests lead to similar qualitative conclusions about the performance of the Mark III VLBI system. There are, however, differences from each of these tests in the quantitative conclusions about the system performance. These tests indicate that the actual statistics of the VLBI group delay observations do not match the theoretically calculated statistics. The differences between the actual and theoretical statistics are of two types. Firstly, there appears to be a proportional error in the calculated standard deviations of between 1.1 and 1.2, which is probably due to correlator non-reproducibility and the effects of instrumental dispersions. Secondly, there appears to be a threshold performance limit which can not be penetrated. This threshold limits the accuracy of the X-band ($\simeq 8.34$ GHz) group delay measurements to between 0.040 nsec ($\simeq 1.2$ cm) and 0.015 nsec ($\simeq 0.5$ cm), depending on the experiment. The S-band ($\simeq 2.3$ GHz) system appears to have a performance limit of 0.15 nsec ($\simeq 4.5$ cm) for all of the experiments analyzed.

We have also used data from the VLBI experiments conducted between July 1980 and June 1983, to estimate the distances between radio telescopes in North America and Europe. The estimate of the Westford, Massachusetts to Onsala, Sweden baseline length, obtained from 40 experiments using these sites, has a weighted root-mean-square repeatability of 2.0 cm and a statistical standard deviation of 0.6 cm. However, studies of these solutions indicate that the actual standard deviation of the baseline length estimate is probably more in accord with repeatability rather than with the statistical estimate of its value obtained from the analysis of the ensemble of data. The estimate of the rate of change of the length of this baseline is 1.6 ± 0.5 cm/yr, which is consistent with the inferred average rate of 1.7 cm/yr.

The determination of the tidal parameters of the earth are also investigated. The results obtained are consistent with the currently accepted values for these parameters.

THE PRECISION AND ACCURACY OF INTERCONTINENTAL
DISTANCE DETERMINATIONS USING RADIO INTERFEROMETRY

by

Thomas Abram Herring

M. Surveying
University of Queensland
(1978)

B. Surveying (Hons)
University of Queensland
(1976)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
at the
Massachusetts Institute of Technology
July, 1983

Signature of Author : _____
Department of Earth and Planetary
Sciences, July 20, 1983

Certified by : _____
Thesis Supervisor

Accepted by : _____
Chairman, Department Graduate Committee

-4-

Acknowledgements

This thesis would not have been possible without the aid of many people.

I wish to especially thank Irwin Shapiro, for not only his guidance and suggestions throughout the preparation of this thesis, but also for his instilling in me the techniques of careful and thorough research.

Much of the contents of this thesis would not have been possible without innumerable conversations with Alan Rogers of the Haystack Observatory. His tolerance in answering my often ill-conceived questions about the Mark III system led to my better understanding and appreciating much of this system.

My thanks also go to Jim Ryan and Chopo Ma at Goddard Space Flight Center, who not only aided this thesis with discussions on VLBI, but also with their hospitality during my visits to Goddard Space Flight Center.

During the past four years I have had many enlightening conversations which have given me new insights into many problems. For these discussions, I am grateful to Alan Whitney, Doug Robertson, Chuck Counselman, Bob King, Jim Davis, Dave Shaffer, Tom Clark, Juan Marcaide and Michael Ratner.

I am especially indebted to Mark Reid, Matt Schneps, and Shoshanna Rosenthal for their allowing my use of the Hewlett Packard computer at the Smithsonian Astrophysical Observatory. Without the use of this computer, much of the computations which went into this thesis, would have been impossible.

-5-

My interest in research was started in Australia. I owe
many thanks to the staff of the Surveying Department of the
University of Queensland. I especially wish to thank Mete
Nakiboglu who guided my progress through two earlier theses.

Of course, it is unlikely that I would have ever reached
this stage without the early guidance and support of my
parents -- Kevin and Vilma.

I have saved to last, the person I wish to thank most --
my wife, Laura. Her support and friendship from my first days
in the United States have made this work possible. Her
efforts were not only in the form of support for these many
years, but also by word processing most of the text of this
thesis.

TABLE of CONTENTS

1. Introduction

In April 1968 the first bandwidth-synthesis, very-long-baseline-interferometry (VLBI) experiment was conducted using the NRAO Mark I recording system (Whitney, 1974). Two experiments later in January 1969 the bandwidth synthesis technique was successfully used to measure accurate group delays (errors <10 nsecs or ≈3 m effective path length). The January 1969 experiment yielded estimates of the relative station locations of two radio telescopes, one in Westford, Massachusetts, the other in Green Bank, West Virginia which were separated by approximately eight hundred kilometers. These estimates agreed with conventional survey results to within 2 m in the distance between the sites (or baseline length) and within 5 m in orientation. Thirteen years after the first successful experiment, baseline lengths of over eight thousand kilometers are now routinely being measured using the Mark III VLBI system with precisions of a few centimeters -- unmatched by any conventional surveying techniques. In this thesis we will investigate the accuracy of the results achievable with the currently available Mark III system and the improvement we can expect in the future.

The measurement of group delays gave the bandwidth synthesis VLBI system a major advantage over other VLBI systems available in 1969. Prior to this time the accuracy of the group delay measurements had been limited by the bandwidth of the recording systems. The maximum recorded bandwidth

available was 4 MHz for the Canadian long baseline interferom-
etry (LBI) system. The NRAO Mark I recording system was
capable of recording bandwidth of only 360 kHz. The bandwidth
synthesis system (in 1969) measured group delays by physically
changing the frequency of the second local oscillator in the
receiver and sequentially sampling the output of the receiver
with the narrow bandwidth recorder. This method allowed a
large bandwidth to be sampled ($\equiv$"spanned") while recording
with a narrow bandwidth. The accuracy of the group delay
determination then became a function of the spanned bandwidth
rather than the recorded bandwidth. With the Mark III VLBI
system it is no longer necessary to switch the frequency of
the local oscillators because signals from up to 28 channels,
each of 2 Mhz bandwidth can be recorded simultaneously,
eliminating the need to switch frequencies.

Even before the first successful bandwidth synthesis VLBI
experiment, the applications of such a powerful geodetic
technique were being envisaged (Shapiro, 1968; Shapiro and
Knight, 1970). Potentially, intercontinental baseline lengths
could be measured with precisions of a few centimeters. The
applications of such results would be numerous. For the first
time, it would be possible to determine the current-day plate
velocities with a time resolution of only a few years. These
rates could be measured between points thousands of kilometers
away from active spreading centers or convergence zones,
thereby being immune to any local effects at these centers or
zones. Also, the technique could be used to monitor earth

rotation and polar motion with unprecedented precision and time resolution. It was envisaged that the position of the rotation axis with respect to the crust of the earth could be determined with a precision of better than 10 cm every day. The system would provide an almost inertial reference frame because of the great distances to the quasars which were being observed. After several decades of operations the precession constant and terms in the nutation series could be accurately measured. This information would then be useful in constraining earth rheology and structure models.

For all of the above goals to be realized three factors are essential: accurate observations, accurate models and appropriate statistics to account for any inadequacy in the models. In this thesis we will concentrate on the first of these factors. We commence in Chapter 2 with a detailed review of the VLBI observables. The fundamental observable of VLBI is the phase of the cross spectrum of the data recorded at two sites. However, this observable is not readily used because in general we have no way of determining the integral number of cycles, or ambiguities, associated with the phase observation. The bandwidth-synthesis technique overcomes this problem by using the group delay, i.e. the derivative of the phase with respect to (angular) frequency. We will investigate the nature of these observables and in particular consider the quantities which affect the phase observable and hence, in general, the group delay. Also in Chapter 2 we will discuss the phase-calibration system. We will associate this

particular calibration with the observations themselves because, when available, it is automatically applied during the processing of the recorded data. The term observation has now implicitly come to mean "phase-calibrated observation."

In Chapter 3 we will consider methods of assessing the accuracy of the VLBI observations. Specifically, we will consider closure of observations around triplets of baselines, analysis of the behavior of the residual phase in each frequency channel and the behavior of the phase-calibration system.

A major advance in eliminating the random error and minimizing some of systematic errors in VLBI observations would be achieved by eliminating the phase-delay ambiguities. With current Mark III observations from two widely separated frequencies, the group delays should be sufficiently accurate to allow this elimination (observations made at two widely seperated frequencies allow the plasma delays to be estimated). This problem is dealt with in Chapter 4. In many respects, the investigations of phase-delay ambiguity estimation can be considered as accuracy checking. We start by evaluating the approximations made in using dual-frequency data to calibrate for the delay through the ionosphere. We then investigate how these formulas should be applied to VLBI data. To check the algorithms used to remove ambiguities, we have applied them to data taken on pairs of sources which were observed repeatedly over many hours. With such closely spaced observations we are able to "phase connect" the data without

use of the group delay. We are then able to compare these connected phase delays with the predicted ambiguities based on the group delays.

The studies of phase-delay ambiguity prediction have produced a technique which can be used to check the accuracy of the VLBI group delay measurements. We then investigate in Sections 4.3 and 4.4 the accuracy of the group delay measurements from both the phase connected experiments and a geodetic schedule (i.e., a sequence of observations which are used to determine estimates of geodetic parameters).

In Chapter 5, we will summarize the estimates of baseline lengths obtained from VLBI experiments spanning the interval from May 1978 to June 1983. In addition, we will also present estimates of the Love numbers h and ℓ and the tidal lag angle.

The weight of this thesis may imply that all aspects of radio interferometry are addressed in this volume. This thesis, however, will address only topics which are applicable to geodetic uses of wideband, very-long-baseline interferometric observations of extragalatic radio soucres.

Many of the questions which VLBI measurements will be able to answer are related to geophysics. Consequently, this thesis will not only be directed towards experts in VLBI, but also towards the potential users of the measurements obtained with VLBI -- earth scientists. Users of the system must be aware not only of the system's advantages but also the system's disadvantages and limitations. This need is most evident in the interpretation of the uncertainty of, and the

correlation between, results. It is critical that VLBI users
(actually users of all advanced geodetic techniques) be aware
of underlying assumptions used in data acquisition and proces-
sing, especially when these assumptions may introduce correla-
tions between results or unrealistically small uncertainties.
Throughout the thesis a conscious effort will be made to make
all asumptions and their consequences explicit. In this way
it is hoped that the full capability and limitations of the
Mark III VBLI system can be understood by all readers.

2. THE VLBI OBSERVABLES

VLBI measurements can be used to obtain accurate esti-
mates of geodetic parameters, e.g. the relative positions of
radio telescopes and radio sources and the location of the
earth's rotation axis with respect to the crust of the earth
and the source coordinate frame. Analysis of changes in these
geodetic parameters over a period of time could be used to
answer questions about the dynamics of the earth's crust and
the properties of the earth's interior. However, the esti-
mates of the geodetic parameters are themselves the end result
of a process which contains many steps. In this chapter we
investigate the first step of this process - the inference of
the values of the VLBI observables. For the geodetic applic-
ations of VLBI these observables consist of group delay, phase
delay and phase delay rate. Although these quantities will be
referred to as the VLBI observables, the measurements made
during an experiment consist of recordings of radio signals
from extragalatic radio sources. These recordings, which not
only contain signals from the radio source but also noise from
the radio receiver and the regions surrounding each antenna,
are not used directly to estimate geodetic parameters, but
they do contain information which is used for this purpose in
a later stage of the analysis.

The information we can use is the difference in the
arrival times of the signals from the radio source at two
sites. We would expect that the time of arrival of the

signals from the radio source would be a function of the
location of the radio telescopes with respect to the radio
source (Figure 2.1). When two radio telescopes observe the
same source we should be able to compare the recordings from
the two sites and deduce the difference in arrival times of
the signals at the two telescopes. In addition, since the
radio telescopes are located on a rotating earth, we should be
able to measure the rate of change of the arrival times. This
latter quantity is the phase delay rate. But how do we obtain
optimum estimates of these quantities from the data recorded
at each telecope? To answer this question we will need to
know the nature of the signals being received from the radio
source and the nature of any noise sources. We will study
this problem in Section 2.1. In Section 2.2 we will investi-
gate the methods used to infer the values of the VLBI
observables from the recordings at each radio telescope, and
in Section 2.3 we will investigate the phase calibration
system which is used to calibrate the propagation delay
through the receiver system of the radio telescope.

2.1 The properties of the signals recorded during a VLBI observation

In this section we will investigate two types of radio
emission which are encountered in geodetic VLBI observations
of extragalatic radio sources. These two types are thermal
emission and electron synchrotron emission. The thermal

Figure 2.1 Generalized and simplified geometries of a VLBI observation

a. Generalized case

b. Simplified case
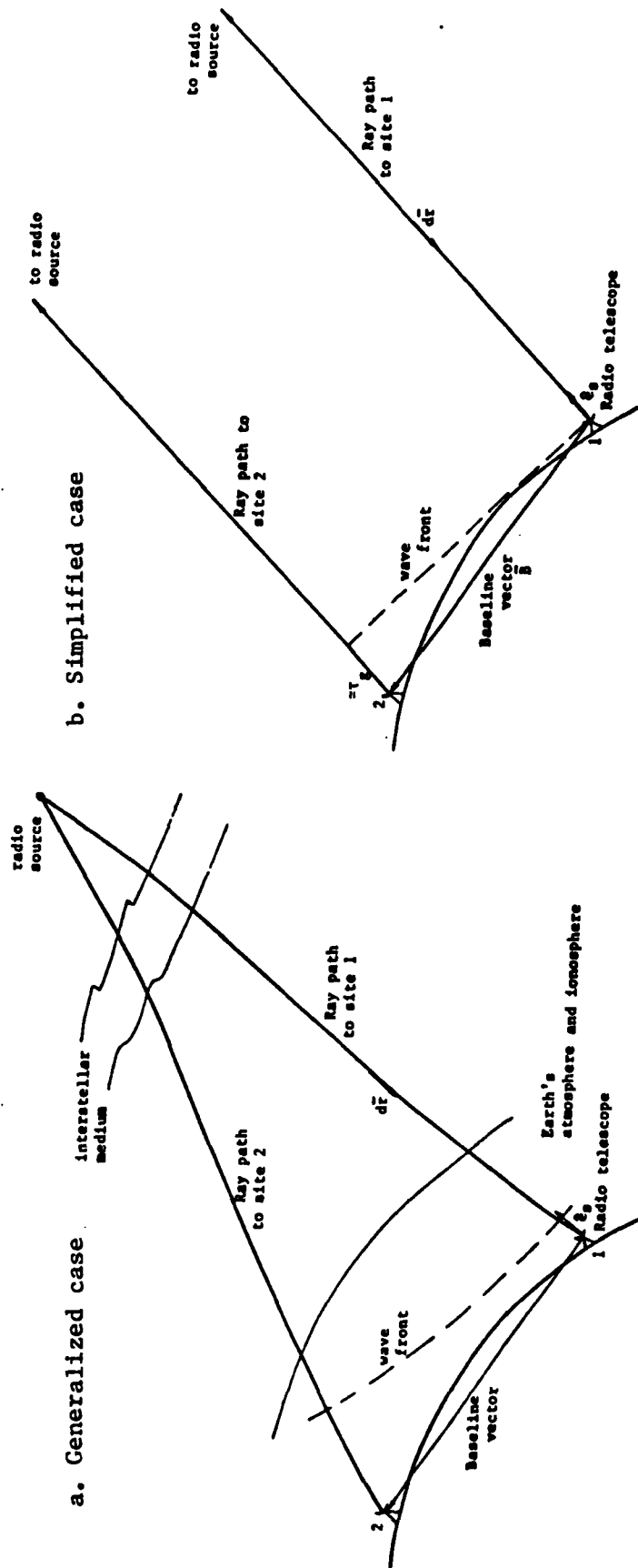
The group delay will be the difference in the arrival times at the two sites, of the signals from the radio source.

The group delay $\tau_g = t_2 - t_1$, where

$$t_2 = \int_{source}^{site\ 2} n_g(\bar{r})d\bar{r}/c \quad \text{and} \quad t_1 = \int_{source}^{site\ 1} n_g(\bar{r})d\bar{r}/c$$
$$\text{along ray path} \qquad \text{along ray path}$$

where $n_g(\bar{r})$ is the group refractive index.

For the simplified geometry case, $n_g(\bar{r})$ is independent of $\bar{r}$ and

$$t_2 = t_1 - n_g(\bar{B}\cdot\hat{e})/c \quad \text{and} \quad \tau_g = - n_g(\bar{B}\cdot\hat{e})/c$$

Expressions for the phase delays may be obtained by substituting the phase refractive index for the group refractive index.

emissions are generally associated with noise sources, e.g.
receiver noise, cable losses and atmospheric emission, whereas
the synchrotron emission is the dominant emission process
occurring in the radio sources (generally quasars) which are
observed. In studying these mechanisms we will mainly be
interested in two of their properties -- the statistical prop-
erties of the electromagnetic field components of the emission
and the power spectral density of the emission. The power
spectral density is of interest to us because its Fourier
transform, when normalized, will be the autocorrelation
function of the signals in the time domain.

Thermal emissions in all bodies are due to changes in the
quantum states at random times. If the body is a perfect
absorber (and hence a perfect radiator), it is referred to as
a blackbody and the brightness of the radiation from the body,
B_r, at a given frequency is a function only of the temperature
of the body (See Table 2.1.1 for definitions of brightness
and other quantities, as given by Krauss (1966)). This
relationship is described by Planck's radiation law

$$B_r = \frac{2hf^3}{c^2} \frac{1}{e^{hf/kT}-1}$$

where $\underline{h}$ is Planck's constant (6.63×10^{-34} Joule sec), $\underline{f}$ is
frequency (Hz), $\underline{c}$ is the velocity of light ($\approx 3 \times 10^8$ m sec^{-1}),
$\underline{k}$ is Boltzmann's constant ($\approx 1.38 \times 10^{-23}$ Joule K^{-1}), and T is
the temperature of the body (K).

This type of radiation characterizes many of the noise

Table 2.1.1 Definitions of Radio Source Properties

Brightness B_r (w/m^2-Hz-rad^2) — Power received per unit area per unit solid angle per unit bandwidth

Power $dW = B_r \cos\theta\, d\Omega\, dA\, d\nu$ (w) — power received in bandwidth $d\nu$ from solid angle $d\Omega$ incident on area dA with angle θ between normal to surface and direction to emission

Spectral power $dw = B_r \cos\theta\, d\Omega\, dA$ (w/Hz) — power per unit bandwidth

Flux density $F = \iint_{source} B_r(\theta,\phi)\, d\Omega$ — flux density of source

(w/m^2-Hz) = (10^{26} Janskys)

Power pattern $P_n(\theta,\phi)$ (dimensionless) — response of an antenna to power from direction θ,ϕ normalized to 1 at $P_n(0,0)$

Observed flux density $F_o = \iint_{source} B_r(\theta,\phi)\, P_n(\theta,\phi)\, d\Omega$

(if source is small and $P_n(\theta,\phi) \approx 1$, then $F_o \approx B_r\, \Omega_s$ where Ω_s is the angular extent of the source)

Poynting vector $\bar{P} = (\bar{E} \times \bar{H})$ (ws/m^2) (MKSA units) — power flow through unit area

Electric field strength $\bar{E}$ v/m

Magnetic field strength $\bar{H}$ a/m

For time harmonic fields $\bar{P}(\omega) = \bar{E}(\omega) \times \bar{H}^*(\omega)$

$P(\omega) = 2F$ if $\bar{E}$ and $\bar{H}$ are interpreted as field densities

Rayleigh-Jean's Law (for $f \ll kT/h$ and black body radiation)

$$B_r = 2kT/\lambda^2$$
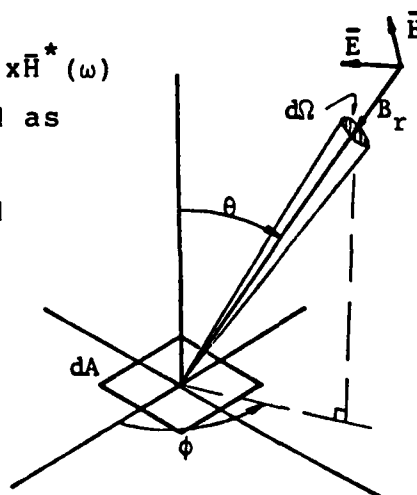
sources encountered in radio astronomy. Since we are gener-
ally interested in determining if a signal is greater than the
noise level, it has become common practice to associate emis-
sions of all types with an equivalent blackbody temperature.
However, this temperature can only be associated accurately
with a physical temperature for blackbodies.

We can deduce some statistical properties of this type of
radiation based only on the nature of the emission process.
Because the radiation is due to the sum of the emissions from
individual electrons which are changing quantum states at
random times we would expect the components of the electric
field (E field) to be random. In addition, since the compo-
nent of the E field, at a given frequency, will be the sum of
the emissions from many individual electrons changing state,
we can use the central limit theorem of probability to deduce
that these components will be Gaussianly distributed. (Since
the E field probability distribution is probably different for
each state change we need to use the relaxed version of the
central limit theorem which does not require identical proba-
bility distributions. This substitution is valid since no
single electron's emission contributes nearly all of the total
energy radiated.) Since the spectrum of the emissions has
components at high frequency (see Planck's radiation law), we
also know that the signals will have a very short correlation
time.

Electron synchrotron radiation is the other major emis-
sion process we will consider. Any accelerating free electron

will emit electromagnetic radiation. Any free electron moving in a uniform magnetic field will travel in a helical path (neglecting collisions between the electron and other atomic particles). If the electron's velocity is much less than the speed of light, the radiation pattern of the emission will extend over a large angular range perpendicular to the acceleration vector of the electron (Figure 2.1.1a). When the velocity of the electron is near that of the speed of light, the radiation from the electron, as seen from the frame of the observer, will appear to "beam", i.e. radiate in a very narrow cone in the instantaneous direction of motion of the electron (Figure 2.1.1b). We now investigate the nature of this type of emission.

If the observer is in the plane of the motion of the electron, he or she will observe a short pulse of radiation each orbit of the electron. In Table 2.1.2 some of the major formulas associated with synchrotron radiation are given (compiled from Jackson 1975, Chapter 14). Since the emission from the electron is periodic (assuming negligible energy loss and hence ensuring that each pulse is nearly identical), the power spectrum of the emission will only have components at discrete frequencies, i.e. the emission will not have a con- tinuous power spectrum. The frequency spacing in the spectrum will be the inverse of the orbital period of the electron. For the case given in Table 2.1.2, the frequency spacing will be approximately 1.4×10^{-5} Hz. From the duration of the pulse we can also calculate approximately the frequency component in

Figure 2.1.1 Radiation patterns from accelerating electrons
in a uniform magnetic field
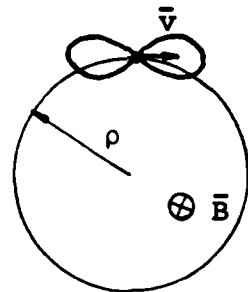
a. Radiation pattern when v<<c
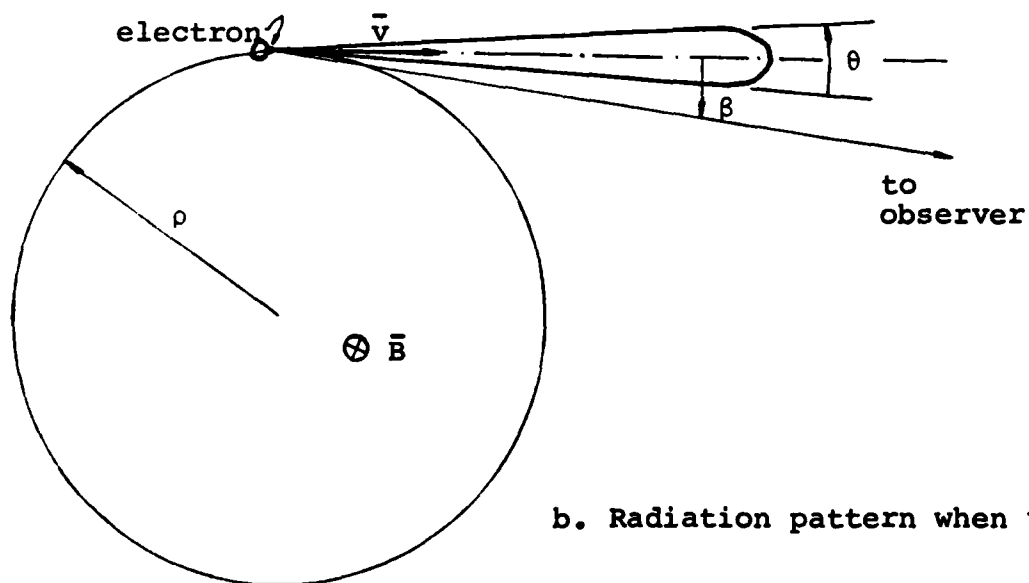
b. Radiation pattern when v≲c

Table 2.1.2 Synchrotron radiation equations (MKS units)

For an electron with rest mass m_o moving at velocity $\underline{v}$ in the region of a uniform magnetic field $\bar{B}_m$ we have (see Figure 2.1.1):

<u>Energy of electron</u> $\xi = m_o c^2 \gamma$

where $\gamma^{-2} = 1 - v^2/c^2$, c = velocity of light

<u>Radius of orbit</u> $\rho = m_o c \gamma/(e B_m)$, e= charge of electron,

<u>Beam width</u> $\theta = 1/(2\gamma)$

<u>Pulse duration</u> $t_p = m_o/(2 e B_m \gamma^2)$

<u>Energy radiated/(rad/sec)-unit solid angle</u>

$$ I = \frac{e^2 c}{3\pi^2 c} \left[\frac{\omega\rho}{c}\right]^2 \left[\frac{1}{\gamma^2} + \beta^2\right]^2 \left[K_{2/3}^2(\psi) + \frac{\beta^2}{(1/\gamma)^2 + \beta^2} K_{1/3}^2(\psi)\right] $$

where $\underline{e}$ is the charge of an electron, ω is frequency (rad/sec), β is the angle between the observer and the direction of the instantaneous velocity of the source, $K_{2/3}$ and $K_{1/3}$ are Airy functions and

$$ \psi = \frac{\omega\rho}{3c} \left[\frac{1}{\gamma^2} + \beta^2\right]^{3/2} $$

For $\omega \ll 1/t_p$

$$ I_{|\beta=0} \approx e^2 c \left[\frac{\omega\rho}{c}\right]^{2/3} \left[\frac{\Gamma(2/3)}{\pi}\right]^2 (3/4)^{1/3} $$

<u>Example</u> For $\xi = 10^{13}$ eV=1.6×10^{-6} J, $B_m = 10^{-4}$ Gauss=10^{-8} Wb/m^2, e = 1.6×10^{-19} Coul., $m_o = 9.1 \times 10^{-31}$ kg, c= 3×10^8 m/sec, d = 10^9 ly (10^{25} m) and $\omega = 50 \times 10^9$ rad/sec, we have $\gamma = 1.95 \times 10^7$, $\rho = 3.33 \times 10^{12}$ m, $\theta = 2.56 \times 10^{-8}$ rad (5.3 mas), $t_p = 7.5 \times 10^{-18}$ sec, orbit period = 7.0×10^4 sec

$I_{|\beta=0} = 8.8 \times 10^{-21}$ Joule/(rad/sec)-rad^2

<u>Average power radiated each orbit</u> = 7.2×10^{-24} w/Hz-rad^2. 10^9 light years from the source the flux density from this single electron would be dF = 7.2×10^{-75} w/m^2-Hz and hence the number of electrons necessary to produce a 1 Jansky (10^{-26} w/m^2-Hz) source is <u>1.4×10^{47}</u>.

the spectrum with the maximum power. This frequency for an electron with an energy of 10^{13} ev is 1.3×10^{17} Hz; see Table 2.1.2. (We have taken a maximum value for the energy of an electron in a quasar to ensure that the calculated number of electrons is a minimum so that we will be sure that the application of the central limit therom will always be valid (see discussion later in this section).)

In Table 2.1.2 we calculate the number of electrons needed to produce a radio source with a flux density of a 1 Jy, 10^9 light years from the observer (a typical distance for quasars). The estimated 1.4×10^{47} electrons is certainly an underestimate because we assumed all electrons were beaming directly at the observer and no account was made for absorption or scattering of the radiation as it propagated from the source to the receiver. Since the emission from at least 1.4×10^{47} electrons sum together to yield the total radiated emission, we may use the central limit theorem to deduce that the spectral components of the radiation will be Gaussianly distributed provided the emissions from the electrons (or at least groups of electrons) are independent. Since a uniform magnetic field is necessary to produce the helical orbits of the electrons we may expect that this field could cause some dependence of the emissions from different electrons. However, two properties of the emission from quasars would suggest that the emissions from the electrons are not strongly correlated. Firstly, the emission from a single electron is linearly polarized if the electron is moving in a plane

circular orbit, but emission from most quasars shows only a small amount of linear polarization (see, e.g., Haves, 1975; Altschuler and Wardle, 1977; Conway et al., 1977; Simard-Normandin et al., 1981; Simard-Normandin et al.,1982) indicating that the magnetic fields in the emission region are not uniform or that the radiation has been depolarized as it propagated from the source to the earth, e.g. by scattering. Secondly, if all the emissions were strongly correlated then it would be highly unlikely that the earth would fall within the beam of the emission and we would probably not be able to see any quasars unless the actual number of quasars were very large.

From the simplified consideration of synchrotron radiation we have learned two important properties of synchrotron radiation: 1) The components of the the Fourier transform of the signals from the quasar (i.e., components of the spectrum) are Gaussianly distributed and 2) the power spectrum of the emission from quasars is effectively continuous and extends to high frequency. (While the radiated power from individual electrons occurs at discrete frequencies, the power spectrum of the sum of the emissions from many electrons with different energies will fill nearly all of the frequencies in the power spectrum.)

The signals emitted from the radio sources are not directly used to infer the values of the VLBI observables. The recordings of these signals are used. The signals from a quasar will propagate along separate paths to each radio

telescope observing the source. The propagation medium along
each of these paths will, in general, have different proper-
ties. For each path there will be possibly different propaga-
tion velocities and different absorption characteristics, i.e.
there will be losses from the high and low frequency compon-
ents of the spectra of the signals by the time the signals
reach the radio telescopes.

The signals will, however, undergo most changes as they
propagate through the receiving electronics at the radio
telescopes. In Figure 2.1.2 we show a schematic diagram of a
VLBI radio receiver and recorder equipment. Along with the
diagram of the receiver we also show schematically the power
spectrum of the signal at a number of locations in the
receiver.

The first major change to the signal's spectrum occurs as
the signal propagates through the feed horns which are design-
ed to receive a relatively narrow range of frequencies (usual-
ly several hundred MHz). For Mark III geodetic VLBI experi-
ments there are generally two feed horns, one for X-band (≈ 8
GHz) and another for S-band (≈ 2 GHz). There will be losses in
the power of the spectral components with frequencies falling
outside the operating range of the feed horns. Also there
will be dispersive propagation delays introduced to the signal
(dispersive is used here to indicate that the derivative of
the phase with respect to frequency is not independent of
frequency). We will investigate the dispersive delays through
the feed horns in Section 3.1.

Figure 2.1.2 Schematic diagram of a radio telescope receiver for VLBI
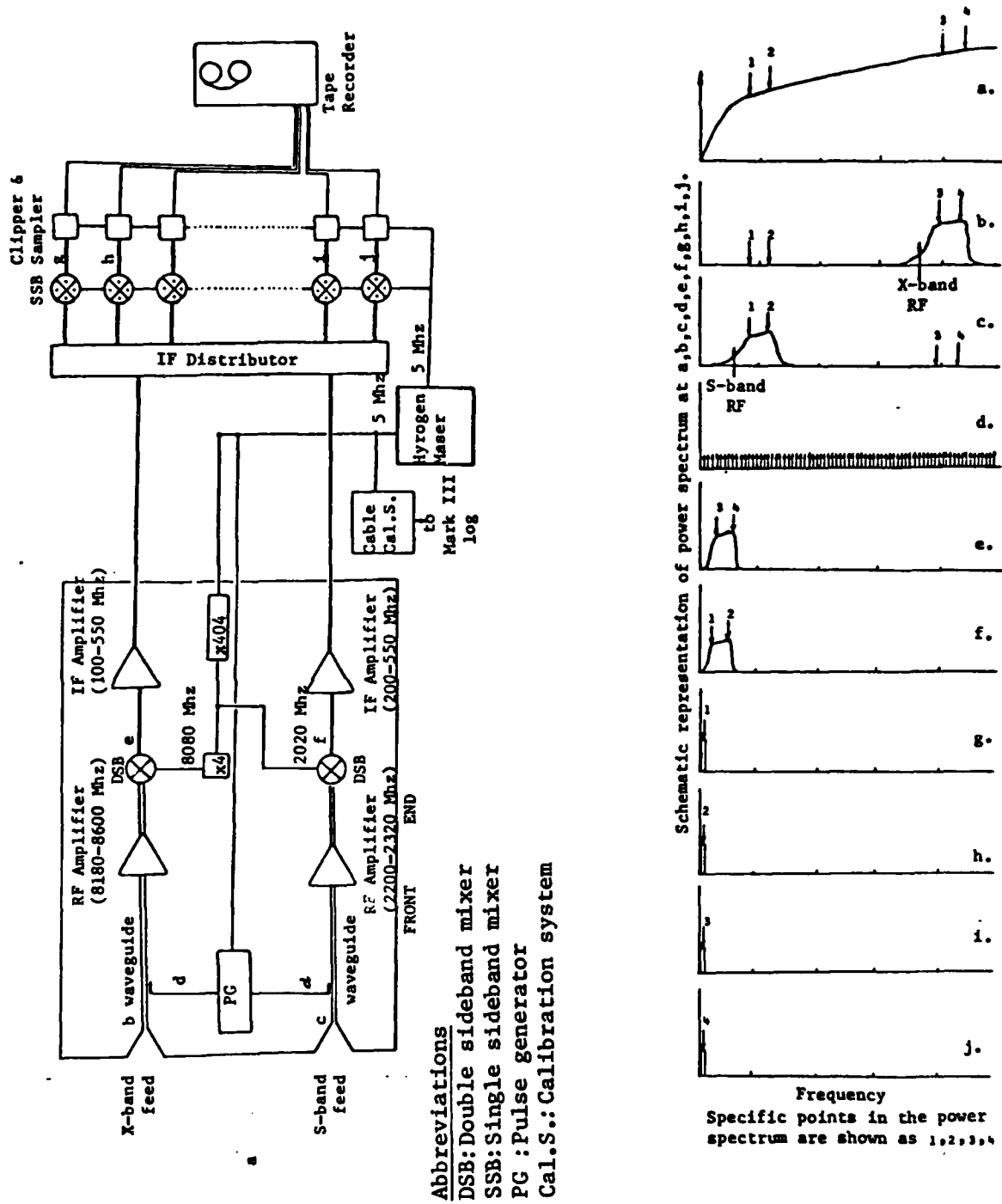
After passing through the feed horn the signals propagate along a waveguide towards the first amplifier. Shortly after entering the waveguide a companion signal is added to the signal received by the antenna. This companion signal is a series of low power pulses, each of less than 50 psec duration and spaced at 1 μsec intervals. These pulses allow the delays through the receiver system to be calibrated. Their operation will be discussed in detail in Section 2.3.

The signal and the phase calibration pulses undergo extensive modification as they propagate through the receiver. Shortly after the injection of the phase calibration pulses the signals are amplified. The amplified signals are high pass filtered and then multiplied by a high frequency signal (8.08 GHz for X-band and 2.02 GHz for S-band). These frequencies are referred to as the first local oscillator (LO) frequencies. The high pass filter is such that all signals with frequencies less than that of the first LO are removed. This multiplication or heterodyning of the signal will shift its spectrum. (Heterodyning is also often referred to as mixing.) If we denote the original spectrum by $\bar{S}(\omega)$ then the heterodyned signal will have a spectrum given by (Bracewell, 1978, p.108 - modulation theorem)

$$\bar{S}_m(\omega) = \bar{S}(\omega+\omega_r) \, e^{+i\phi_r} + \bar{S}(\omega-\omega_r) \, e^{-i\phi_r}$$

where ω_r and ϕ_r are the frequency and phase of the local-oscillator signal, and $\bar{S}_m(\omega)$ is the spectrum of the signal after mixing. We have dropped the factors of one half which

would normally appear in this equation because of the amplification of the signal which accompanies the mixing operation.

This signal is band pass filtered to remove the frequency components near 16 GHz for X-band and near 4 GHz for S-band. This intermediate frequency (IF) signal is now amplified and sent via coaxial cable to the IF distributor. The signals we wish to observe are now in the frequency ranges of 100-550 MHz for X-band and 200-550 MHz for S-band (see Figure 2.1.2).

The IF distributor, as its name implies, separates the IF signals into 28 separate channels which are input to 28 video converters. These video converters will mix a frequency (different for each video converter) between 100-550 MHz with the IF signal. This operation will shift the spectrum of the signal into the range 0-2 MHz. Each video converter will shift a different part of the spectrum into this range. The mixers in the video converters are single sideband mixers and will output two channels, one for the upper sideband and the other for the lower sideband. The output from either or both channels may be recorded on magnetic tape. (If both upper and lower sideband signals are recorded, they are recorded on separate tracks on the tape. When signals from both sidebands are recorded the total number of widely spaced frequency channels which can be sampled is reduced by one half from the number which could be sampled if only one (single) sideband were recorded. The advantages of recording both sidebands will be discussed in Appendix B.)

In Figure 2.1.3 the effects of the mixing of the signal

Figure 2.1.3 Schematic representation of the effects of mixing

Original spectrum

Re

Im

$S(\omega)$

ω

$-\omega_o$ ω_o

$x \cos(\omega_o t + \phi_o)$

Re

Im

$S(\omega+\omega_o)e^{i\phi_o}$
$+S(\omega-\omega_o)e^{-i\phi_o}$

ω for all ω

Upper sideband

Re

Im

$S(\omega+\omega_o)e^{i\phi_o}$ $\omega>0$
$S(\omega-\omega_o)e^{-i\phi_o}$ $\omega<0$

ω

Lower sideband

Re

Im

$S(\omega-\omega_o)e^{-i\phi_o}$ $\omega>0$
$S(\omega+\omega_o)e^{i\phi_o}$ $\omega<0$

ω

Legend
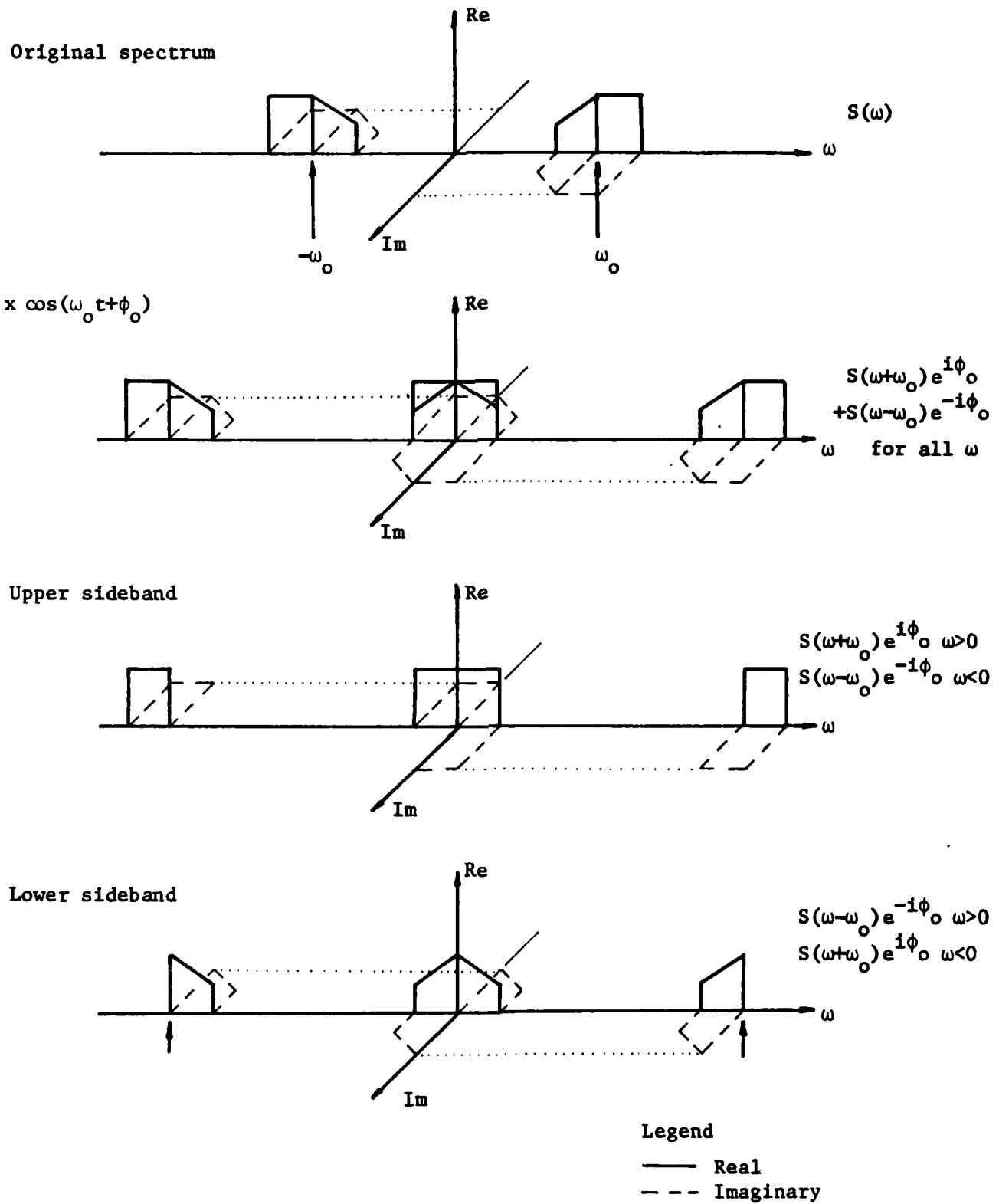
——— Real
— — — Imaginary

are shown graphically. (While examining this figure it is important to remember that, in general, spectra contain both positive and negative frequencies. For real signals (complex sense), the negative frequencies must be present to ensure that the imaginary parts of the Fourier integral will be zero.) From the figure it can be seen that components of the spectrum from above and below the LO frequency will add together when the signal is mixed with only a cosine signal. (This mode of mixing is referred to as double sideband mixing.) To avoid this overlap of upper and lower sidebands a single sideband mixer is used. This overlap of sidebands does not occur in the first mixing operation (which only mixed a cosine signal) because the signals with frequencies less than that of the first LO are removed (by filtering) before mixing. With a single sideband mixer, the signal is split into two channels, and a sine is mixed with one and a cosine wave with the other. The sine channel (i.e., the channel mixed with the sine wave), after mixing, is rotated by -90° of phase and added to the cosine channel. Their sum will contain only the. frequency components from above the LO frequency (Rogers, 1971). This sum is referred to as upper sideband (USB). If the 90° rotation is applied to the cosine channel, the sum will contain only frequency components from below the LO frequency. This sum is the lower sideband (LSB).

When single sideband mixers are used in the video con- verters, the output signal of the mixers for upper sideband

will be related to the input IF signal by

$$\bar{S}^u(\omega_v) = \bar{S}_m(\omega_v + \omega_I) \ e^{i\phi_I^u} \qquad ; \qquad \text{if } \omega_v > 0$$

and

$$\bar{S}^u(\omega_v) = \bar{S}_m(\omega_v - \omega_I) \ e^{-i\phi_I^u} \qquad ; \qquad \text{if } \omega_v < 0,$$

where $S^u(\omega_v)$ is the spectrum at video frequencies (i.e, the video spectrum), and ω_I and ϕ_I^u are the frequency and phase of the local oscillator in the video converter. We may combine the effects of the two local oscillators into a single expression. For upper sideband the relationship between the RF spectrum and the video spectrum will be

$$\bar{S}^u(\omega_v) = \bar{S}(\omega_v + \omega_1) \ e^{i\phi_1^u} \qquad ; \qquad \text{if } \omega_v > 0 \qquad (2.1.1.a)$$

and

$$\bar{S}^u(\omega_v) = \bar{S}(\omega_v - \omega_1) \ e^{-i\phi_1^u} \qquad ; \qquad \text{if } \omega_v < 0, \qquad (2.1.1.b)$$

where $\omega_1 = \omega_r + \omega_I$ and $\phi_1^u = \phi_r + \phi_I^u$. For the lower sideband, the relationship between the RF spectrum and the video spectrum will be

$$\bar{S}^\ell(\omega_v) = \bar{S}(\omega_v - \omega_1) \ e^{-i\phi_1^\ell} \qquad ; \qquad \text{if } \omega_v > 0 \qquad (2.1.2.a)$$

and

$$\bar{S}^\ell(\omega_v) = \bar{S}(\omega_v + \omega_1) \ e^{i\phi_1^\ell} \qquad ; \qquad \text{if } \omega_v < 0, \qquad (2.1.2.b)$$

where $\phi_1^\ell = \phi_r + \phi_I^\ell$ is the lower sideband phase shift. Ideally ϕ_1^ℓ will equal ϕ_1^u but this equality may not always hold (see Appendix B).

Before the video signals are recorded, they are hard-limited or clipped. The clipping is achieved by amplifying the signal to the saturation point of the amplifier, i.e. the output of the amplifier is plus or minus its saturation

voltage. This amplified signal is then uniformly sampled at the Nyquist rate for a 2 MHz bandwidth signal. The output of the clipper is either plus or minus one depending on whether or not the input voltage was positive or negative. In this mode a single bit on the tape can be used to represent the data sample.

Up until the clipping of the signal the operation of the receiver has only rotated, shifted in frequency, and added noise spectra to the spectrum of the signal from the radio source. The clipping operation will however change the shape of the spectrum of the signal and noise. These changes in the spectra will change the correlation function of the signals. The changes in the correlation function of a signal which has been clipped have been previously studied (Van Vleck and Middleton, 1966; Whitney, 1974). For a stationary process, the autocorrelation function, $A_c(\tau)$, of a clipped signal will be given by

$$A_c(\tau) = (2/\pi) \sin^{-1} A(\tau) ,$$

where $A(\tau)$ is the autocorrelation function of the original signal, and τ is the lag introduced for correlation.

The signal plus noise which is recorded on magnetic tape is bandlimited to the frequency range of 0-2 MHz (the actual recorded signals do not cut off instantaneously at 2 MHz nor do they have a zero frequency component; this DC component is filtered out before recording). The spectral components of the signals in this 0-2 MHz range are the same components that occurred in the spectrum of the signals from the radio source

in the range 0-2 MHz around their radio frequency (RF), (they

have however been amplified) with added noise which also came

predominantly came from 0-2 MHz band around the RF. The major

noise source is the first amplifier in the receiver chain.

Over the 2 MHz range at RF the signal and noise power have

constant amplitude and hence the video signals will also have

a constant power spectrum. A signal whose power spectrum is

constant in the range 0-2 MHz will have an autocorrelation

which is given by $\sin(2\pi B\tau)/(2\pi B\tau)$, where B is the bandwidth

(2 MHz in our case). The autocorrelation function of the

clipped signal will therefore be

$$A_C(\tau) = (2/\pi) \sin^{-1}\{\sin(2\pi B\tau)/(2\pi B\tau)\}$$

If the clipped signal is sampled at the Nyquist rate (1/2B

time interval between samples), we see that $A_C(n/2B)$ is zero

for all integral n and hence that the samples of the clipped

signal will be uncorrelated.

2.2 Cross spectrum phase and its relationship to group delay, phase delay and phase delay rate.

In Section 2.1 we examined the signals which are recorded

on magnetic tape at each radio telescope during a VLBI obser-

vation. In this section we will investigate the method used

to extract the values of the VLBI observables from these

recordings. The recorded data consist of three components:

the signals from the radio source (which will be present on

the recordings at each site); the noise, mainly from the

receivers at each site; and the pulses generated by the phase calibration system. The statistical properties of the recorded data were discussed in Section 2.1. The spectral components of both the radio source signal and the receiver noise should be Gaussianly distributed. For the radio sources observed in geodetic VLBI experiments, the power spectrum of the signals over the 2 MHz recorded bandwidth will be constant (except at the edges of the band where it rapidly, but not instantaneously, decreases). The power spectrum of the noise should also be constant over the 2 MHz recorded bandwidth.

If we were to cross correlate the tape recordings from each site, we would expect that the signals from the radio source would correlate and cause a peak in the correlation function at a lag between the tapes corresponding to the difference in the arrival times of the signals (provided the integrated signal strength is sufficently greater than the noise that this peak can be distinguished from the noise). (For the moment we will neglect the additional complication that the Earth is rotating and hence that the delay is not constant.) This attribute could be used to estimate the difference in arrival times of the signals from the radio source.   An alternative technique for estimating the delay would be to find the cross spectrum of the data recorded at each site and use the phase of the cross spectrum to estimate the delay: the phase of the cross spectrum divided by the angular frequency will be the delay.

There would appear to be a difference in information

content between these two estimates. If all the data are cross correlated, then only a single estimate of the delay is obtained. On the other hand, if the cross spectrum approach is used, then an estimate of the delay is obtained for each frequency at which the cross spectrum is computed. The reason for this apparent difference in information content is that we have been discussing two different types of delay. When the data are cross correlated and the delay determined by maximizing the cross correlation function we are measuring the difference in arrival time of packages of energy from the source (see, e.g., Jackson, 1975, pp 209-303). These packages of energy will propagate at a group velocity and the measured difference in arrival times will be a difference in group delays. When the cross spectrum approach is used, we are decomposing the energy packets into their individual "monochromatic" waves. These waves will propagate at their respective phase velocities. At each frequency the phase of the cross spectrum is a measure of the phase delay at that frequency. However, the phase can only be measured modulo 2π radians and hence from the cross spectrum we do not obtain unique estimates of the phase delays.

The apparent difference in information content of the two techniques can be resolved by noting that the group delay is the derivative of the phase with respect to angular frequency. Hence, the group delay which is obtained from cross correlating data, can be obtained from the cross spectrum results by determining the derivative of the phase of the cross spectrum

with respect to angular frequency. This derivative will be the group delay.

In practice the use of either of the techniques discussed above is complicated by the rotation of the earth and hence by the delay changing with time. To overcome this problem we could cross correlate the data from two tapes while running the tapes at different speeds however with digital recordings this procedure would be very difficult to implement. With the cross spectrum technique, the spectra of the data recorded at each site will be Doppler shifted if the stations are moving with respect to the source and hence the phase of the cross spectrum computed by Fourier transforming the data changes its meaning. If the Doppler shifts of the signals are known then the spectra of the data from each site can be corrected and a more useful cross spectrum constructed.

A method for overcoming the problems associated with the rotation of the earth and for determining the maximum likelihood estimates of the group delay, phase delay and phase delay rate when the delay is a function of time was developed by Rogers (1970). Rogers constructed a function, called the delay resolution function, which depended upon the cross spectral components of the recorded data determined from short segments of data and trial values of the phase delay, group delay, and the phase delay rate. The values of these latter two quanities which maximize the value of this function can be shown to be the maximum likelihood estimates of the VLBI observables (see Appendix A).

To overcome the problem associated with the rotation of the earth, we can segment the recorded data from each channel into short intervals (e.g., intervals of 4 μsec duration) and compute the cross spectra of these 4 μsec segments of data (sixteen data samples from each channel are used to determine the cross spectrum for that channel). Only 8 complex components of the cross spectrum are determined from the 4 μsec span of data (i.e., the cross spectrum is computed at eight equally spaced frequencies spanning the 2 MHz recorded bandwidth). Over such a 4 μsec interval, the change in phase of the cross spectrum due to the rotation of the earth will be small, and the computed cross-spectral components will approximate the cross spectrum which would have been obtained had the two sites been moving with equal velocity (see Appendix A for details). The cross-spectral components from all of the 4 μsec data segments, from all of the channels, could then be used in the delay resolution function to estimate the values of the VLBI observables. (This technique of using 4 μsec segments of data is not the actual technique used by the Mark. III correlator, but it will provide us with a method for study of the actual techniques used; see Appendix B for more details.)

In Appendix A we review the development of the delay resolution function and show that when the statistical properties of the signal and noise are independent of time and frequency, the maximization of the delay resolution function is the maximum likelihood estimator of the VLBI observables

and is equivalent to a least squares estimator. The data which would be used in the least squares estimator are the phases of the cross-spectral components computed at 7 frequencies from each 4 µsec data segment or from coherent averages of these cross-spectral phases from many 4 µsec data segments (see Equation (A.16)).

The delay resolution function $D(\tilde{\tau}_g, \tilde{\tau}_p)$ is given by Equation (A.10) (see Appendix A for details):

$$D(\tilde{\tau}_g, \tilde{\tau}_p) = \frac{1}{BT} \sum_{j=1}^{BT} \bar{X}_{1j}\bar{X}_{2j}^* \exp[-i(\omega_j-\omega_o)\tilde{\tau}_g - i\omega_j\tilde{\tau}_p\Delta t_j] \quad ,$$

where $\bar{X}_{1j}\bar{X}_{2j}^*$ are the components of the cross spectrum, at frequency ω_j, of the data recorded at each site at time Δt_j relative to the center of the observation (Δt_j refers to the epoch of the center of the short segment of data used to determine the cross spectrum), $\tilde{\tau}_g$ and $\tilde{\tau}_p$ are trial values of the group delay and phase delay rate which are varied until the magnitude of $D(\tau_g, \tilde{\tau}_p)$ is maximized. The phase of $D(\hat{\tau}_g, \hat{\tau}_p)$, where $\hat{\tau}_g$ and $\hat{\tau}_p$ are the values of τ_g and $\tilde{\tau}_p$ which maximize the magnitude of the delay resolution function, is the estimate of the visibility phase referred to the frequency ω_o. (The visiblity phase is an estimate of the cross-spectral phase at ω_o; see Appendix A for more detailed discussion.)

The input for the delay resolution is not the data recorded during the observation, but rather the components of the cross spectrum of the recorded data. We have discussed one method for obtaining these components: by cross correlating short segments of data and finding the Fourier transform

of the cross correlation functions from each data segment. In Appendix B we investigate the methods used by the Mark III correlator to generate the cross-spectral components from the recorded heterodyned signals.

The actual implementation of maximizing the delay resolution function is not as straightforward as the above equation (Equation A.10) implies. (The material discussed here is treated extensively in Appendix B; in the following discussion we will try to highlight the major points from this appendix.) The summation given in Equation (A.10) is two summations, one over time and the other over frequency. In the implementation of Equation (A.10) these summations are treated separately. The cross-spectral components within each frequency channel are coherently averaged over a period of time referred to as an accumulation period (usually ≃2 sec). The coherent average corresponds to initially carrying out the time summation in sections, each of a sufficiently short duration that the _a priori_ phase-delay rate can be used in the summation rather than the trial values of the phase-delay rate. The trial values of the phase-delay rate are used to sum the coherent averages. The frequency summation is also separated into two parts (but for reasons different from those for separating the time summation). The frequencies of the cross-spectral components are clustered in groups in each 2 MHz bandwidth channel, with the groups separated from one another by several tens to several hundreds of megaHertz. The group delay will be more precisely estimated the using the cross-spectral phases from

widely spaced frequencies. Hence, two group delays are estimated -- a "multiband" group delay, which is estimated from the slope of the "video DC" phases from the widely spaced 2 MHz channels and a "single band" group delay which is estimated from the average of the slopes of the cross-spectral phases for the individual 2 MHz channels. (The "video DC" phases are the estimates of the cross spectral phase at $\omega_v=0$ (video DC), which would be obtained from least-squares estimates of a slope and an intercept (at $\omega_v=0$), from the cross-spectral phases at 14 or 7 discrete video frequencies (depending upon whether dual or single-sideband data were recorded, respectively.) The slope is the singleband delay, and the intercept is the video DC phase; see Section B.2, for more details.) The singleband group delay estimate is much less precise then the multiband group delay estimate.

Thus, there are three quanities for which the delay resolution function is maximized: the multiband group, the singleband group delay and the phase delay rate. The estimate of the visibility phase is the phase of the delay resolution function after the estimates of the above three quantities are substiuted.

The estimate of the multiband group delay is, in general, not unique. The frequency spacings of the widely separated channels are chosen to be multiples of the minimum frequency spacing (this type of frequency spacing ensures low "sidelobe" levels in the delay resolution function) and hence it is possible to find equally good representations of the observed

(video DC) phases by shifting the group delay by $2\pi/\Delta\omega_{min}$ sec,
where $\Delta\omega_{min}$ is the minimum (angular) frequency spacing. This
change in the group delay, called the group-delay ambiguity,
will change the estimates of the video DC phases of all of the
(widely spaced) frequency channels, by exact multiples of 2π
radians. Hence, the difference between the observed video DC
phases and the estimates of their values calculated from the
group delay, visibility phase and phase-delay rate will be
unaffected. (Remember that we have no way, other than from a
priori information, of knowing which multiple of 2π radians
the phases in each channel should be located on; see Appendix
A for more details.)

The signals received at each site may be recorded in one
of two ways. In geodetic Mark III experiments, before June 24,
1981, the signals were recorded with dual sidebands, i.e.,
both the upper and the lower sidebands were recorded. This
recording technique offers the advantage that some of the
systematic errors which are introduced by the approximations
made in the correlator algorithms, cancel with this mode of
recording (see Appendix B for details). The disadvantage of
this recording mode is that each channel requires a pair of
tape tracks, limiting the maximum number of frequency channels
which can be recorded in Mode B to seven. (Mode B is a
recording mode for which 14 tracks on the tape are recorded
during the forward motion of the tape and the other fourteen
are recorded in the reverse motion of the tape (Rogers et al.,
1983)). When only single sideband data are recorded (i.e.,

when only one output from each video converter is sent to the tape recorder), 14 frequency channels can be recorded (in Mode B). These additional frequency channels can be used to reduce the sidelobe levels in the delay resolution function (see Whitney, 1974, and Robertson and Cater, 1983, for diagrams of the delay resolution function).

When the algorithms which are used to generate the cross-spectral components were derived in Appendix B, several approximations were made. The effects of these approximations are also studied in this appendix. We found that when the phase delay rate is greater than 10^{-9} sec/sec, the biases introduced into the visibility phase should be less than 10°, when only single sideband data are used. (The phase-delay rate will exceed 10^{-9} sec/sec most of the time on baselines greater than 10 km long if ω_o >2.3 GHz.) When upper and lower sideband data are used, in equal proportions, the visibility phase errors should be zero. Only one approximation (accumulating the cross correlation function with an erroneous phase-delay rate) was found which could affect the multiband group delay. Even in the extreme situation studied the error in the multiband delay was only 20 psec (≈0.7 cm) at X-band. In general, we would expect this approximation to introduce errors in the multiband group delay of less than one picosecond (≈0.3 mm).

There are of course many more components in the Mark III system which could affect the VLBI measurements. It is beyond the scope of this thesis to investigate all of the possible

interactions between these components and the VLBI measurements. In Chapter 3 where we investigate the quality of the VLBI measurements, we will discuss some of those errors due to equipment limitations rather than to the approximations made in applying the estimation procedures.

Before finishing our discussion of the VLBI observables we need to discuss one final aspect of the measurement technique -- the phase-calibration system.

2.3 The phase-calibration system

In Section 2.1 we discussed the system components which operate on the signal from the radio source as the signal propagates through the receiver to the tape recording. These components (waveguides, mixers, amplifiers, cables, filters) will greatly affect the phase of the signal, e.g., 1.5 m of waveguide, with a cutoff wavelength of 5.7 cm, will have a phase delay of 3.9 nsec at 8.3 GHz which corresponds to 32.4 cycles (see Section 3.1). If the propagation properties of the receiver were time independent, then these delays through the receiver would pose no serious problem. (The propagation delays would be constant and could not be differentiated from the epoch offset of the hydrogen maser.) Unfortunately, the properties of the components in the receiver are not time independent. There are two different types of phase variation which can occur in the receiver. These variations are changes in the local oscillator phases which will add direct-

ly to the signal phases (see Equations (2.1.1) and (2.1.2)),
and variations in the propagation delays. Both of these
types of change can be monitored by the phase-calibration
system.

The phase-calibration system consists of pulses which
are injected into the receiver every microsecond (see Figure
2.1.2). In the frequency domain the pulses will form a
"comb" of Dirac delta function "rails" with 1 MHz spacings
between the rails. The pulses are of finite duration ($\approx$50
psec) and hence the comb will extend to approximately the
inverse of the pulse duration ($\approx$20 GHz), (Rogers, 1980).

In the frequency domain, there will be two phase-cali-
bration rails appearing in each 2 MHz channel. The frequency
boundaries of the channels are set such that these rails will
appear at 10 kHz and at 1010 kHz for upper sideband (USB)
channels. The rails will appear at 990 kHz and 1990 kHz in
the lower sideband (LSB) channel when both sidebands are
being recored. If only LSB is being recorded, the
calibration rails would be located at 10 kHz and 1010 kHz.
The frequency boundaries of each channel are determined by
the sum of the local-oscillator frequencies which are mixed
with signals from the radio source, and the recorded
bandwidth.

The phase-calibration pulses will propagate through the
receiver and their spectral components will undergo the same
phase shifts as the spectral components of the signals from
the radio source at the corresponding frequencies. When the

phase-calibration phases are subracted from the video DC

phases (see Section 2.2), the calibrated phases become refer-

red to the phase of the pulses at injection into the pre-

amplifier of the front end.

The phase of only one of the calibration rails is

extracted during correlation (the rail occurring near zero

frequency (DC) in the video spectrum, i.e. usually 10 kHz).

The major reason for extracting a single calibration phase

concerns the use of dual sideband recordings. We discussed

in Section 2.2, the two types of group delay which are esti-

mated during processing of VLBI data, the multiband and

single band group delays. The extraction of a second cali-

bration phase would predominantly affect the estimation of

the singleband group delay since the second calibration phase

would calibrate phase changes over the 2 MHz recorded band-

width. When the Mark III system was being designed the

recording made was expected to be both upper and lower side-

band so that each channel in the synthesized band would have

effectively a 4 MHz bandwidth with the reference frequency

(i.e., the radio frequency corresponding to $\omega_v=0$), for the

channel in the center of the 4 MHz bandwidth. Because the

multiband group delay (this is the delay used in geodetic

data processing), is estimated from the phases at the refer-

ence frequency, any error in the derivative of the phase with

respect to frequency across the 4 MHz (USB + LSB) channel

will not affect these phases (because of symmetry) and hence

will not affect the estimates of the multiband group delay

(see Appendix B for more details). Consequently there was no need to determine the phase of the second calibration rail. However, in recent Mark III VLBI experiments (since June 24, 1981) only single sideband recordings have been used. There were two reasons for the change in recording mode. Firstly, the assumption of the insensitivity of the phase at the reference frequency to the derivative of the phase with respect to frequency across the 4 MHz bandwidth is only valid if the same weight is given to the data from each sideband. This (weight) equality was often not satisfied because of tape reading errors; hence many of the advantages of using both sidebands were compromised. Secondly, the tape density upgrade (Hinteregger, 1980; Hinteregger, 1982) was taking longer than expected and there was a desire for more frequency channels in the synthesized bandwidth. (The additional frequency channels were needed to allow a wider bandwidth to be spanned while having a smaller minimum frequency spacing (this smaller spacing increases the group-delay ambiguity; see Appendix A). The additional frequencies also made the group delays much less sensitive to phase errors in individual channels (see Section 3.1). With single sideband recordings, the extraction of a second calibration phase would reduce errors due to changes in the system delay (over the 2 MHz recorded bandwidth). A 1 nsec change in the delay through the system will introduce an error in the phase of the reference frequency of $0°36$. (This number is easily calculated by noting that the reference frequency is 1 MHz

away from the center of the recorded channel and hence a change in the delay of $\Delta\tau_c$ will lead to a phase error of $\Delta\tau_c \times 1$ MHz; see, also, discussion associated with Figure B.3.1). In Section 3.1 (Figure 3.1.5), we show evidence which indicates that such changes may be possible at some antennas (although the phenomenon shown in Figure 3.1.5 could be explained by other mechanisms as well). At this time, the correlator modules would need to be redesigned to allow both calibration phases to be extracted. When the tape density upgrade is completed, dual sideband observations should be adopted again, in order to minimize errors due to system delay changes.

The extraction of the calibration phases and amplitudes is carried out in the correlator module before the data streams are cross correlated. Ideally, to extract the phase-calibration phase we would multiply each data stream by $\cos(\omega_c t)$ and $\sin(\omega_c t)$, where ω_c is the frequency of the calibration rail, and then low pass filter the output. The calibration phase would then be $\tan^{-1}(C_s/C_c)$, where C_s and C_c are the correlation coefficents from the sine and cosine multiplications. The phase calibration amplitude is calculated from $\sqrt{C_s^2 + C_c^2}$. This amplitude is usually expressed in units of 10^{-4} of the (on-source) system temperature, and hence is a function of the power in the calibration pulses and the system power. The calibration amplitudes are not normally used, but we will see in Section 3.2 that these amplitudes can provide a useful diagnostic tool.

The multiplication by sine and cosine functions is carried out in the correlator, but the cosine and sine functions actually used are two-level approximations, i.e. they are either plus or minus one depending upon the argument of the function. During a two second accumulation period the cosine and sine will rotate through 20,000 cycles and the errors introduced by using the two-level cosine and sine function will cancel to a very high degree. (The error in the calibration phase introduced by the use of the two-level sine/cosine approximation will be $\approx 5 \times 10^{-5}$ rad; see discussions on the three-level sine/cosine function in Appendix B.)

The use of a two-level cosine/sine function is not without its problems however. The signal-to-noise ratio (SNR) of the phase determination is lowered by $4/(\pi\sqrt{2}) \approx 0.90$ (this factor is the ratio of the Fourier series coefficient of the two-level cosine function to the Fourier series coefficient for a pure cosine function divided by the expected increase in noise level, $\sqrt{2}$). This reduction in SNR is not a serious problem because the effective temperature of the phase-calibration rails at 10 GHz is $\approx 5 \times 10^4$ K (Mark III Documentation, 1980, p. DC-8) and after a 1 second integration time the SNR will be 300 if the system temperature is 167 K. (The 5×10^4 K temperature of the calibration rail corresponds to the phase calibration pulse power being $\approx 1\%$ of the total system power in a 2 MHz bandwidth.) With such large SNRs a reduction of 10% in the SNR is not a serious problem.

The second, and more serious, disadvantage of using the two-level sine/cosine function is the harmonics which are generated by this type of function. Since the two level sine/cosine function is periodic, it can be expanded in Fourier series with the fundamental frequency being the calibration frequency at video, i.e. 10 kHz. The Fourier coefficients decay as $1/n$, where n is the harmonic number. These harmonic terms should cause no serious problems unless there are non-random signals at these frequencies. There should not be any non-random signals at these frequencies, except for the 101st harmonic which corresponds to the frequency of the second phase calibration rail in the 2 MHz bandwidth. Approximately 1% of the signal in this second rail will be included in the computation of the calibration phase. This leakage of the second calibration rail into the first calibration rail could introduce errors of order 0.6 (0.01 rad) in the estimated calibration phases. These errors should be kept in mind as the precision of VLBI observations improves. At this time there seems to be more serious problems which need solving (see Chapter 3).

The phase-calibration pulses provide a series of time "ticks" at the injection point, to which the signal arrival times can be referred, i.e. the calibrated cross-spectral phases are the phase differences between the spectra of the signals at the injection points of the pulses, at the two sites in the interferometer, referred to the phases of the spectral components of the calibration pulses, again at the

injection points at the two sites. (See discussion in
Appendix B concerning the singleband delay, Equation
(B.3.3).) Ideally, we would like the signal arrival times
referred to the time "ticks" generated by the hydrogen maser
frequency standard which serves as the master clock at each
site. If the propagation delay through the cable which
carries the 5 MHz signal used to generate the phase-calibra-
tion pulses remains constant, then there would be a constant
offset between the hydrogen maser time ticks and the time
ticks provided by the phase-calibration pulses. The propaga-
tion delay through this cable is, however, a function of
temperature and tension (and possibly other time-dependent
quantities). To ensure that this offset is always known, the
delay through the cable carrying the 5 MHz signal is moni-
tored using a cable-calibration system. The method used to
measure the cable delay relies on reflecting a modulated
version of the 5 MHz signal from the cable/pulse generator
connector back down the cable. (The modulation allows this
reflection to be separated from other possible reflections of.
the 5 MHz signal.) The phase difference between the outgoing
and incoming signals measures the changes in the propagation
delay through the cable. Except for clock synchronization
experiments (Clark et al., 1979), there is no need to measure
the absolute length of the cable.

Before completing this discussion of the phase calibra-
tion system we should discuss one possible problem with this
system. The phase-calibration system injects pulses into the

receiver which will be almost identical in shape and spacing at each site. There should be a very high correlation between the phase-calibration signals when the recordings from each site are cross correlated. (The phase-calibration signals are not pulses on the recordings because the signals have been heterodyned to video frequencies (see Section 2.1).) The VLBI estimation algorithm relies on the correlation between the signals from the radio source. Will this additional correlated signal affect the measurements? In general, there should not be any major effect because of the difference in fringe rate (i.e., the rate of change of phase; see also discussions on maximizing the delay resolution function, Appendix B.3), between the phase calibration pulses and the signals from the radio source. The correlation between the phase-calibration pulses will quickly dissipate as the delay between the tapes is changed and the rotation corrections are applied during correlation of the tapes. We should notice that observations with small fringe rates will be most affected by the correlation between the phase-calibration pulses.

We may calculate the expected magnitude of the contribution of the calibration pulse's cross correlation by summing the cross-spectral components of the calibration pulses, i.e. we may decompose the cross spectra of the recorded signals into contributions from the signal (from the radio source), the noise, and the phase-calibration signals. The cross correlation between the calibration pulses will introduce an

additional phasor to the signal cross-spectral phasor. As
the cross-spectral components are counter-rotated, to account
for the changing phase of the signal part of the cross
spectrum, this additional phasor (due to calibration pulses)
will also be rotated. However, the phase of the cross-
spectral components due to the calibration signals will be
constant (almost) before the rotation and, hence, after the
rotation, they will no longer align. We may calculate the
expected amplitude of the summed, calibration pulse cross
spectral components, by replacing the summation of these
contributions to the cross spectrum, with an integration (see
Equation (A.15) for a discussion of the coherent sum of the
cross-spectral components). This amplitude, after T seconds
of integration, will be

$$A_{cal} = \left| \int_0^T A_o \exp[-i\omega_r t_{ap} t] \, dt \right| \qquad (2.3.1)$$

where ω_r is the (angular) RF of the video channel before
heterodyning, A_o is the amplitude of the calibration phasor
($\approx 1\%$ of the total system power, i.e. $A_o \approx 0.01$), and $\dot{\tau}_{ap}$ is
the a priori phase-delay rate. The integration of Equation
(2.3.1) yields

$$A_{cal} = 2A_o \sin^2(\omega_r t_{ap} T/2)/(\omega_r t_{ap}) \quad . \qquad (2.3.2)$$

The integration of the signal term over the same interval of
time, assuming that $\dot{\tau}_{ap}$ is close to the actual phase delay
rate (see Appendix B for details), will be (relative to the

system power)

$$A_{signal} = \sqrt{Ta_1 Ta_2 / (Ts_1 Ts_2)} \; T \qquad\qquad (2.3.3)$$

We may now calculate the magnitude of the error in the cross-spectral phases due to cross correlation of the phase calibration pulses. For an observing frequency of 2.3 GHz, $(\omega_r = 14.5 \times 10^9$ rad/sec), $\dot{\tau}_{ap} = 100$ psec/sec (small value even for baselines less than 1 km), $T=100$ sec, $A_o = 0.01$, and $\sqrt{Ta_1 Ta_2 / (Ts_1 Ts_2)} = 10^{-4}$, the ratio of the amplitudes of the integrated calibration phasor to the signal phasor would be $A_{cal}/A_{signal} = 0.014$. The phase calibration cross correlation in this case could lead to phase errors of up to $0\overset{o}{.}8$ ($=0.014$ rad). This corruption of the signal by the phase-calibration pulses is a problem for observations on short baselines (<1 km). The problem can be solved by offsetting the phase-calibration rails at one site. However the current Mark III correlator cannot process data recorded in this manner.

We have now completed a basic discussion of the VLBI observables and the techniques used to estimate their values. The discussion in Appendix A has presented a variety of methods of visualizing and studying the estimation of the values of the VLBI observables. We have applied some of these methods in Appendix B to study some of the limitations of the methods used to estimate the values of the observables. We will now investigate other procedures which can be used to assess the quality of the VLBI measurements.

3. The Quality of the Observations

In Chapter 2 we studied the methods used to estimate the group delay, phase delay and phase-delay rate from signals recorded at two radio telescopes. In this chapter we will investigate the quality of these measurements. We have already seen that the observed values of the group delay, phase delay and phase delay rate are related to the radio telescope's position and the radio source's position. In Chapter 5 we will study in more detail the relationship between these quantities and the observations. For the moment the exact relationships are not important. Given such a relationship, one method of checking the quality of the observations would be to compare them with estimates of their values calculated from such a model.

There are, however, some major problems with this technique of studying observation quality. Firstly, this technique does not allow the errors in the observations to be separated from the inadequacies of the models. Secondly, because the models depend on parameters such as station and source positions, any observational errors will likely be at least partially absorbed into the estimates of these parameters. (Ideally, we would like to check the observation quality at the subcentimeter level. In order to achieve this level of uncertainty in the model values we would need to estimate corrections to the model parameters, based on the observations themselves.) If the observational errors were

independent and Gaussianly distributed then we could easily compute the expected reduction in the magnitude of the postfit residuals (i.e., the differences between the observations and the estimates of their values from the model, after parameters have been estimated from the observations), due to some of the errors in the observations being partially absorbed into the estimated parameters (see Appendix E). Unfortunately, we are interested in possible systematic errors in the observations and the statistical arguments which apply to "random" errors are likely to be erroneous when applied to systematic errors. In particular, one question we would like to answer is: what are the fractions of the systematic errors which are absorbed into the estimates of the station and source positions? The analysis of postfit residuals will never fully answer this question.

How then are we going to study the quality of the observations? A VLBI observation is very redundant and we will use this redundancy to study quality. The number of bits recorded at each radio telescope during a 100 second observation is typically 3.2×10^9. These bits are reduced to three numbers which are used for geodetic data analysis - the multiband group delay, phase-delay rate and phase delay (the visibility phase divided by ω_o; see Appendix A).

A review of the methods used to achieve this reduction in information volume will reveal several methods which may be used to study the quality of the observations. We commence by studying the behavior of the "residual phases" from each

frequency channel. The residual phases are the differences between the coherently averaged cross-spectral phases in a channel and the estimate of the cross-spectral phase for that channel, calculated from the group delay and the visibility phase (see discussion associated with Figure 3.1.1). If more than two channels are used in the synthesized band then we may use the residual phases to study how well the phases in each channel are modelled by the group delay and the visibility phase. We will study the residual phases in Section 3.1. After examining the residual phases, we will examine, in Section 3.2, the consistency of the phase and cable calibrations. These studies will aid in interpreting the behavior of the residual phases. Finally, in Section 3.3, we will investigate the sum of delays around a triplet of baselines as a measure of quality.

Many of the studies of the quality of the observations were motivated by early attempts to predict phase-delay ambiguities from the group delays. These prediction techniques, which probably provide one of the most stringent quality checks of the VLBI observations, will be discussed in detail in Chapter 4. In the following sections we will attempt to outline the properties of each of the quality checks. Specifically, we will be most interested in the limitations of each of the checks, i.e. the nature of errors which a specific check can or cannot measure. For each type of quality check we will give specific examples. We will not, however, be able to present all of the results thus far obtained. Instead, we

will compile the results of the quality checks and make some
conclusions about the performance and the limitations of
currently available Mark III systems (Section 4.4).

Some checking of the quality of the observations is
carried out before the data analysis stage. The program FRNGE
checks a number of error conditions during correlation, e.g.,
non-detection of the phase calibration signal, loss of time
synchronization while reading the data tapes, and low correla-
tion amplitudes, are some of the checks which are made.
Observations which fail these checks are flagged. These
observations will be shown with lower case letters when data
are plotted.

3.1 Residual phases

In Chapter 2 we investigated the methods used to estimate
the values of the VLBI observables from the clipped signals
recorded at each site during an observation. The data
recorded in each frequency channel, at each site, are cross
correlated in segments. These segments, referred to as
accumulation periods, are typically of 2 seconds duration.
From these complex cross correlation functions, cross spectral
components are computed at 7 discrete frequencies in each 2
MHz bandwidth channel. In Appendix A, we show that the esti-
mates of group delay, phase-delay rate and visibilty phase are
obtained by estimating the parameters of surface, which is
linear in both frequency and time, from the cross-spectral

phases. (This estimation process is performed by maximizing the delay resolution function. Also, separate group delays are used to model the phase variations across the 2 MHz channels and across the synthesized band (see Appendix B for details), the first of these delays being referred to as the singleband group delay and the second as the multiband group delay.)

The observed cross spectral phase for each accumulation period and frequency component will not exactly match the phase calculated from the group delay, phase delay rate and visibility phase. We would expect these differences to be in accord with the statistics of the signals and the noise. It is these differences we wish to study.

In studying these differences we are, however, faced with a dilemma. For a four-channel, dual sideband observation of 100 seconds duration, there will be 2800 phase differences (50 accumulation periods x 4 channels x 14 cross spectral frequencies). In a typical multi-antenna intercontinental VLBI experiment there will usually be over 1000 such observations per day. Even if the phase differences are stored as single precision numbers, one day of data would take 11.2 Mbytes of storage. Detailed analysis of this volume of data is impractical (with currently available computers).

In this section we will study averages of these phase differences. In this way we can reduce the volume of data to be analyzed to tractable size. These averages will be taken over time and frequency within each channel (both upper and

lower sidebands). For a four-channel dual-sideband observa-
tion, the 2800 phase differences are reduced to four phase
differences, one phase difference for each frequency channel
in the synthesized band. We will refer to these averaged
phase differences as "residual phases". These phases are
shown schematically in Figure 3.1.1. The averaging is carried
out in the program "FRNGE" which estimates the group delay,
phase delay rate and visibility phase for each observation
(see Appendix B). The residual phases are stored in the
geodetic data base (Ma, 1979) and hence can be easily accessed
and analyzed.

For the statistics of the signals and the noise sources
given in Chapter 2, we would expect the residual phases to be
zero mean, Gaussianly distributed, random quantities.

To introduce the discussion on residual phases we show in
Figures 3.1.2 and 3.1.3 plots of residual phases as a function
of time from observations of the radio source 3C 345 during
the June 1981 experiment. These plots show the residual
phases in each of the four frequency channels used in these
observations. The residual phase in Channel 1 (frequency
8.34099 GHz) is shown with a letter which represents the
source being observed (these plots show the residual phases
for a single source 3C345, denoted by the letter E). The
residual phases in the remaining channels are denoted by
numbers (i.e., the numerals 2, 3, and 4 denote the 8.24099
GHz, 8.49099 GHz, 8.54099 GHz channels, respectively).

At approximately 7:00 hr UT, typical error bars, based on

Figure 3.1.1 Graphical representation of the residual phases
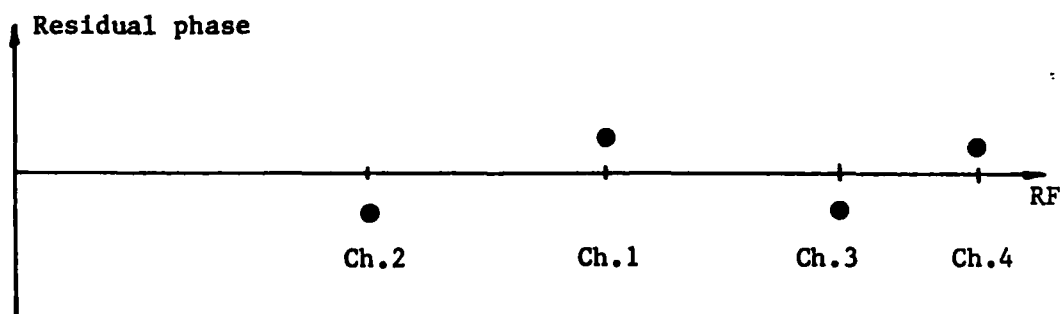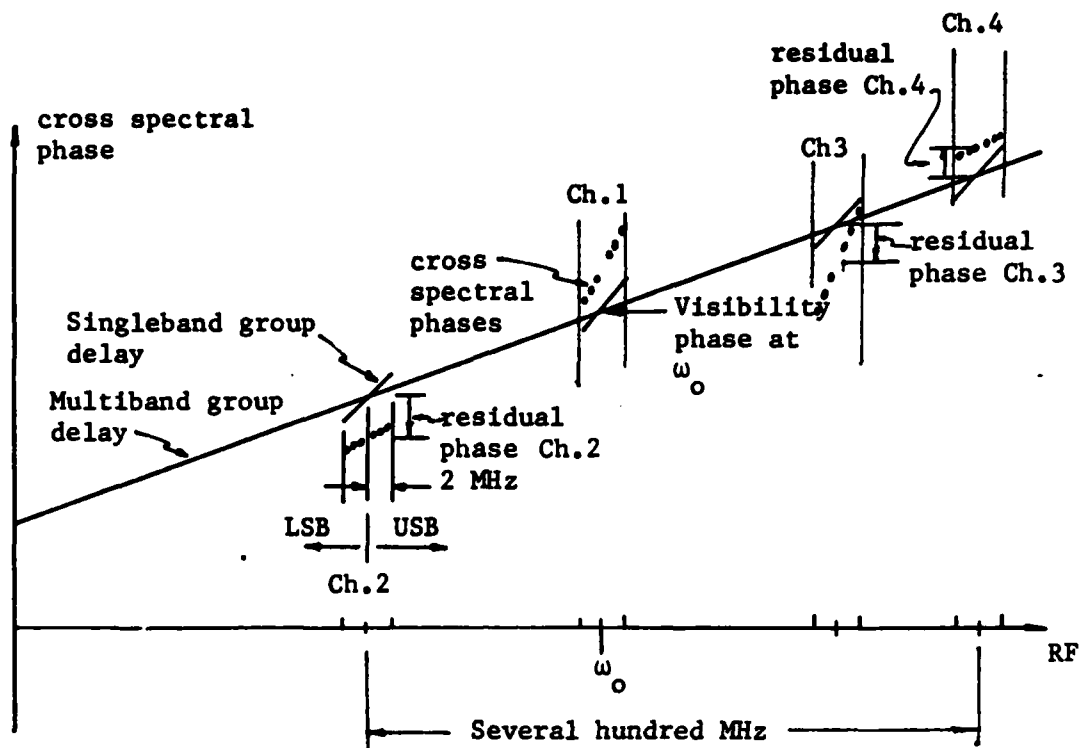
Figure 3.1.2 Residual phases for the Haystack to Ft. Davis baseline
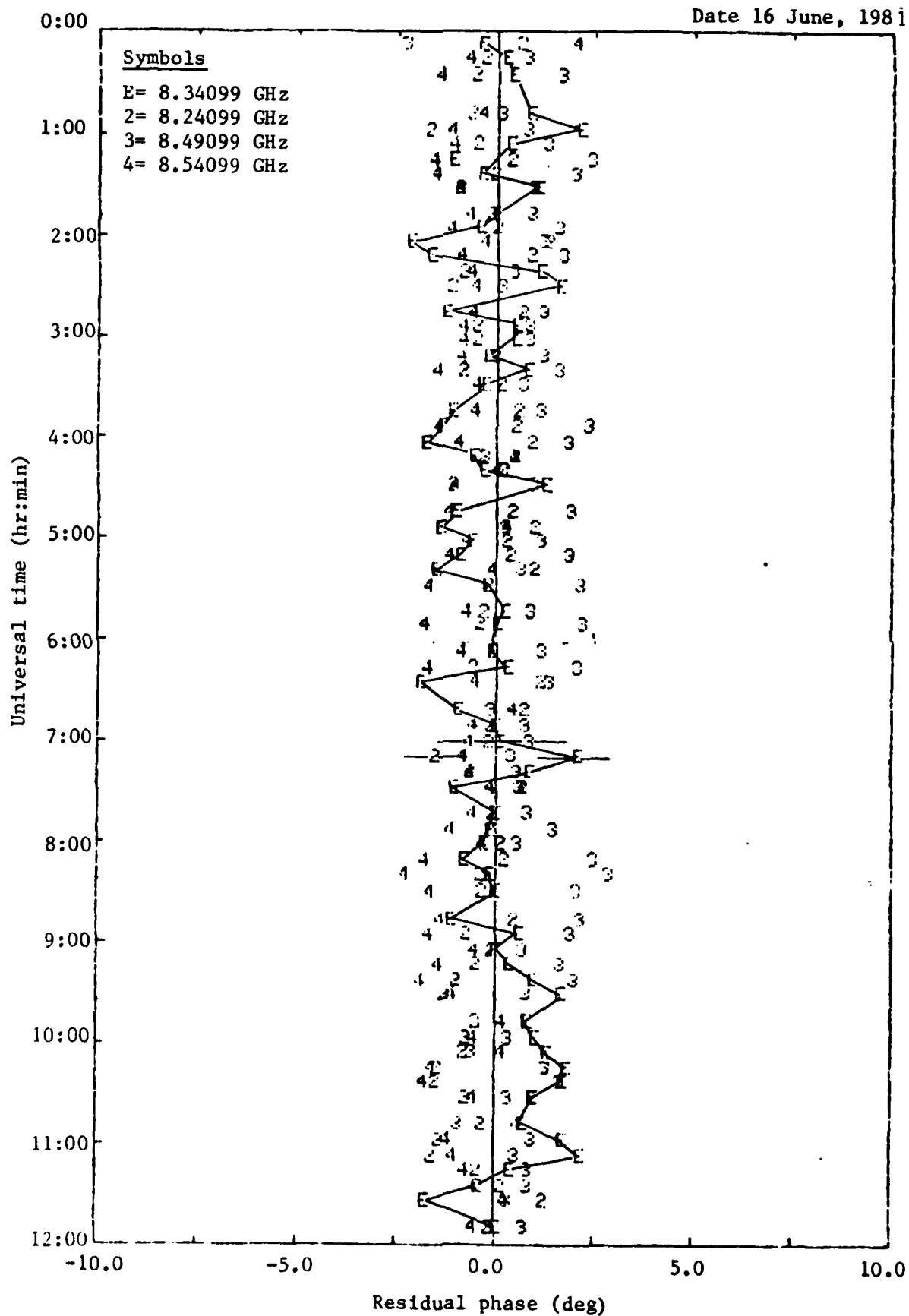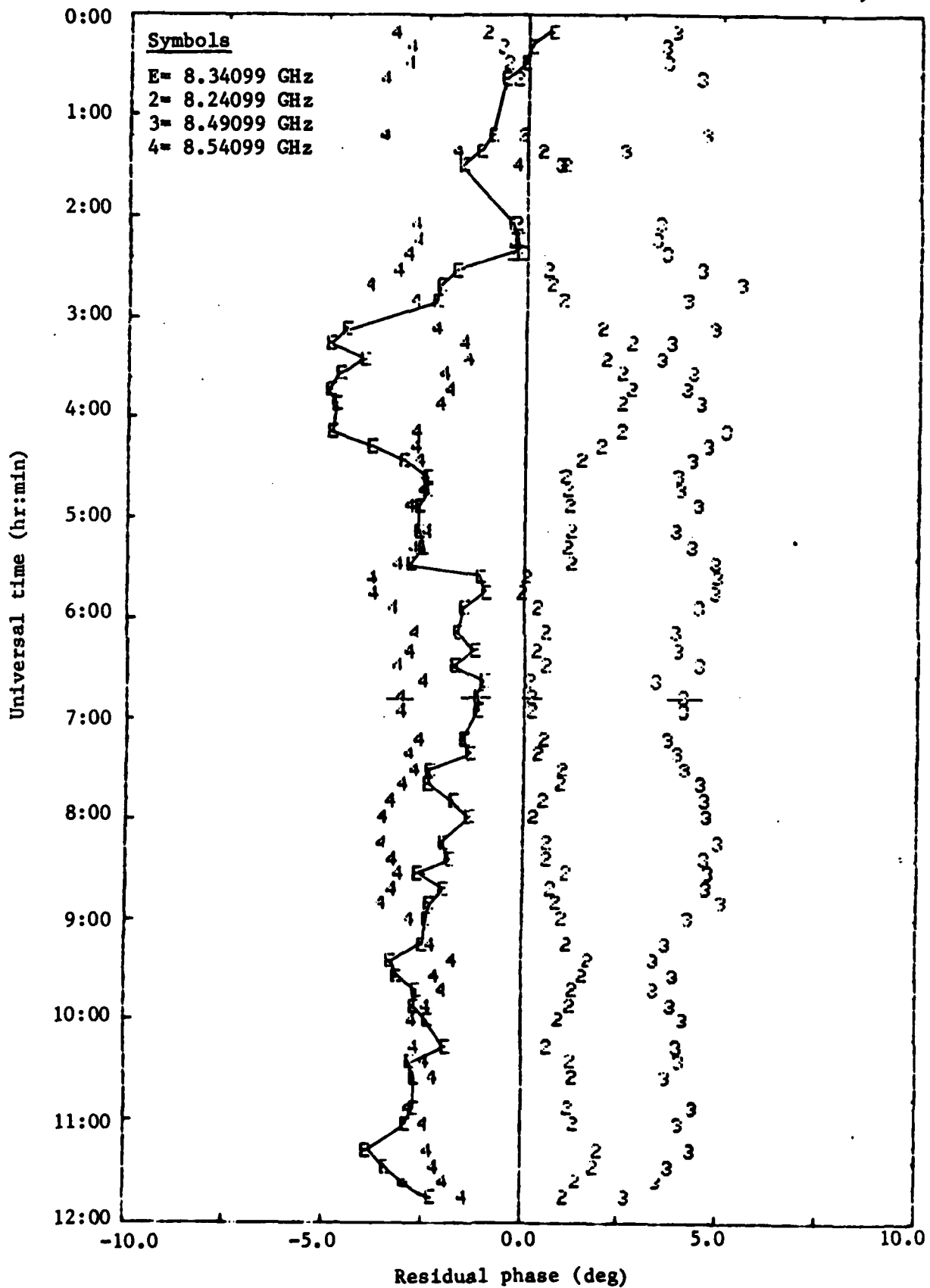

Date 16 June, 1981

Figure 3.1.3 Residual phases from the Haystack to Owens Valley baseline

Date 16 June, 1981

the signal-to-noise ratio (SNR) of these observations are shown. The error bars take into account that residual phases are the differences between the observed phases in each channel and the calculated values for these phases, after estimating the multiband group delay and visibility phase. The uncertainties of the residual phases in the frequency channels at the edges of the synthesized band are approximately 20% smaller than those in the center band, because errors (either random or systematic) in these outer channels tend to be absorbed into the estimate of the group delay (see Appendix E for a discussion on the methods used to calculate these uncertainties).

Figure 3.1.2 shows approximately the behavior of the residual phases we would expect, i.e. zero mean and random scatter, although the scatter of the values is approximately 1.5 times larger than we would expect given the signal-to-noise ratio (SNR) of these observations (we discuss this topic in more detail later in this section). Figure 3.1.3 shows a very different character. The residual phases do not have zero means and there are systematic variations with time. We believe we understand the origin of most of these systematic variations. We will discuss this topic in Section 3.2 when we investigate the behavior of the phase calibration system. We will concentrate in this section on the mean values of the residual phases and their scatter about these mean values.

We hope to answer several questions by studying the behavior of the residual phases: a) what is the origin of

the nonlinear behavior of the visibility phase with frequency?
b) what is the long term stability of the mean values of the
residual phases? and c) what can we deduce about the uncer-
tainties of the group and phase delay measurements based on
the scatter of the residual phases?

The non-zero mean values of the residual phases indicate
that the Mark III system is dispersive. We would expect the
Mark III receivers to be dispersive, but, we should also
expect the phase calibration system to remove most of this
dispersion (we will discuss below the meaning of "most"
shortly), because the phase calibration pulses travel with the
signals through nearly all of the system. The parts of the
receiving system that are not calibrated by the phase cali-
bration system are the feed horn and the waveguide up to the
point where the phase calibration pulses are injected. After
this injection point any dispersions in the system should be
calibrated and removed by the phase calibration system.

So let us start our investigations of the origin of the
dispersion with a discussion of the dispersive properties of
feed horns and waveguides. Propagation through waveguides is
discussed in most books on electromagnetic radiation. The
phase velocity of a wave propagating through a hollow cylin-
drical waveguide with cut-off frequency ω_λ (i.e., the lowest
frequency wave which will still propagate through the wave-
guide) is (Jackson, 1975, p. 344, MKS units),

$$v_p = (1/\sqrt{\mu\epsilon}) \left(1-(\omega_\lambda/\omega)^2\right)^{-1/2}$$

where ϵ and μ are the permittivity and permeability of the material inside the waveguide (in our case this material is air and hence $1/\sqrt{\epsilon\mu} = c$, the free space velocity of propagation), and ω is the frequency of the propagating wave. The change in phase of a signal propagating through a waveguide with a free space length of τ_o will be

$$\Delta\phi_w = \tau_o \omega \left(1-(\omega_\lambda/\omega)^2\right)^{1/2} . \qquad (3.1.1)$$

The group delay, $\Delta\tau_w$, for propagation through the waveguide may be obtained by differentiating Equation (3.1.1) with respect to angular frequency. This differentiation yields

$$\Delta\tau_w = \tau_o \left(1-(\omega_\lambda/\omega)^2\right)^{-1/2} . \qquad (3.1.2)$$

Clearly the propagation through the waveguide is dispersive. But is this dispersion sufficiently large to cause the offsets in the residual phases seen in Figure 3.1.3? To investigate this possibility, the phase changes due to the propagation of a signal through a 10' long waveguide ($\tau_o \simeq 10$ nsec) with a cutoff frequency of 5.3 GHz (cutoff wavelength of $\simeq 5.7$ cm) were calculated. (This cutoff frequency is typical for the X-band waveguides in the Mark III receivers (Rogers, 1983, private communication).) In Figure 3.1.4 we show these phases computed at the frequencies used in the June, 1981 experiment. In the lower part of this figure we show the residual phases which would be seen after the group delay and visibility phase had been estimated. (We have assumed that this waveguide precedes the injection point of the phase

Figure 3.1.4 Waveguide dispersion and expected residual phases
(10' of waveguide with $\omega_\lambda = 5.3$ GHz)

Legend

● Expected residual
  phases

■ Residual phases from
  intercept and slope
  at 8.34 GHz
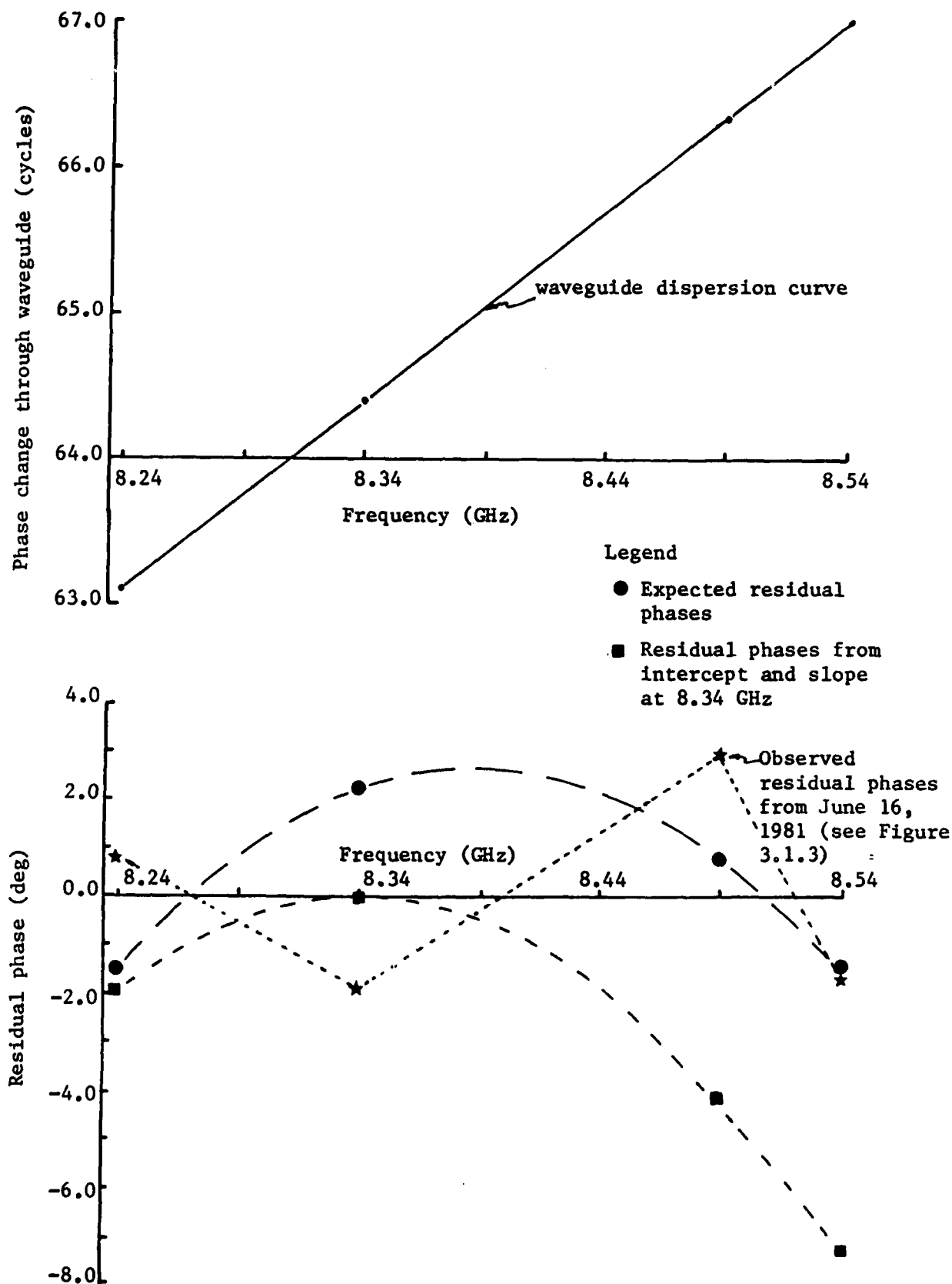
calibration pulses.)   We can clearly see in Figure 3.1.4 that

the waveguide dispersion cannot explain the residual phase

pattern (as a function of frequency) observed on the Haystack

to Owens Valley baseline.

(We should note that the phase delay through this wave-

guige computed from the least squares fit to the calculated

phases is 7.8 nsec while the group delay is 12.9 nsec. There

is a large offset between the group and phase delay. This

offset will need to be accounted for when the ionospheric

delay corrections are computed from group and phase delays (we

will discuss this topic in detail in Section 4.2.2).)

Calculations of feed horn dispersions yield similar

results to those shown in Figure 3.1.4.   The change in phase

due to the dispersion may be large (over one cycle for the

waveguide case), but the residual phases will be small because

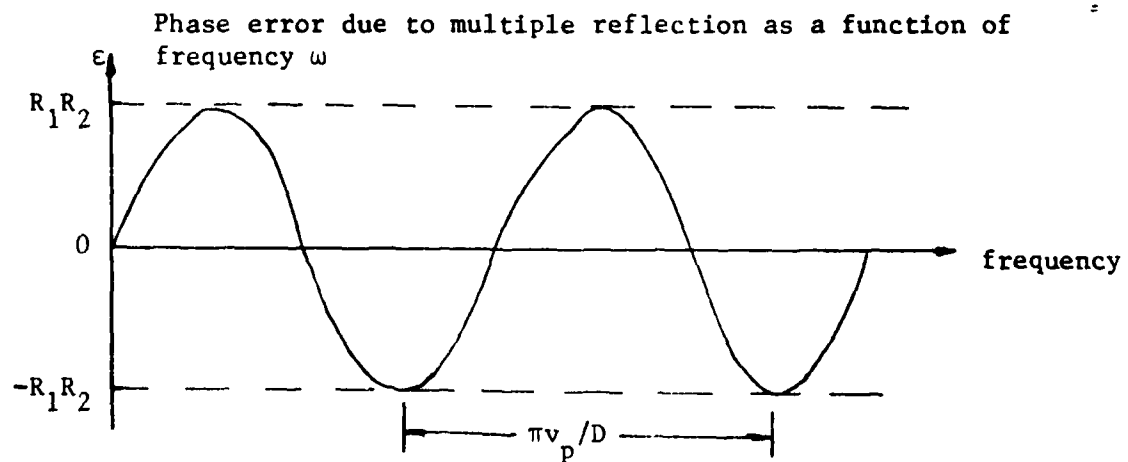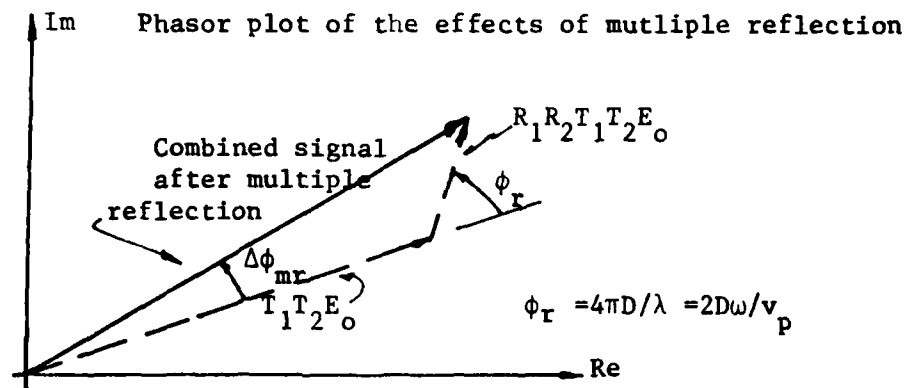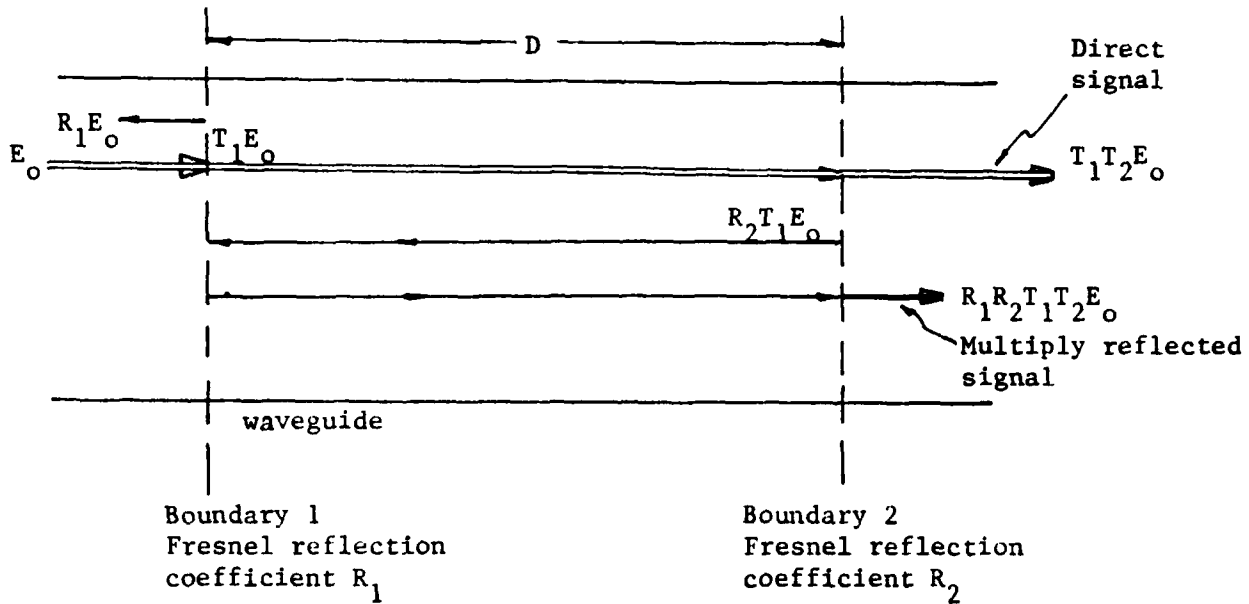over a limited spanned bandwidth, the dispersion is very close

to linear.

The ionospheric dispersion (see Chapter 4 for details)

will produce residual phases at X band of less than 1°, again,

because over the limited spanned bandwidth the inverse

frequency relationship is almost linear. (For a differential

electron content of 10^{18} electrons/m^2 (corresponding to an

ionospheric delay at X-band of ≈1.3 nsec) and the X-band

frequency sequence shown in Figure 3.1.4, the maximum residual

phase would be $-0^{\circ}\!.1$ in the 8.34 GHz channel. At S-band, where

the ionospheric effects are much larger, the residual phases

could reach $-0^{\circ}\!.4$ in the 2.70 GHz channel (assuming the channel

sequence is 2.295 GHz, 2.270 GHz and 2.220 GHz). This
frequency sequence was used in our July 1980 and June 1981
VLBI experiments.

The next possible origin for the dispersion which we will
investigate is multiple reflections. We will study a simple
example of a multiple reflection that will allow us to
characterize the general properties of multiple refections.
For a multiple reflection to exist there must be at least two
reflecting boundaries in the propagation path of the signals
(the signal in this case could be either the radiation from
the radio source or the phase calibration pulses). In our
simple example, a multiply reflected signal will be generated
whenever a signal is partially reflected from one boundary,
reversing its propagation direction, and then again partially
reflected from another boundary, again reversing its
propagation direction (hence the name "multiple reflection").
In Figure 3.1.5 we show the geometry for multiple reflections.
The characteristics of the multiple reflection can be
specified by three parameters: the distance between the
reflecting boundaries, and the reflection characteristics of
each of the boundaries.

The specification of all of the possible reflecting
boundaries in the Mark III system is beyond the scope of this
thesis. We will, instead, try to characterize generally the
properties of the multiple reflections in the Mark III system
based on the available residual phase data. The examination
of all of the reflecting boundaries in the Mark III system is

Figure 3.1.5 Geometry of a multiple reflection
(see text for symbol definitions)

Boundary 1
Fresnel reflection
coefficient R_1

Boundary 2
Fresnel reflection
coefficient R_2

Phasor plot of the effects of mutliple reflection

$$\phi_r = 4\pi D/\lambda = 2D\omega/v_p$$

Phase error due to multiple reflection as a function of
frequency ω

actually not necessary. Any multiple reflection which occurs after the injection point of the phase calibration pulses will equally affect the phases of both the signals from the radio source and the phase calibration signals. When the phase calibration phases are subtracted from the video DC cross spectral phases (see Appendix B), the changes in cross spectral phases due to the multiple reflection will be removed provided that the multiple reflection does not occur over such a large distance that phase "ripples" occur over the 2 MHz recorded bandwidth (see further discussion later in this section).

In the lower section of Figure 3.1.5 we plot as a function of frequency the effect of a multiple reflection on the phase of a signal which has been multiply reflected. The amplitude of the phase error will be a function of the reflection properties of the boundaries. The "wavelength" of the error, i.e. the change in frequency necessary for the error to complete one cycle, will be a function of the distance between the reflecting boundaries.

We will derive the relationship between the phase errors and frequency as a function of the distance, D, between the reflectors, and the Fresnel reflection coefficients of the boundaries, R_1 and R_2 (see Figure 3.1.5). We will assume that the reflection coefficients are sufficiently small that we can use small angle approximations in calculating the phase errors. (The magnitude of the dispersion ($< 10°$) indicates that these approximations should be adequate for our calculations.)

We will also assume that the reflecting boundaries are normal to the direction of propagation of the radiation.

In Figure 3.1.5 we show the amplitude of the electric field after each interaction of the signal with a boundary. The amplitude of the directly transmitted signal will be $E_t = T_1 T_2 E_o$ where T_1 and T_2 are the two Fresnel transmission coefficient of the reflection boundaries, and E_o is the amplitude of the electric field before any reflections. These transmission coefficients are related to the reflection coefficents by $T_1=1+R_1$ and $T_2=1+R_2$ (See, for example, Kong, 1975, Chapter 4). The amplitude of the multiply reflected signal's electric field will be $E_r = R_1 R_2 T_1 T_2 E_o$ and its phase, relative to the directly transmitted signal, will be $\phi_r = 2D\omega/v_p$, where D is the distance between the reflectors, and v_p is the phase velocity of propagation through the medium between the reflecting boundaries. The phase error, $\Delta\phi_{mr}$, at frequency ω, due to the muliple reflection will be (see Figure 3.1.5)

$$\Delta\phi_{mr} \simeq R_1 R_2 \sin(2D\omega/v_p) \quad . \tag{3.1.3}$$

Does this expected error match the observed residual phases?

The widely spaced frequencies used in geodetic experiments do not allow us to answer this question because the phase error could oscillate many times in the frequency intervals between the channels. We can, however, calculate a minimum value for the reflection coefficients. For a 4° phase error to be generated by multiple reflections (this value is the maximum mean residual phase on the Haystack to Owens Valley baseline in June 1981; see Figure 3.1.3), the product
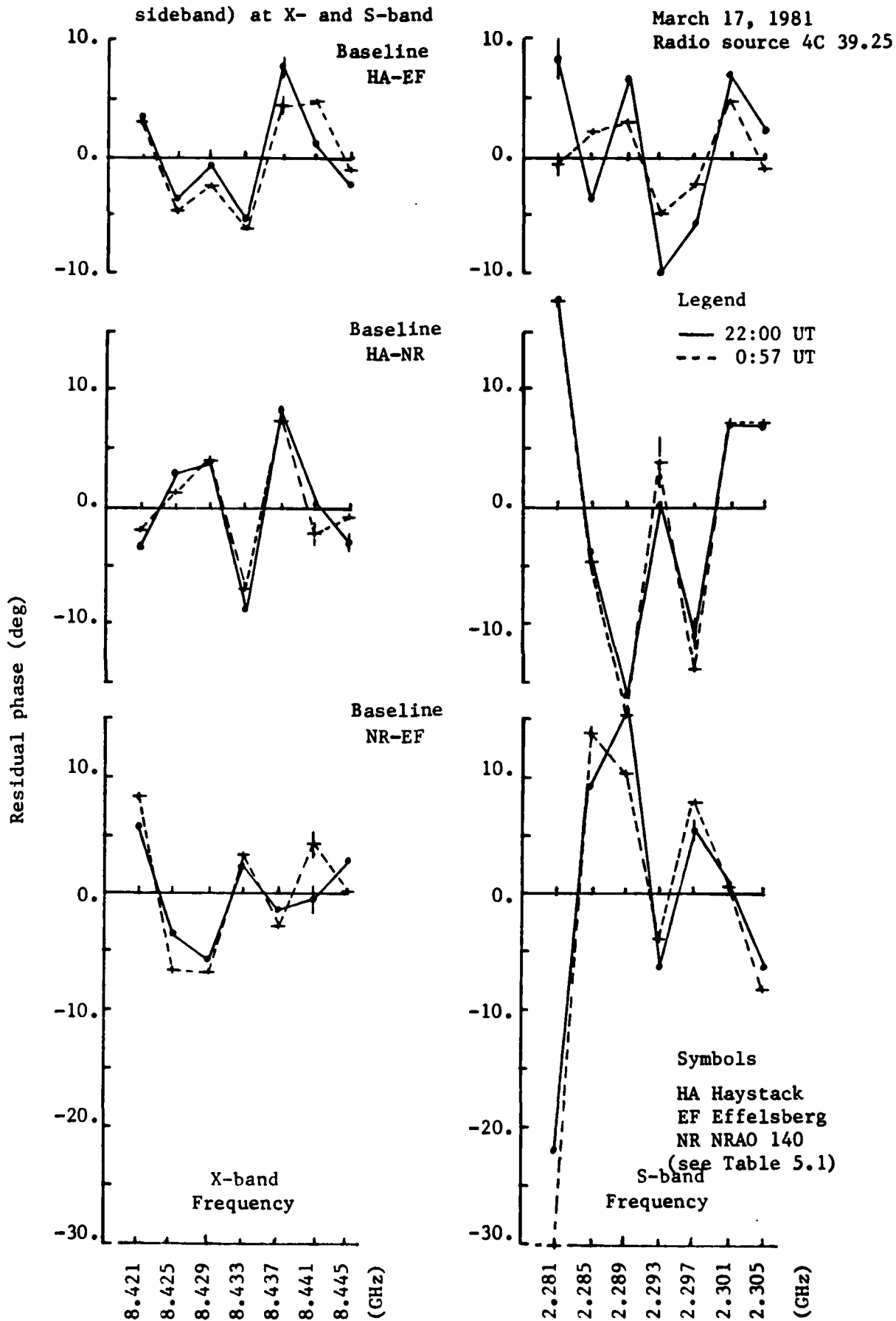
$R_1 R_2$ must be at least 0.07. If we assume that $R_1 = R_2$, then the power reflected at each boundary is 7%. However, mean residual phases of up to 16° have been observed for other baselines in other experiments (see Figure 3.1.6), which would imply power reflections of up to 28%.

To study the frequency dependence of the multiple reflections we need to look at observations with uniformly spaced frequencies so that the sinusoidal signature of a multiple reflection can be seen. This type of observation is generally only used for astronomy observations of weak sources (where spanned bandwidth is not important). We do, however, have one set of observations on a strong radio source. (Strong sources or large antennas are necessary to ensure that the random errors in the residual phases are small.) In Figure 3.1.6 we show plots of the residual phases at both X- and S-band from the triplet of baselines involving Haystack, NRAO and Effelsberg. (These stations were chosen because in all cases the standard deviations of the residual phases were less than 1°.) Each plot shows results from observations of the radio source 4C39.25, at two epochs separated by 2 hr 27 min. This radio source was only observed twice in this experiment. The primary observations in this experiment were of 1038+528A and B (Marcaide, 1982).

The residual phases in Figure 3.1.6 do not match the expected sinusoidal variation expected from a single multiple reflection. This lack of agreement could be due to there being more than one multiple reflection in the system (e.g.,

-73-

Figure 3.1.6 Residual phases in seven contiguous frequency channels (dual
sideband) at X- and S-band

March 17, 1981
Radio source 4C 39.25

Baseline
HA-EF

Baseline
HA-NR

Baseline
NR-EF

Legend

—— 22:00 UT
- - - 0:57 UT

Symbols

HA Haystack
EF Effelsberg
NR NRAO 140
(see Table 5.1)

Residual phase (deg)

X-band
Frequency

S-band
Frequency

8.421
8.425
8.429
8.433
8.437
8.441
8.445
(GHz)

2.281
2.285
2.289
2.293
2.297
2.301
2.305
(GHz)

possibly both the signal from the radio source (before injection of the phase calibration pulses) and the phase calibration pulses (before being injected) are multiply reflected. If the variations of the residual phases are due to multiple reflections, we can calculate a minimum distance between the reflecting surfaces. For each baseline in Figure 3.1.6, the residual phases seem to "cycle" at least once over the 28 MHz. From Equation (3.1.3), we see that for $\Delta\phi_{mr}$ to rotate through one cycle in 28 MHz, the minimum distance, D_{min}, between the reflectors must be

$$D_{min} = v_p/(2\Delta f) \approx 6.25 \text{ m}$$

where we have taken $v_p = c$. The phase velocity through the waveguide could be as much as 28% greater than $\underline{c}$ (see Equation 3.1.1), thus increasing D_{min}.

The residual phases for the two epochs agree well for the Haystack to NRAO baseline and reasonably well for the X-band band results involving baselines to Effelsberg. The S-band results for the baselines involving Effelsberg, however, show differences of up to 4°. (This is approximately 10 times the standard deviation of the difference on the NRAO to Effelsberg baseline.) There are no apparent errors in the processing of any of these S-band results and we are led to suspect that the difference is due to a change in the antenna or the receiver between these two observations. The Effelsberg antenna is 100 m in diameter (it is the largest fully steerable antenna in the world), and possibly we are seeing some changes in the geometry of the antenna (the observations were at elevations

of 72° and 50°). There is also a possibility that these variations are due to the phase calibration error which will be discussed in Section 3.2. Unfortunately with only two observations it is not possible to identify this error (see the method used to detect phase calibration errors in Section 3.2).

Another possible cause of the dispersion is asymmetry of the phase calibration pulse shape. This problem is very similar to multiple reflection of the phase calibration pulses, because a multiple reflection results in a distortion of the pulse shape. The effects of pulse shape on the phase of the calibration pulses can be studied using the convolution theorem of Fourier transforms (Bracewell, 1978, p.108). The phase calibration pulses, $p_c(t)$, can be represented (in the time domain) by

$$p_c(t) = \int_{-\infty}^{\infty} h(u) \; III(t/t_p-u) \; du \qquad (3.1.4)$$

where $h(u)$ is the pulse shape function, t_p is the time interval between pulses, and $III(t/t_p-u)$ is the sampling or replicating function (Bracewell, 1978 pp. 77-79) and represents an infinite series of Dirac delta functions separated by t_p.

The Fourier transform of Equation (3.1.4) will be given by (using the convolution theorem)

$$P_c(\omega) = H(\omega) \; III(\omega/\omega_p) \qquad (3.1.5)$$

where ω_p is the frequency spacing between the comb elements of the phase calibration rails (see Section 2.3) and equals

$2\pi/t_p$. From Equation (3.1.5), we can see that the phase of $P_c(\omega)$ will depend on the Fourier transform of the pulse shape. We should notice that if the pulse is very sharp, i.e. if the duration of the pulse is very small, then the phase of $H(\omega)$ cannot vary rapidly with frequency. (This result can be deduced from the bandlimited nature of the pulse in the time domain.) From Figure 3.1.6 we can deduce properties of the pulse shape, if the pulse shape is the cause of dispersion. Since we see deviations of $5°$ in the residual phases which seem to cycle in 28 MHz, we can deduce that ≈ 0.1 $(\sin 5°)$ of the power in the pulse must occur at least 0.035 µsec (1/28 MHz) away from the main pulse. Given that the pulse is supposed to be less than 0.00005 µsec duration these numbers seem very unlikely.

These results do suggest a possible cause for the dispersion. The Mark III pulse generator produces a pulse every zero crossing of the 5 MHz input signal, i.e., 10 pulses/µsec are produced. Nine of the ten pulses should be suppressed, or gated, before they enter the Mark III receiver. However, if the gating electronics are not working properly, then some of these additional pulses could be entering the receiver and corrupting the phase calibration signal extraction. An additional pulse could be 0.1 µsec away from the primary pulse which is in accord with the minimum separation which we calculated from Figure (3.1.6) (0.035 µsec). The gating electronics should be carefully checked.

We have now studied several possible causes of the dis-

persion in the Mark III system and none of the causes (individually) could fully explain the observed ampltitude verus frequency pattern of the dispersion. All of the possible causes studied probably contribute to the dispersion to some extent. An analysis of the Mark III hardware is now being commenced to ascertain, if possible, the origin(s) of the dispersion. We now investigate the seriousness of the dispersion on the measurement of the visibility phase and the group delay.

The dispersion in the Mark III system can cause two problems. Firstly, if the relative weights of the data in each frequency channel change with each observation then the contribution of the dispersion to the group delay will change because the group delay estimation is a weighted least squares fit of a slope to the cross spectral phases (see Appendix A). Secondly, if the dispersion is time dependent then there will also be variations of the group delay.

The first of these points may be addressed by simply calculating the weighted least squares fits to a set of hypothetical cross spectral phases using different weights in each channel. In Figure 3.1.7 we show, as an example, the results of such a series of calculations. We have chosen the S-band frequency sequence which was used in experiments up to June 24, 1981. The frequency channels are located at 2.295 GHz, 2.270 GHz, 2.220 GHz. This sequence was chosen because in Chapter 4 we will be analyzing extensively data taken at these frequencies. In addition the S-band group delays will be most

Figure 3.1.7 The effects of a 5° dispersion on the group delay estimates when the weight of the 2.220 GHz channel is varied
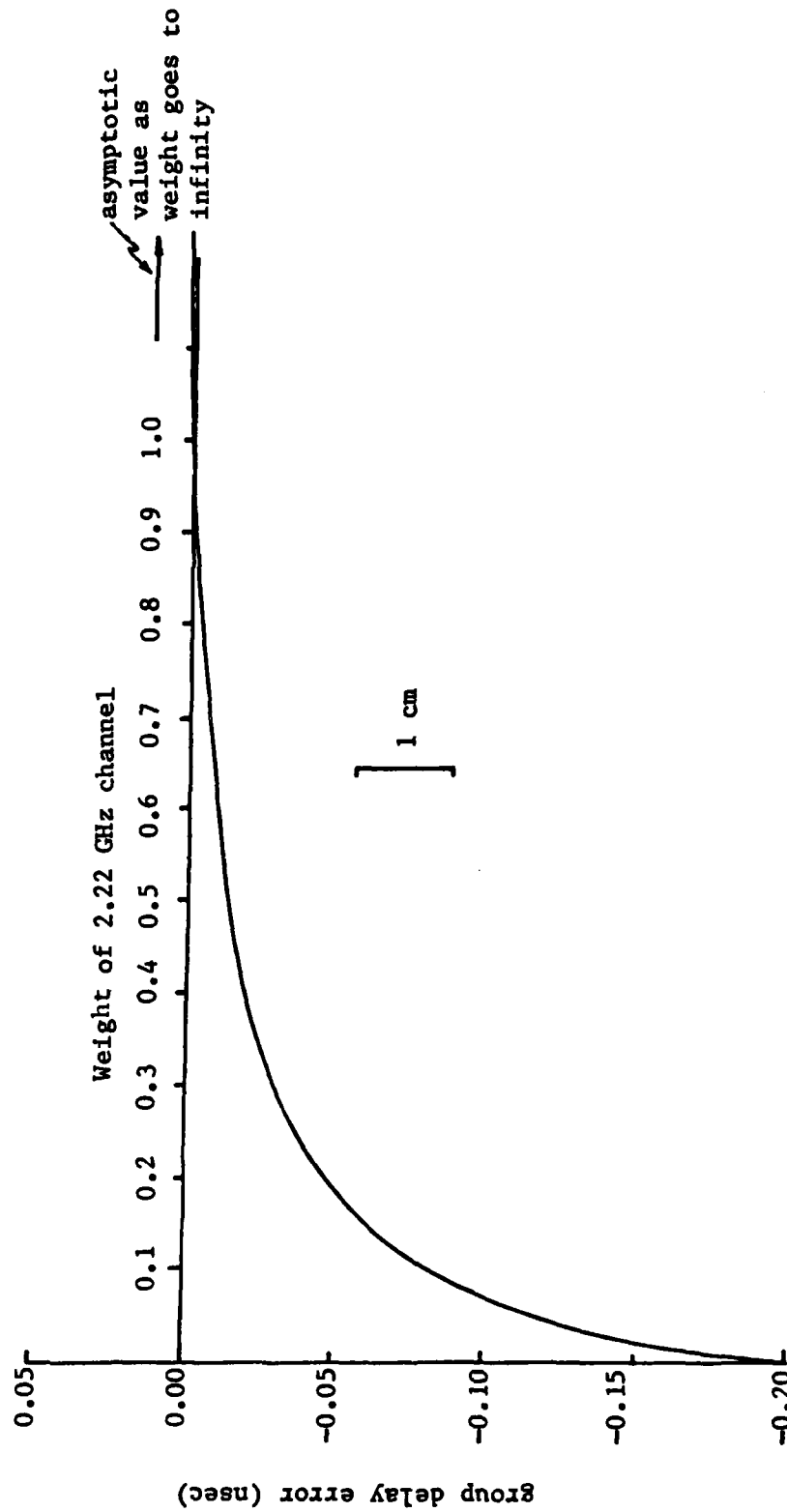
Based on frequency sequence 2.295 GHz, 2.270 GHz, and 2.220 GHz, and a 5° dispersion in the 2.220 GHz channel

The group delay errors are referred to the group delay obtained when the weights in all of the channels are the same (see text).

sensitive to weight changes because of the narrow spanned bandwidth (75 MHz), and the use of only three frequency channels. To generate Figure 3.1.7 we assumed that there was a 5° dispersion in the 2.220 GHz channel and that the weight of this channel relative to the 2.295 Ghz and 2.270 GHz channels, whose weights remained constant, changed between 0 and infinity. The group delay errors are referred to the case when the weights in all channels are equal because this should be the typical value for most observations. (The change in weight of the 2.220 GHz could be due either to changes to the amplitude of the signal or to changes in the number of accumulation periods used from the channel.) In Figure 3.1.7, we see that even for a small dispersion the errors in the group delay can be as large as 0.2 nsec. However, we should notice that large errors (>2 cm) only develop for very disproportionate weights among the frequency channels. (Note that the multiband group delay does not approach the single-band group delay as the weight of the 2.220 GHz channel approaches infinity because these two delay types are treated as independent parameters by FRNGE; see Appendix B.3 for details.)

We now address the constancy of the dispersion. Even though the precise origin of the dispersion is not known, we can study the constancy of the dispersion by accumulating their statistics from several experiments. There are two data sets which we can use for these studies - the short MERIT campaign and the POLARIS experiments (see Chapter 5 for

experiment details). We study these two sets of data because for long periods of time (one month for MERIT and six months for POLARIS) the Mark III systems were left undisturbed at the radio telescopes.

In Figure 3.1.8, we show one-day averages of the residual phases from the short MERIT campaign for three baselines: Haystack-OVRO 130, Haystack-Onsala 60, and OVRO 130-Onsala 60 (see Figure 5.1 and Table 5.1 for station locations and definitions for terminology). Each point plotted is an average of between 60 and 100 residual phase values and hence the uncertainty of each mean value is quite small, typically $<0.^{\circ}5$. The uncertainties of the mean values are plotted at representative points. These uncertainties were calculated from the formal standard deviation of the mean, scaled by the normalized-root-mean-square (NRMS) scatter of the individual values about the daily mean. (The NRMS is the square root of the χ^2 per degree of freedom of the fit of the observations to the mean value; see detailed discussion later in this section.)

One feature in Figure 3.1.8 is very noticeable. On October 20, there is an apparent change in the dispersion at OVRO 130. We should notice that the error bars for this day are much larger than those for adjacent days, although the difference between the mean residual phase for this day and for the adjacent days is still many standard deviations, $\approx 6\sigma$. The reason for the increase in the uncertainty is related to the apparent change in the dispersion. In Figure 3.1.9 we

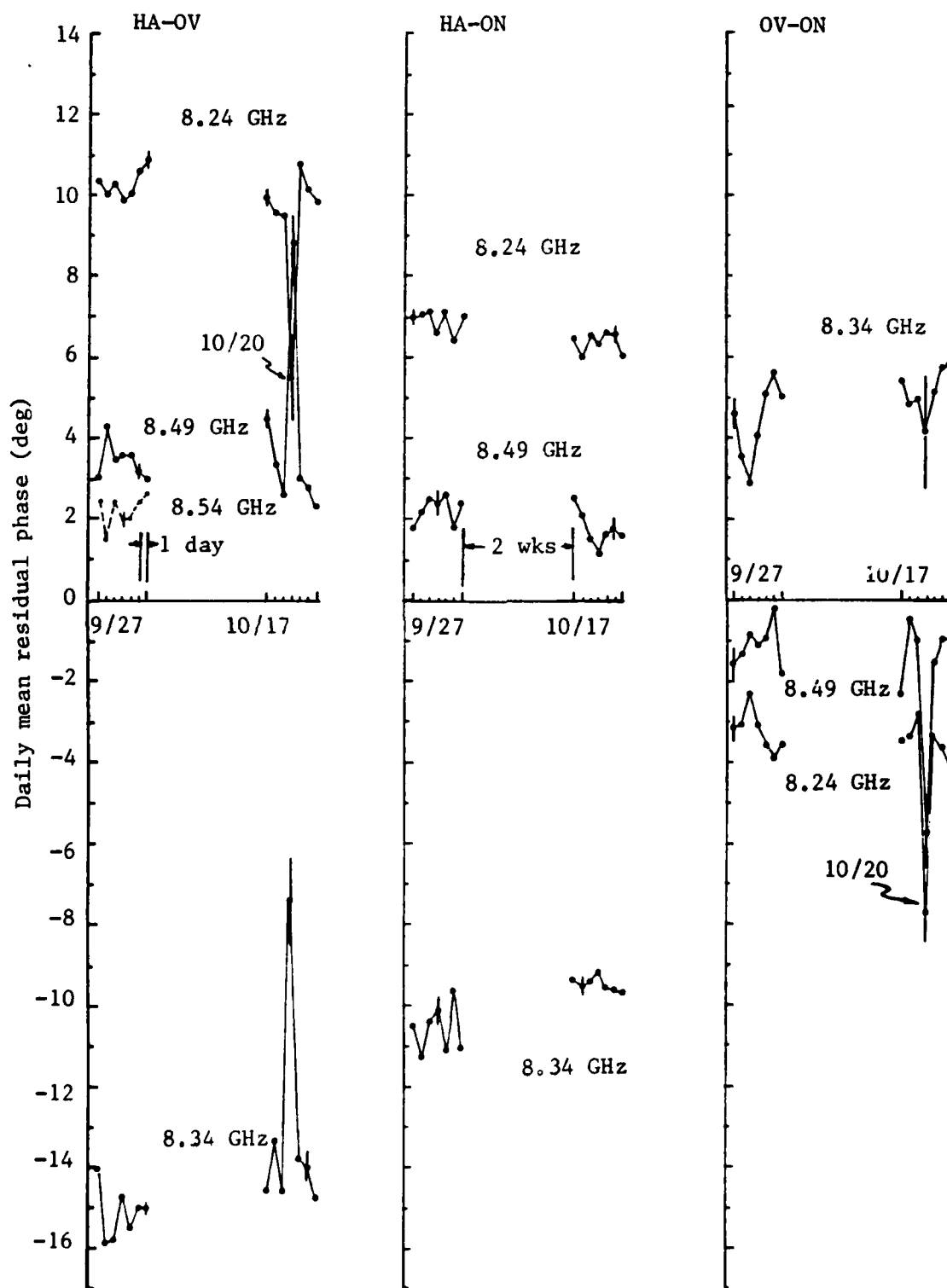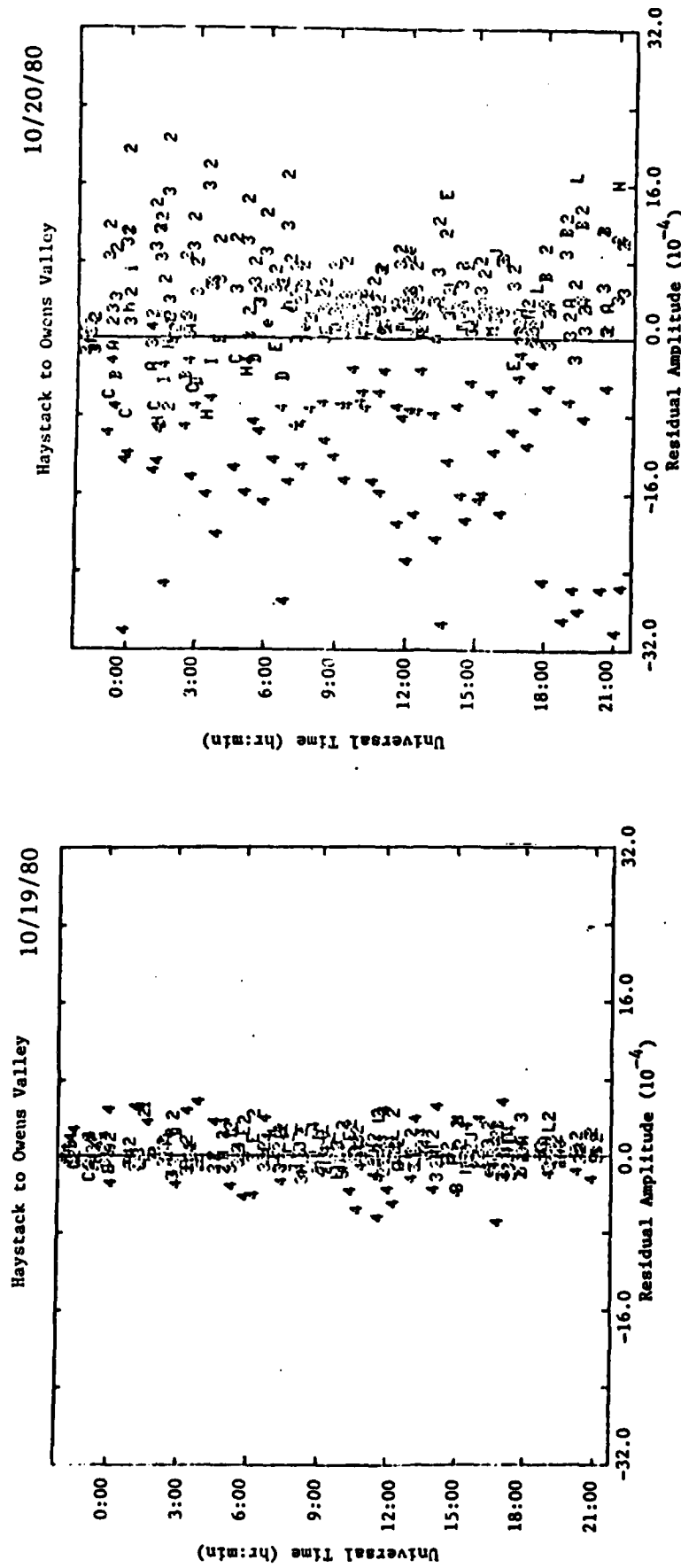Figure 3.1.8 Daily mean residual phases from the short MERIT campaign

Figure 3.1.9 Residual correlation amplitudes for the Haystack-Owens Valley baseline on October 19 and 20, 1980

show the residual amplitude for the anomalous day and the preceding day. (The residual amplitudes are the differences between correlation amplitudes in each channel and the mean correlation amplitude for the observation.) The symbols in these plots have the same meanings as those in Figures 3.1.1 and 3.1.2.

It is clear from Figure 3.1.9 that the residual amplitudes show much more scatter on 20 October than on 19 October. (The scatter on 21 October was the same as that on 19 October.) Also Channel 4 (8.24 GHz) has systematically lower amplitudes than do the other frequencies.

We mentioned earlier in this section that one of the major problems associated with the dispersion is the apparent change in the group delays, and consequently the apparent change in the dispersion, when the weight given to a channel is changed. The anomalous behavior on 20 October gives an indication of how large these effects may be.

If we exclude 20 October from Figure 3.1.8, we see that there are still significant variations of the dispersion between the daily averages. These variations seem to be less than 1° per day in most cases. What does this indicate about the constancy of the dispersion? To answer this question, we must remember that the group delay and visibility phase have been estimated from cross spectral phases and hence any variations in the dispersion will be partially absorbed into these estimates. Consequently, the real change in the dispersion will appear smaller in the residual phases. (The phase delay

rate is also estimated from the cross spectral phases.
However, since the group and phase delays are referred to an
epoch near the (temporal) center of the data, the phase delay
rate estimate is decoupled from the delay measurements. We
may therefore study the variations in the dispersion by
considering only estimates of the group delay and visibility
phase.)

We may calculate the expected reduction in the varia-
tions of the dispersion due to the estimation of the group
delay and visibility phase based on the formulas given in
Appendix E. In Table 3.1.1 we give the numerical values of
matrices applicable to studying dispersion variations (see
Appendix E for a discussion of the techniques used to generate
Table 3.1.1). We have generated this table for the frequency
sequences used up to June 24, 1981 because most of the data we
will be analyzing were obtained at these frequencies.

Of most importance to us at this time are the numerical
values associated with Equations (E.5) and (E.6). Equation
(E.5) relates the change in the dispersion (denoted by vector
$\underline{\varepsilon}_r$ in Table 3.1.1) to the changes in the group delay and
visibility phase. Equation (E.6) relates the change in the
dispersion to the change in the residual phases, $\underline{v}$, i.e. to
the apparent change in the dispersion as deduced from residual
phases such as those shown in Figure 3.1.8. (We should note
that the matrix which relates $\underline{\varepsilon}_r$ to $\underline{v}$ is singular and hence it
is not possible to obtain unique values for $\underline{\varepsilon}_r$ given $\underline{v}$; see
Appendix E for further discussion.)

Table 3.1.1 Numerical values for studying the effects of changes in dispersion

For X-band

Frequency sequence 8.34, 8.49, 8.54, 8.24 GHz.

From Equation (E.2), $\underline{\phi}_r = A\underline{x}_p$

$$\underline{\phi}_r = \begin{bmatrix} 1.0 & 0.00 \\ 1.0 & 0.15 \\ 1.0 & 0.20 \\ 1.0 & -0.10 \end{bmatrix} \underline{x}_p \qquad \underline{x}_p = \begin{bmatrix} \phi_t \\ \tau_g \end{bmatrix}$$

where $\underline{\phi}_r$ and $\underline{\phi}_t$ are in cycles, and τ_g is in nsec.

From Equation (E.5.a), $\hat{\underline{x}}_p = A^-\underline{\phi}_r$

$$A^- = \begin{bmatrix} 0.319 & 0.154 & 0.099 & 0.429 \\ -1.099 & 1.538 & 2.418 & -2.857 \end{bmatrix} \begin{matrix} \text{cycle/cycle} \\ \text{nsec /cycle} \end{matrix}$$

From Equation (E.6), $\underline{v} = \underline{\phi}_r - \hat{\underline{\phi}}_r = (I - AA^-)\underline{\varepsilon}_r$

$$\underline{v} = \begin{bmatrix} 0.681 & -0.154 & -0.099 & -0.429 \\ -0.154 & 0.615 & -0.462 & -0.001 \\ -0.099 & -0.462 & 0.417 & 0.143 \\ -0.429 & -0.001 & 0.143 & 0.285 \end{bmatrix} \underline{\varepsilon}_r$$

where $\underline{\varepsilon}_r$ is the change in the dispersion and $\underline{v}$ is the change in the residual phases which would be produced by $\underline{\varepsilon}_r$

Table 3.1.1 Continued.

For S-band

Frequency sequence 2.295, 2.270, 2.220 GHz.

From Equation (E.2), $\underline{\phi}_r = A\underline{x}_p$

$$\underline{\phi}_r = \begin{bmatrix} 1.0 & 0.000 \\ 1.0 & -0.025 \\ 1.0 & -0.075 \end{bmatrix} \underline{x}_p \qquad\qquad \underline{x}_p = \begin{bmatrix} \phi_t \\ \tau_g \end{bmatrix}$$

where $\underline{\phi}_r$ and $\underline{\phi}_t$ are in cycles, and τ_g is in nsec.

From Equation (E.5.a), $\underline{\hat{x}}_p = A^-\underline{\phi}_r$

$$A^- = \begin{bmatrix} 0.714 & 0.429 & -0.143 \\ 11.428 & 2.857 & -14.286 \end{bmatrix} \quad \begin{array}{l} \text{cycle/cycle} \\ \text{nsec /cycle} \end{array}$$

From Equation (E.6), $\underline{v} = \underline{\phi}_r - \underline{\hat{\phi}}_r = (I - AA^-)\underline{\epsilon}_r$

$$\underline{v} = \begin{bmatrix} 0.286 & -0.429 & 0.143 \\ -0.429 & 0.642 & -0.214 \\ 0.143 & -0.214 & 0.072 \end{bmatrix} \underline{\epsilon}_r$$

where $\underline{\epsilon}_r$ is the change in the dispersion and $\underline{v}$ is the change in the residual phases which would be produced by $\underline{\epsilon}_r$

If we now reconsider the variations in the mean residual phases shown in Figure 3.1.8, we can make some estimates of the stability of the Mark III system. As mentioned earlier we cannot obtain unique estimates of the changes in the dispersion from the residual phases. One estimate of the variations in the dispersion is exactly the variations in the residual phases, i.e. one solution to Equation (E.6) is $\underline{\varepsilon}_r = \underline{v}$. There would be no variations in the delay introduced by the dispersion if the $\underline{\varepsilon}_r$ were equal to $\underline{v}$. This solution is very unlikely to be relevant. We may, however, use this solution to obtain an estimate of the minimum variation of the dispersion. The weighted root-mean-square (WRMS) scatter of the mean residual phases about the mean values for each channel and baseline, is $0^\circ\!.7$ (excluding 20 October). Variations of the dispersion of at least this amount occur in the Mark III system. We cannot place an upper bound on the variations of the dispersion (based on the residual phases) because any variation in the dispersion which mimics a delay or a phase offset will be absorbed into the group delay and visibility phase estimates and hence not be seen in the residual phases.

The RMS variation of $0^\circ\!.7$ in the dispersion implies variations in the group delays of up to 8 psec/day at X band and 36 psec/day at S-band (these values are obtained from the numerical values of Equation (E.5) given in Table 3.1.1). We will compare these estimates with other estimates of the stability of the Mark III system, obtained from predicting phase delay

ambiguities from group delays, in Section 4.3.

We may further study the stability of the Mark III system on a time scale of 6 months by analyzing the mean residual phases from the POLARIS experiments. In Figure 3.1.10 we show, for a six-month interval, the weekly values of the mean residual phases from 24 hr duration observing sessions from the eight channel, single sideband Westford-HRAS baseline observations. Again, we see that the dispersion is large, up to $7°$ for some experiments.

In general, the dispersion remains constant or varies smoothly over the six month period except for one discontinuity (in some channels) which occurred between 9/30/81 and 10/15/81. The mean residual phases from these two experiments are plotted as a function of frequency in Figure 3.1.11. Clearly, there seem to be changes in this dispersion curve at both high and low frequencies. These changes would give us an indication of the origin of the dispersion, if we could isolate a specific change in the configuration of the Mark III system at one of the antennas, because then we could deduce that this change affects the dispersion.

The station logs at Westford indicate that no changes were made between the two experiments which show the discontinuity (Webber, 1983, private communications). At HRAS, no detailed station logs were being kept (at that time), and hence we could not establish whether any changes had been made to the receiver (Sebring, 1983, private communication). (Shortly after this time a detailed station log was started.)

Figure 3.1.10 Weekly values of the mean residual phases from 24 hour observing sessions for the Westford–HRAS baseline
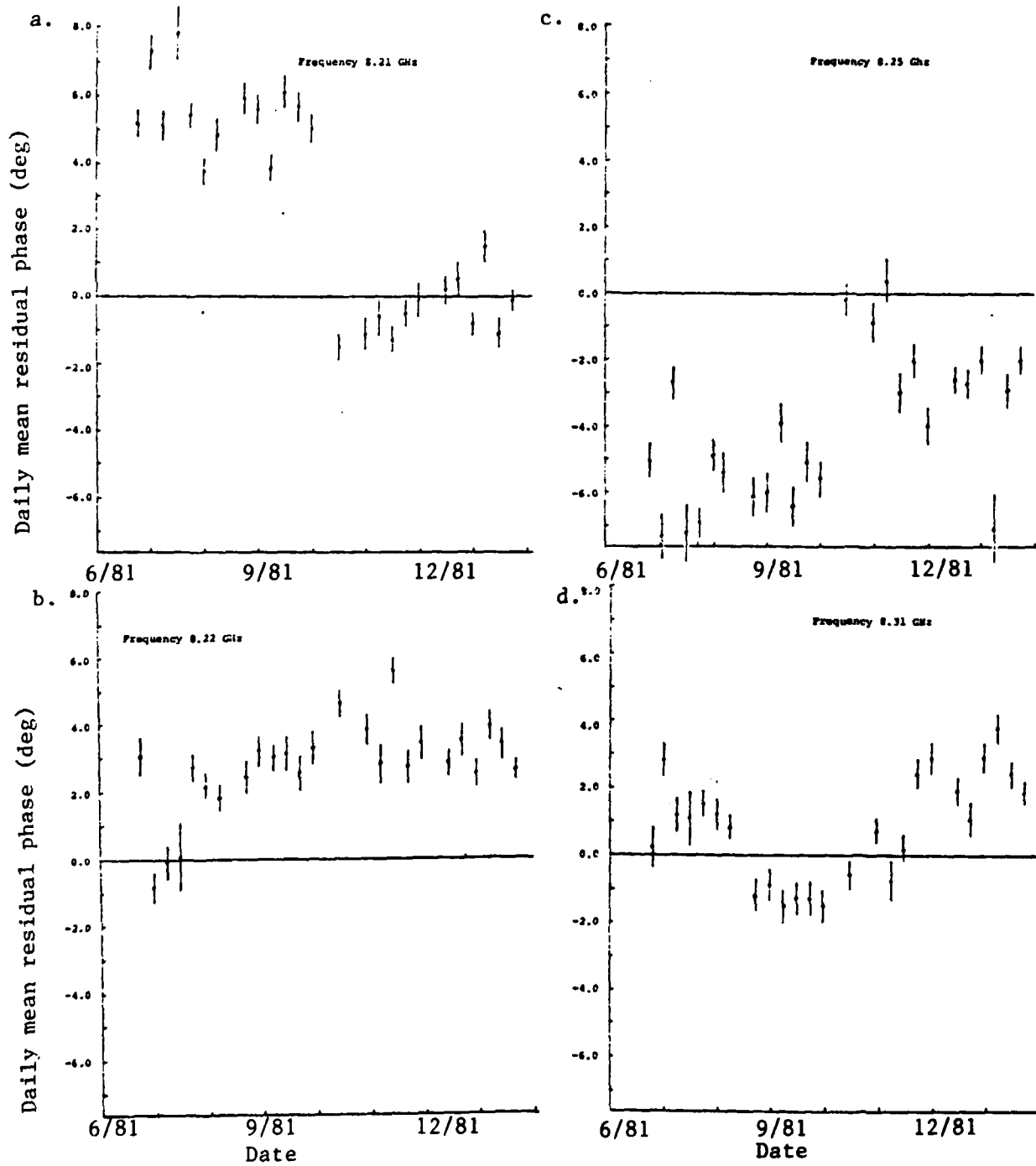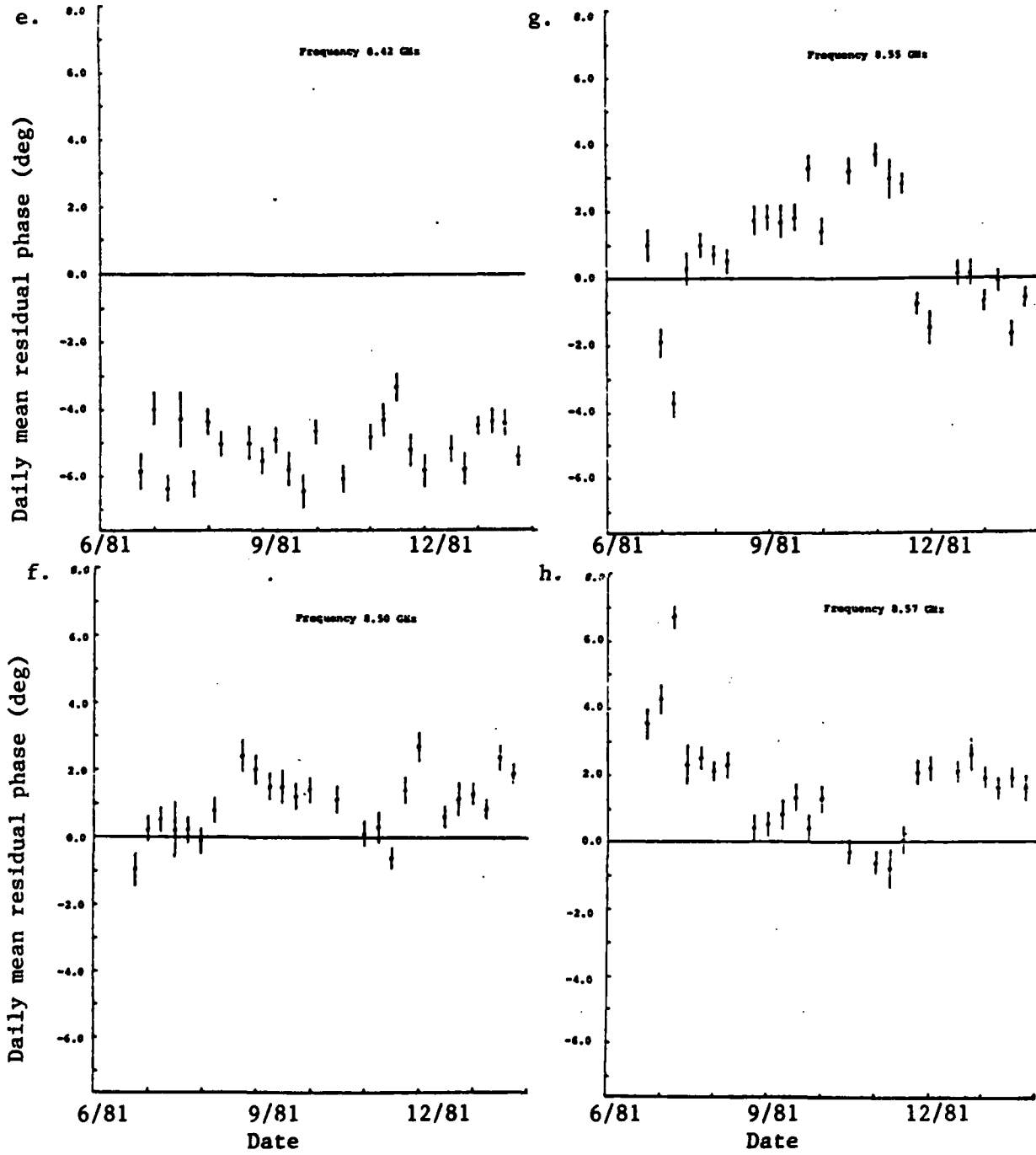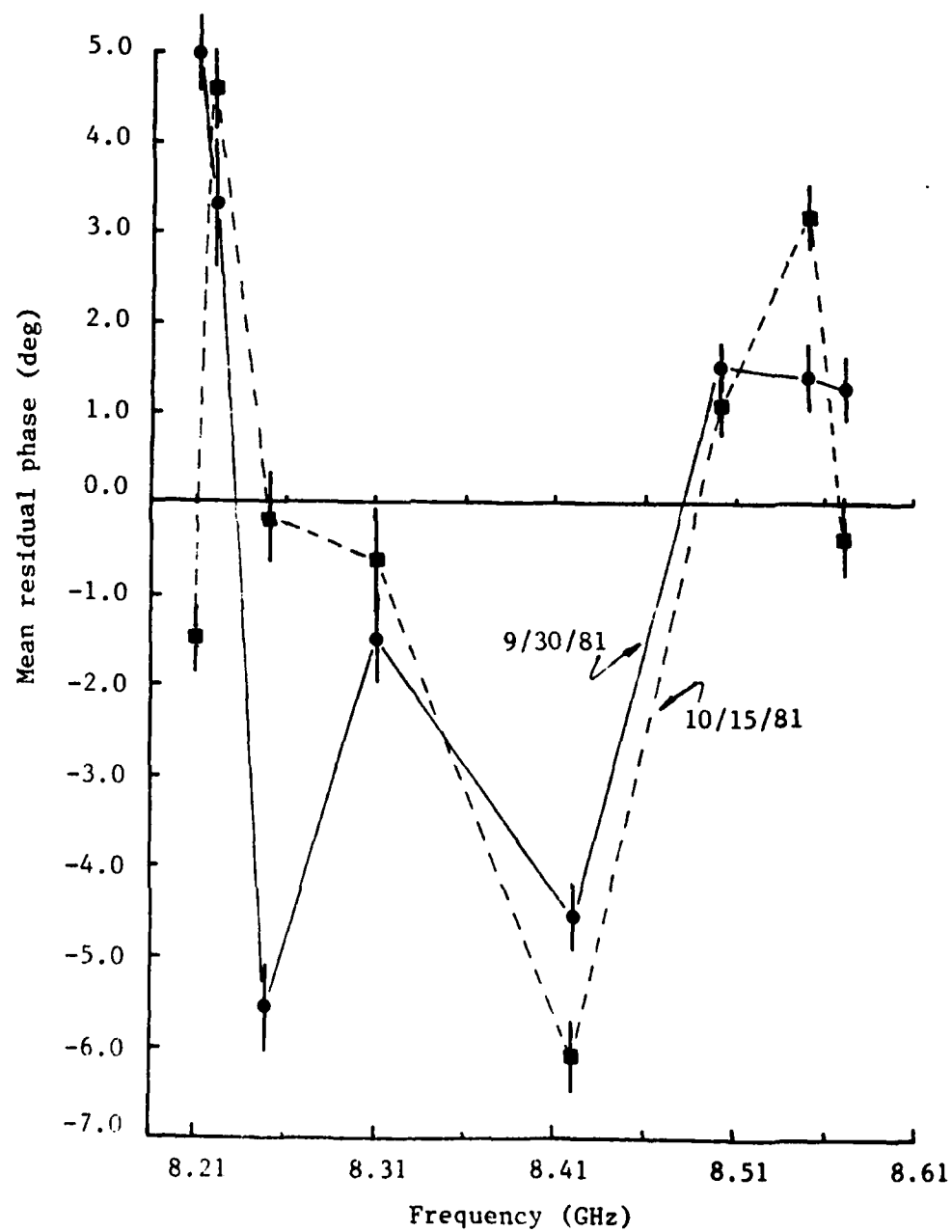
Figure 3.1.10 Continued

Figure 3.1.11 Changes in the mean residual phases for the Westford-HRAS
baseline between 9/30/81 and 10/15/81

For most of this section we have discussed the behavior of the mean of the residual phases from 24 hr observing sessions in an attempt to study the long term behavior of the dispersion in the Mark III system. We now turn our attention to the scatter of the residual phases from each observation about the 24 hr mean values. These results are compiled in Figure 3.1.12 where we have plotted the normalized root-mean-square (NRMS) scatter as a function of the expected scatter of the residual phases. (The NRMS is the square root of the χ^2 per degree of freedom of the residual phases about their mean values. We plot this quantity, rather than the χ^2 per degree of freedom, in order to decrease the dynamic range of the plot.)

We see clearly from this figure that the noise in the residual phases does not match the expected statistics of this noise. The statistics of the residual phases indicate two properties of the performance of the Mark III system. Firstly, there appears to be a limit on the accuracy of the Mark III system, i.e. even for observations of very strong radio sources the residual phases still show scatter which presumably arises from instrumental noise that is independent of the radio source being observed (e.g., the effects of variation in the weights of each channel when there is a large dispersion present; variations in the dispersion (see earlier discussions); and "noise" in the phase calibration system; see Section 3.2). Secondly, the NRMS scatter does not approach unity as the signal strengths decrease. This behavior could

Figure 3.1.12 Normalized RMS scatter of the residual phases as a function of their theoretical scatter
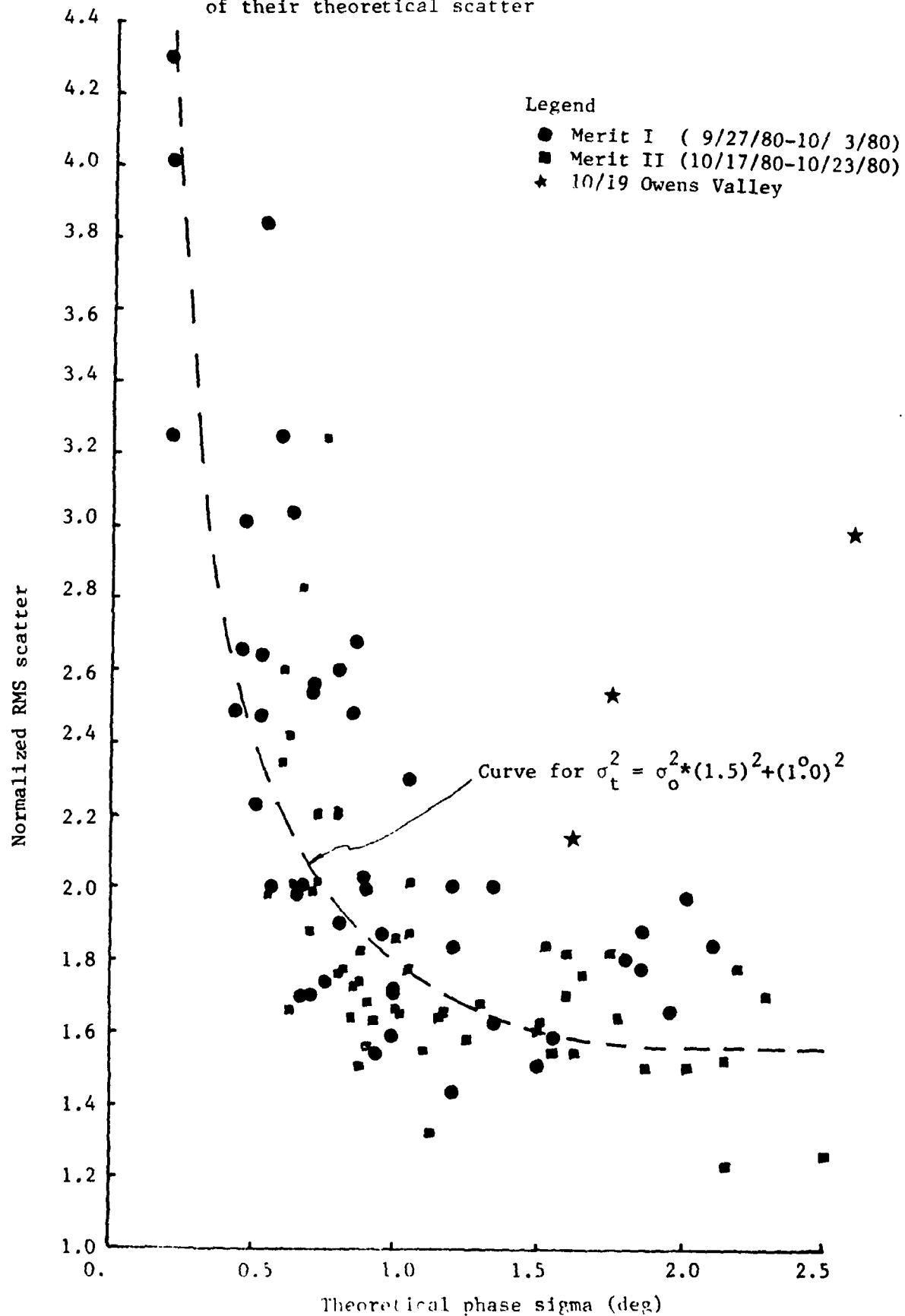
be due to instrumental noise sources which are a function of signal strength, e.g. the correlator non-reproducibility seems to behave in this fashion (see Appendix B.4).

We have empirically fitted a model to the behavior of the NRMS scatter which is shown as the solid line in Figure 3.1.12. This empirical model relates the "actual" standard deviation of the residual phases, σ_t (as deduced from their scatter), to their formal standard deviations, σ_o, by

$$\sigma_t^2 = \alpha_r^2 \sigma_o^2 + \sigma_c^2 \qquad (3.1.6)$$

where α_r and σ_c are the parameters to be estimated. The solid curve shown in Figure 3.1.12 corresponds to $\alpha_r = 1.5$ and $\sigma_c = 1^o$. (These values were approximately calculated to match the results shown in Figure 3.1.12.) These results will form the first part of our compilation of the performance of the Mark III system.

In Figure 3.1.13, we show a compilation of the statistics of the residual phases similar to that shown in Figure 3.1.12, for the intercontinental experiment conducted in November 1981 and the transcontinental experiment in June 1981. In this figure we also show the empirical law for the relationship between the NRMS and the formal estimate of the scatter which was deduced from the MERIT data. It is clear that in the eight- to twelve-month period between these experiments and the MERIT experiments, the scaling factor in the empirical relationship seems to have come closer to unity. This improvement is probably related to the improvement in the tape
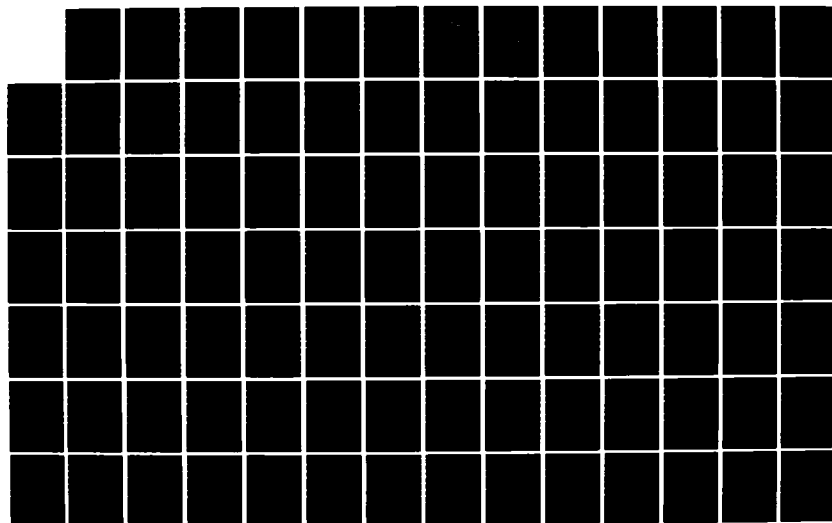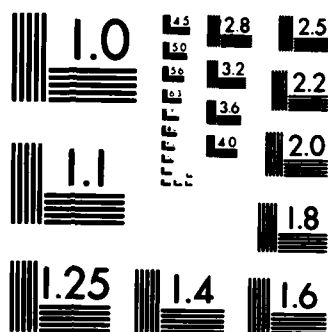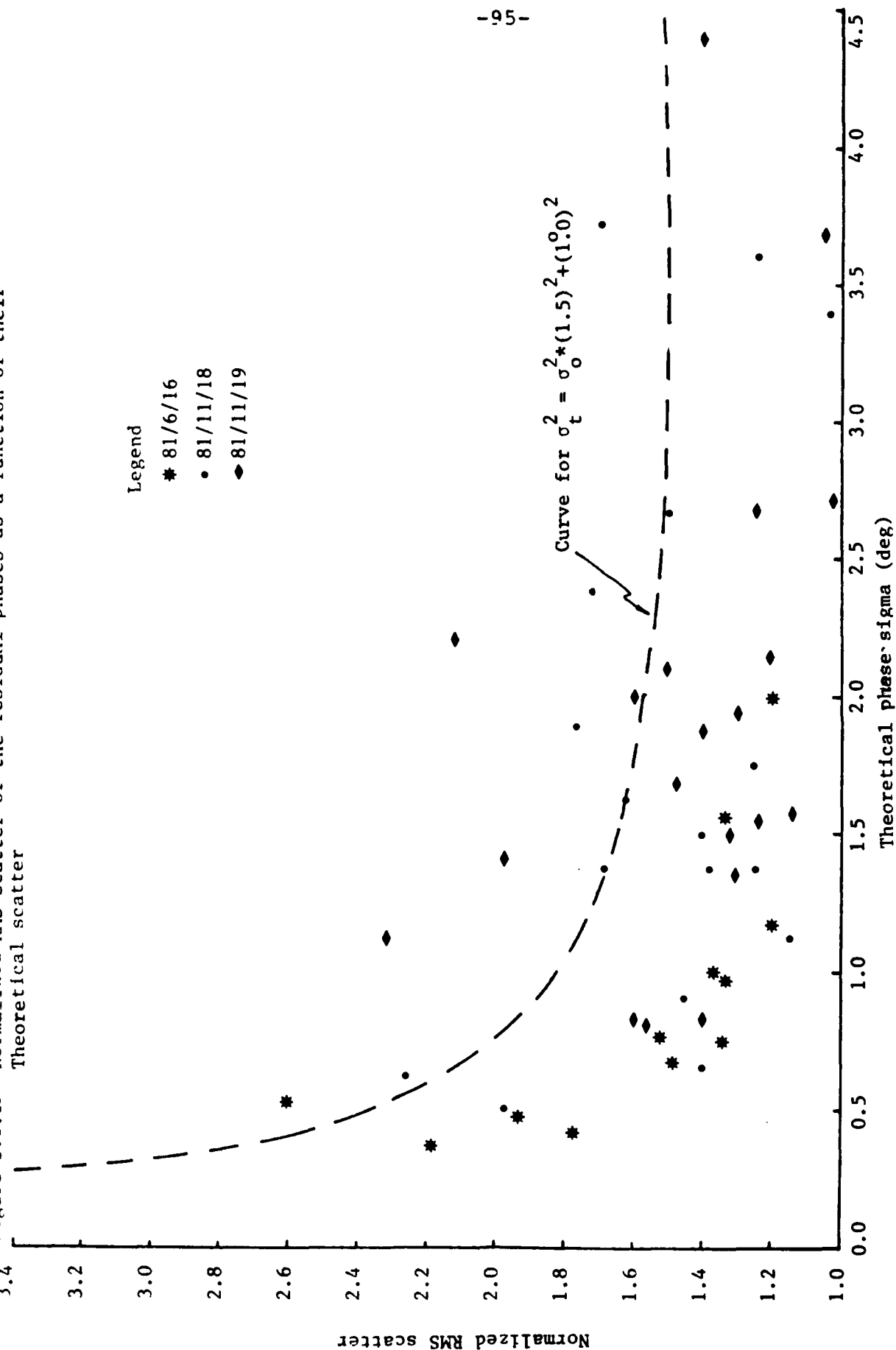
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Figure 3.1.13 Normalized RMS scatter of the residual phases as a function of their Theoretical scatter

quality between these experiments. Unfortunately, there are no very small standard deviation observations in the recent experiments which would allow us to assess the effect on σ_c. (The formal uncertainties of the residual phases are larger in November 1981 than in June 1981 and MERIT, because the November experiment used single sideband recording. We would expect an increase in the formal errors of $\approx\sqrt{2}$ because less data are used to form a channel with single sideband recording.)

We will return to these discussions of equipment performance after we have discussed the phase calibration system performance and the delay closures around triplets of baselines.

3.2 Phase calibration system

In Section 2.3, we discussed the operation of the phase calibration system. We now discuss the analysis of the phase calibration measurements. Firstly, we consider the components of the Mark III system which affect the phase of the calibration signals. We will then discuss the interpretation of the calibration phases from some VLBI experiments. As with the residual phase behavior, we will defer detailed analysis of the individual observations until Section 4.3, when we discuss the phase delay ambiguity elimination results.

In Figure 3.2.1 we show the calibration phases for a six-hour period, from the 8.34 GHz channel at Onsala, on July

Figure 3.2.1 Phase of the calibration signal in the 8.34 GHz channel
at Onsala during observations of 3C 345 and NRAO 512
on July 27-28, 1980.

STATION ONSALA60
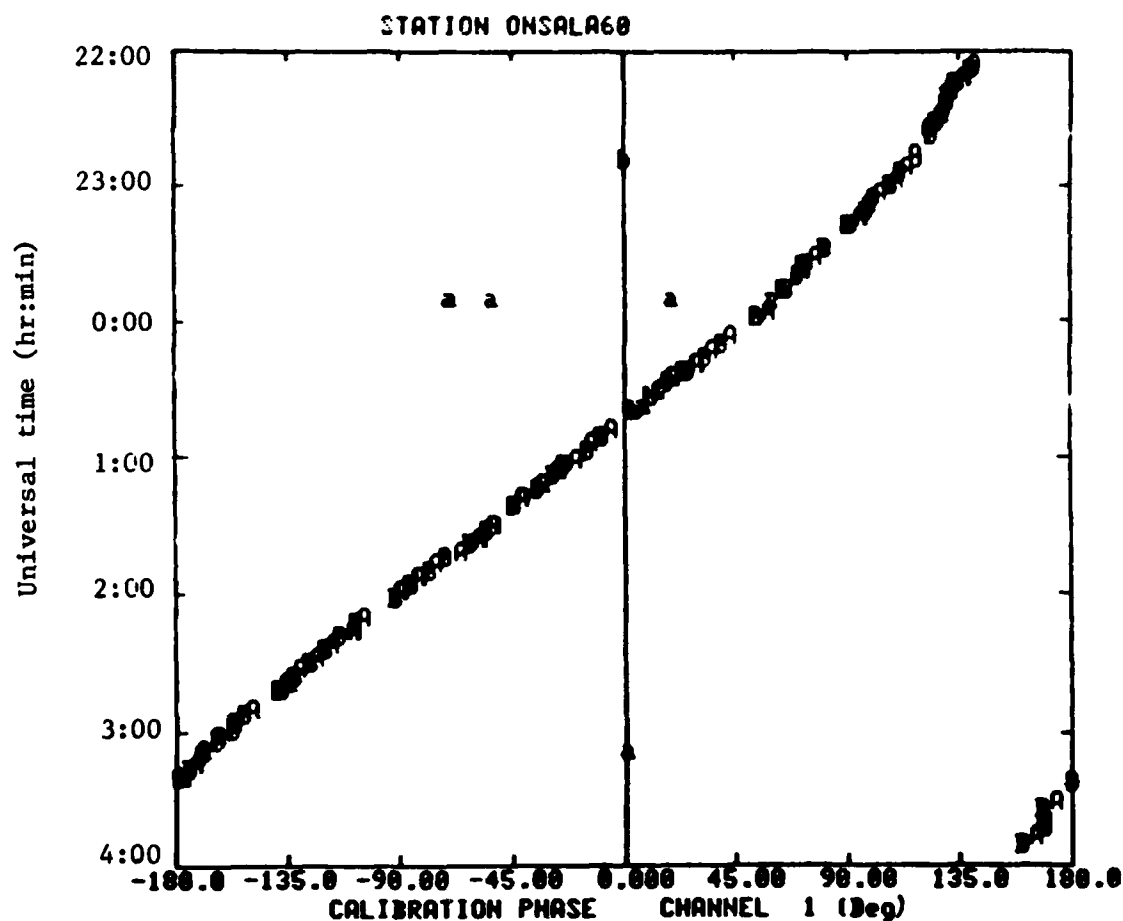
Legend

The letters A and B denote observations of 3C 345 and NRAO 512,
respectively. The lower case letters denote observations on
each radio source which have been flagged as defective by
FRNGE (see discussion at beginning of Chapter 3).
See Table 5.2 for IAU convention names for these radio sources.

28, 1980. (The lower case letters in this and other plots in this thesis, denote observations which have been flagged as defective by FRNGE; see beginning of Chapter 3.) These phases drift between $\pm 180^\circ$ over the six-hour period. What do we hope to learn from these phases? To answer this question we firstly need to consider the components of the Mark III system which affect these phases. In Section 2.3, we discussed the operation of the phase calibration system. A 5 MHz signal is sent, via coaxial cable, to the receiver where it is used to generate the LO frequencies for the first mixers and the pulses for the phase calibration system. These phase calibration pulses are injected into the signal train before the first mixing and amplification operations. The pulses then propagate, with the signals from the radio source, through the same components of the Mark III system. Based on this flow of the calibration signals, we can write an expression for their phase:

$$\phi_{cal} = \phi_{cable} + \phi_{LO} + \phi_{vc} + \phi_{instr} , \qquad (3.2.1)$$

where ϕ_{cal} is the phase of the calibration signals, ϕ_{cable} is the contribution to ϕ_{cal} from the length of cable carrying the 5 MHz signal (usually we only monitor changes in the cable length; see Section 2.3), ϕ_{LO} is the contribution of the phase of the first local oscillator, ϕ_{vc} that of the oscillator in the video converter for the channel, and ϕ_{instr} that of the instrumental delays.

Of the four contributions to ϕ_{cal}, the phase due to the first local oscillator is the most rapidly varying. However,

this term is common to all channels within one frequency band (either X- or S-band; see discussion on heterodyning in Section 2.1), and hence may be removed by differencing the calibration phases from two channels. In Figure 3.2.2, we show such differences for the Onsala system, after removing mean differences. We see now that the rapid variations seen in Figure 3.2.1 no longer exist, indicating that most of the variation in that figure was due to the phase changes of the first LO.

We may use these differenced phases to compute the contribution of the phase calibration signals to the group delays by estimating the derivative with respect to angular frequency, of the calibration phase differences (see discussions in Appendix B on the use of the phase calibration phases). These delays will be referred to as the "phase calibration group delays." We show the contribution of the phase calibration signals to the group delays at Onsala in Figure 3.2.3.

We now examine in some detail the phase-calibration results at other sites. In Figure 3.2.4 we show the contribution of the calibration phases to the group delay at Haystack during July 27-28, 1980. (This day was devoted mostly to the 3C345/NRAO512 difference experiment which will be analyzed in Section 4.3.) We should notice that for the last three hours of this experiment, the phase calibration group delays show an oscillating pattern. The residual phases for baselines involving Haystack for this same interval of time also show a

Figure 3.2.2 Differences between the calibration phases at 8.54 GHz and 8.24 GHz, and the 8.34 GHz calibration phase, respectively, for the observations in Figure 3.2.1. (See Figure 3.2.1 for details of symbols.)

Differenced calibration phase (deg)
8.34 GHz channel minus 8.24 GHz channel
Mean difference 32.2

Differenced calibration phase (deg)
8.34 GHz channel minus 8.54 GHz channel
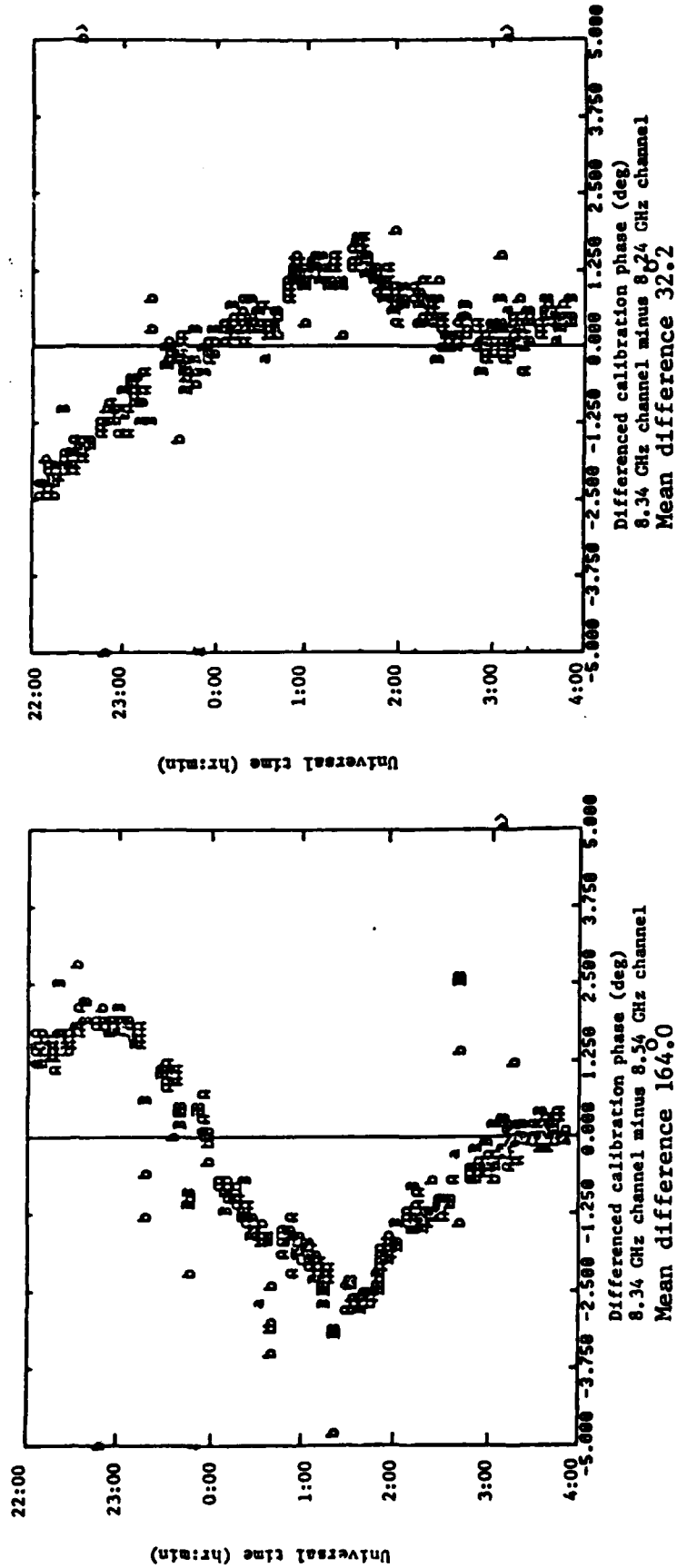Mean difference 164.0

Figure 3.2.3 Contribution of the calibration phases to the group delays for baselines to Onsala (see Figure 3.2.1 for details of symbols).
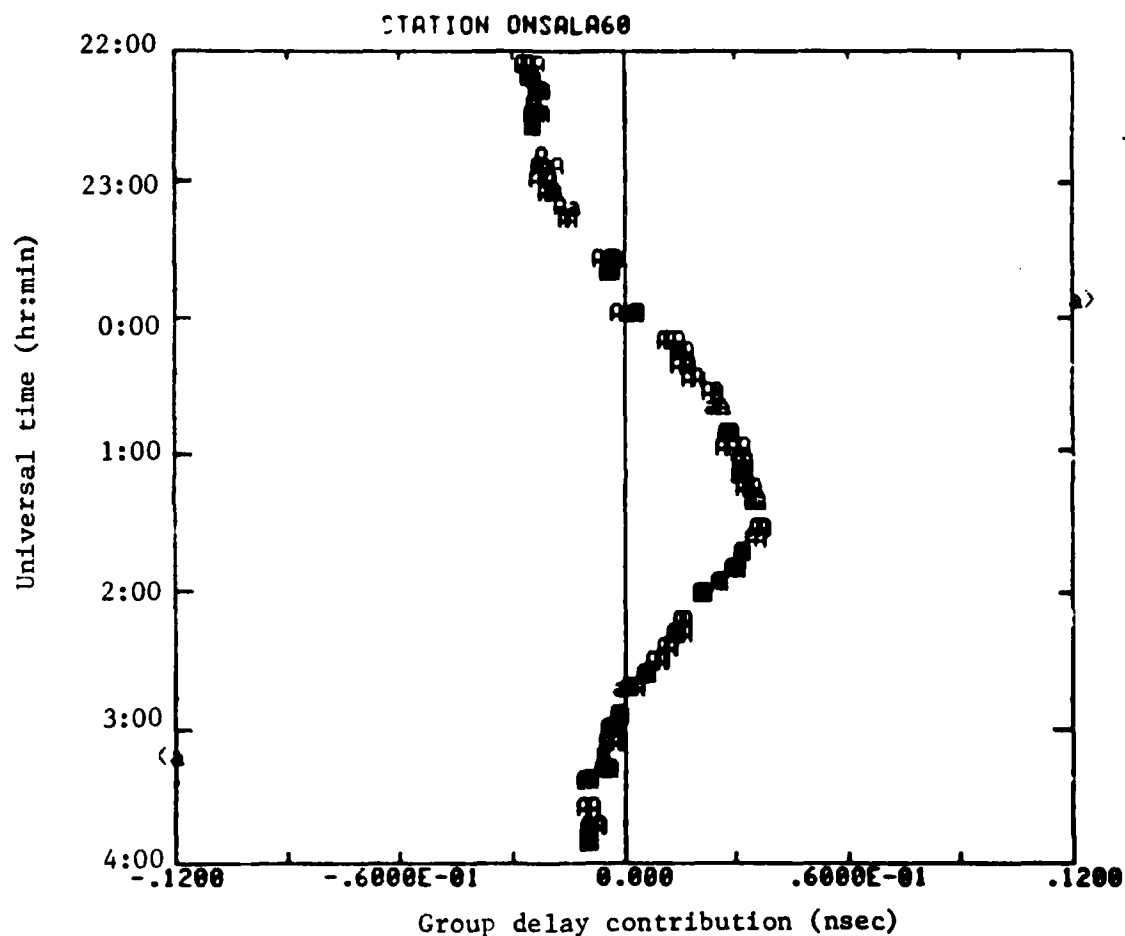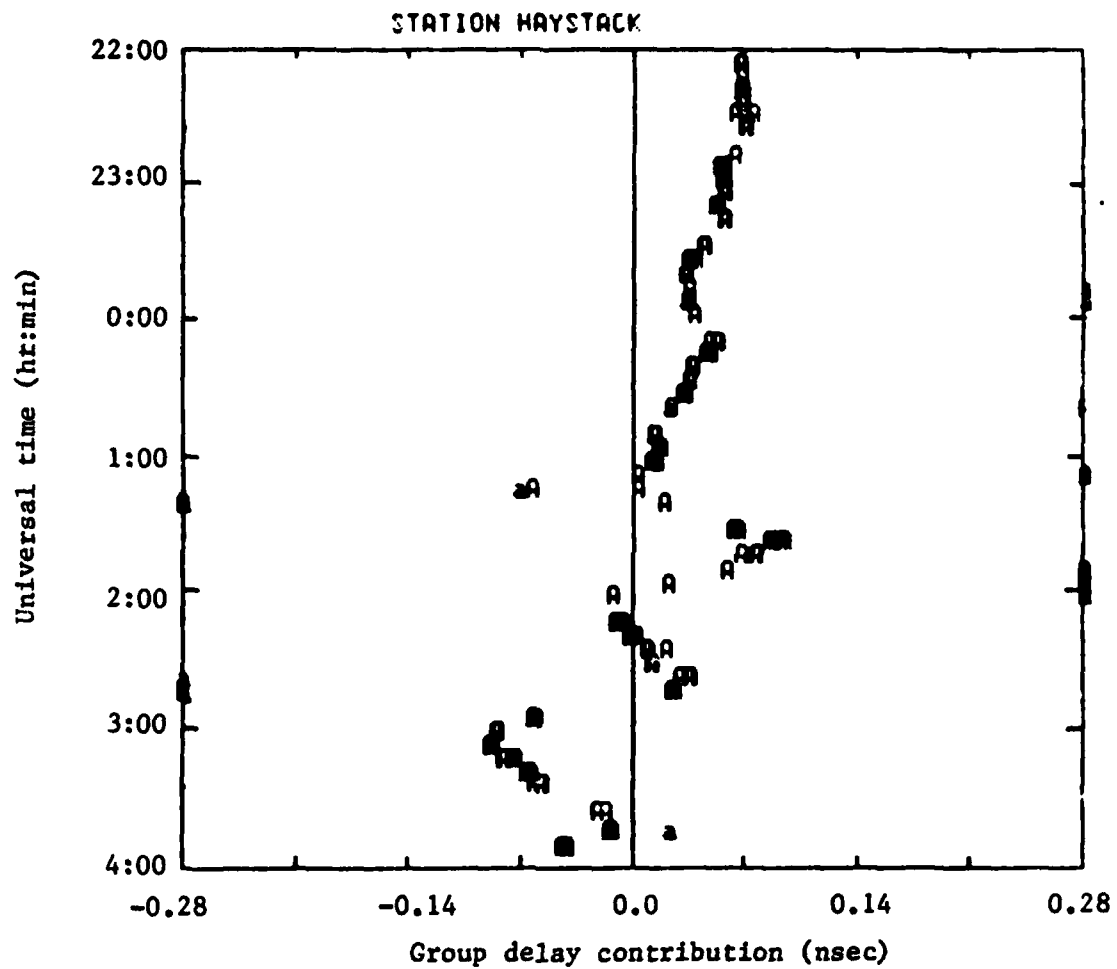
Figure 3.2.4 Contribution of the calibration phases to the group delays
for baselines to Haystack (see Figure 3.2.1 for details of
symbols).

STATION HAYSTACK

similar behavior. When these data were further analyzed, it was found that both the residual phase data and the oscillation of the phase calibrated group delays, were artifacts of the phase calibration system, i.e. there were errors in the calibrations not in the quantities which were being calibrated.

The major evidence for this conclusion is shown in Figure 3.2.5. We plot for each frequency channel, the amplitude of the phase calibration signal as a function of its phase. There are clearly sinusoidal variations of the amplitude with the signal phase. This type of pattern is consistent with the addition of a coherent signal to the phase calibration signal. We show graphically the effects of such an additional signal in Figure 3.2.6. (We shall call this additional signal, the "spurious signal.") If we assume that the spurious signal has a constant amplitude and phase relative to the phase calibration signal we can write an expression for the observed amplitude as a function of the phase of the calibration signal. From the geometry of Figure 3.2.6, we have

$$A^2_{obs} = A^2_{cal} + A^2_s + 2A_{cal}A_s\cos(\phi_{cal}-\phi_s) \quad , \qquad (3.2.2)$$

where A_{obs}, A_{cal}, and A_s are the observed calibration amplitudes, the actual calibration amplitude, and the spurious signal amplitude, respectively, and ϕ_{obs}, ϕ_{cal}, and ϕ_s are the observed calibration phase, the actual calibration phase and the spurious signal phase, respectively. We now wish to develop an algorithm which will allow us to calculate A_s and

Figure 3.2.5 Amplitude of the phase calibration signal at Haystack as a function of the phase of the calibration signal (see Figure 3.2.1 for details of symbols).
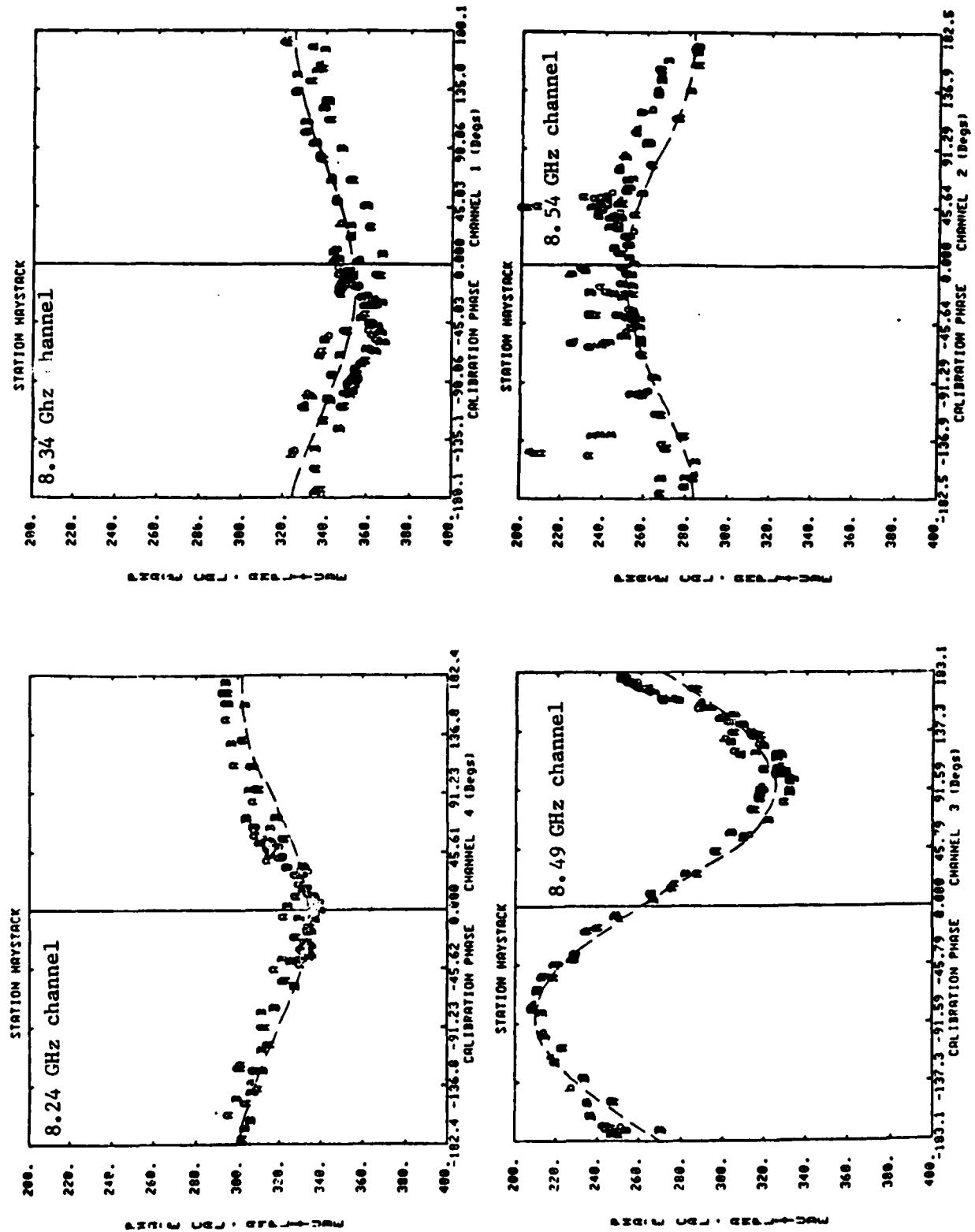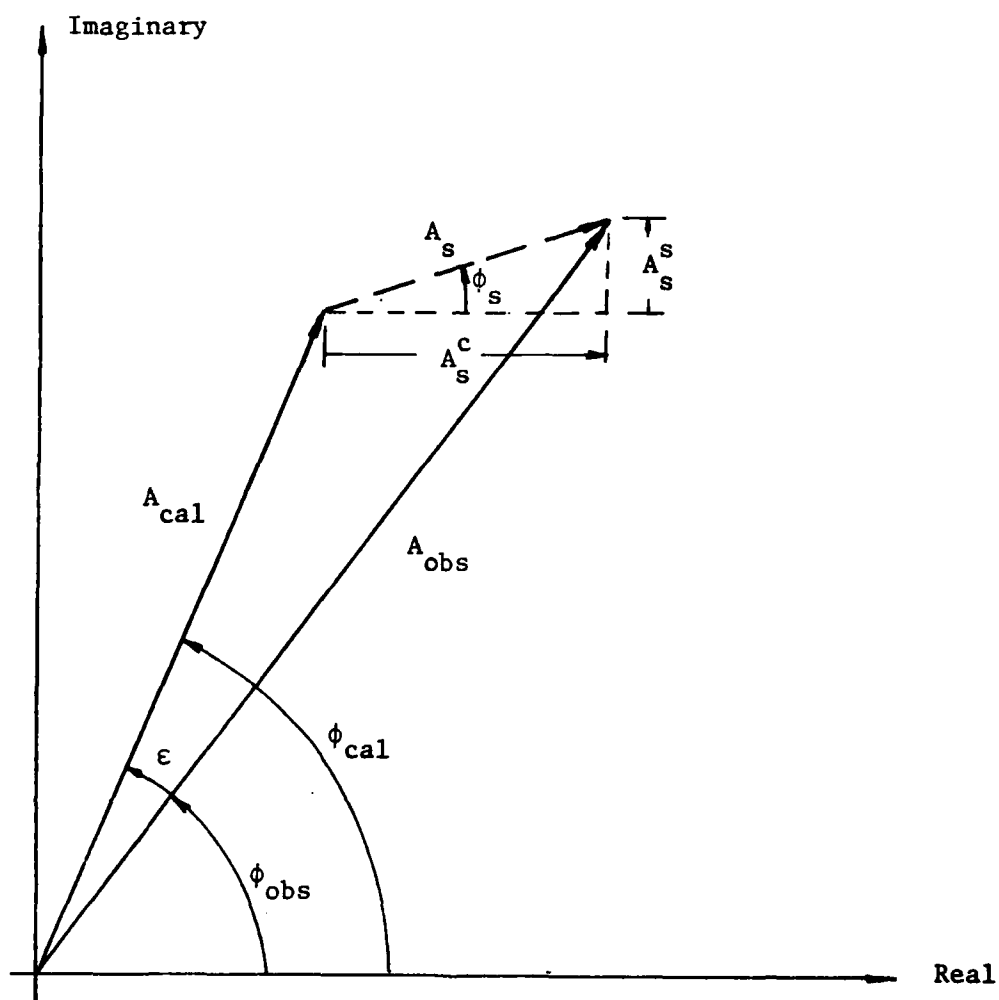
Figure 3.2.6 Graphical representation of the effects on the calibration signal of a spurious signal with constant amplitude and phase.

ϕ_s from a set of A_{obs}'s at different values of ϕ_{obs}. Once A_s and ϕ_s are known we can then reconstruct A_{cal} and ϕ_{cal} from the observed calibration signals.

If we assume that $A_s \ll A_{cal}$, then we can linearize Equation (3.2.2), which will make the estimation of A_s and ϕ_s much simpler. The linearization of Equation (3.2.2) yields

$$A_{obs} \simeq A_{cal} + A_s \cos(\phi_{cal} - \phi_s) \quad ,$$

which may be rearranged to yield,

$$A_{obs} \simeq A_{cal} + A_s^C \cos(\phi_{cal}) + A_s^S \sin(\phi_{cal}) \quad , \qquad (3.2.3)$$

where $A_s^C = A_s \cos\phi_s$ and $A_s^S = A_s \sin\phi_s$. If $A_s \ll A_{cal}$, then ε, the error in the calibration phase, will be small, and we may approximately equate ϕ_{cal} and ϕ_{obs} in the cosine and sine terms in Equation (3.2.3). This substitution yields

$$A_{obs} \simeq A_{cal} + A_s^C \cos(\phi_{obs}) + A_s^S \sin(\phi_{obs}) \quad . \qquad (3.2.4)$$

Also based on the geometry of Figure 3.2.6, we may write an expression for the error in the calibration phase, ε. Hence

$$\varepsilon \simeq \tan^{-1}\left[\, (A_s^C \sin\phi_{obs} - A_s^S \cos\phi_{obs})/A_{cal}\right] \qquad (3.2.5)$$

and the corrected calibration phase will be given by

$$\phi_{cal} = \phi_{obs} + \varepsilon \quad . \qquad (3.2.6)$$

In Figure 3.2.5, we have superimposed the expected amplitude variations with changes in the phase calibration signals. (The differences between the predicted curve for the amplitude versus phase and the observed values, for the 8.34 GHz chan-

nel, are due to a variation of the calibration amplitude with
time; see discussion in Section 2.3.)

The Haystack radio telescope was not the only antenna to
be affected by a spurious signal during the July 1980 experi-
ment. The Onsala antenna also was affected. We show the
phase calibration signal amplitude versus phase values for
these observations in Figure 3.2.7 Again we see a very dis-
tinctive dependence of the amplitude on the phase, especially
in the 8.34 GHz and 8.54 GHz channels.

We show in the Figure 3.2.8, the phase calibration group
delays at Haystack and at Onsala before and after correcting
for the spurious signals. We notice that, at Onsala, the
phase calibration contribution to the group delays is almost
totally removed when the correction for a spurious signal is
made. At Haystack, however, there still appear to be residual
oscillations in the later part of the experiment. There are
several reasons for this behavior. Firstly, the formulation
we presented for estimating the magnitude of, and correcting
for, a spurious signal was linearized. The ratio of the
spurious signal strength to the calibration signal strength in
the 8.54 GHz channel at Haystack was 0.21. This ratio is
sufficently large for the linearization to introduce sizable
errors. (The second order terms will be $O\left((0.21)^2\right)$, implying
that the correction could be in error by as much as $\simeq 20\%$.)
Secondly, we have assumed that the spurious signal is constant
in both amplitude and phase; this need not be the case. A
comparison of Figures 3.2.8.a and b indicates that at Haystack

Figure 3.2.7 Phase calibration signal's amplitude as a function of its phase
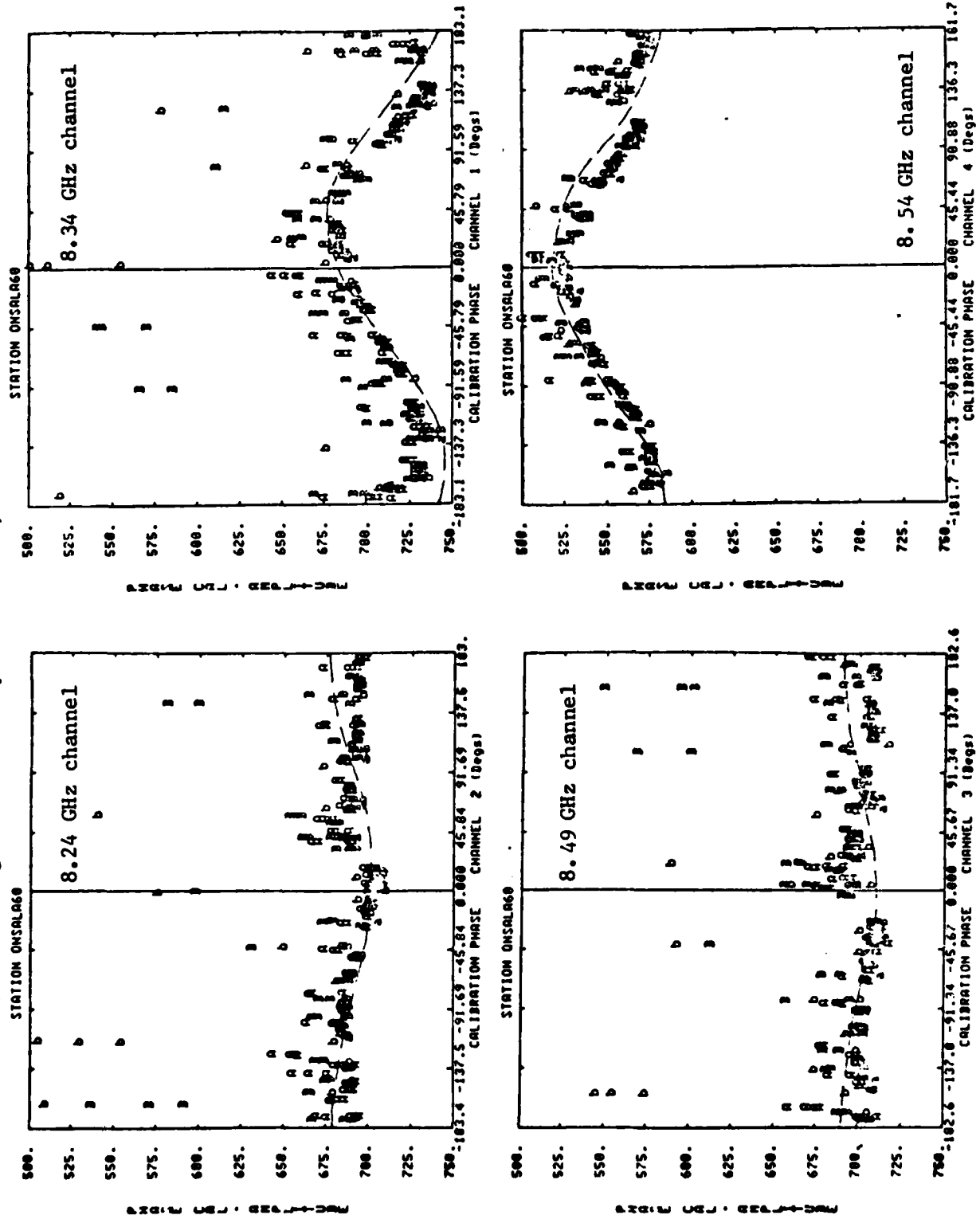(same as Figure 3.2.5 except for Onsala)

Figure 3.2.8 The effects of the spurious signal on the phase calibration group delay (see Figure 3.2.1 for details of symbols).
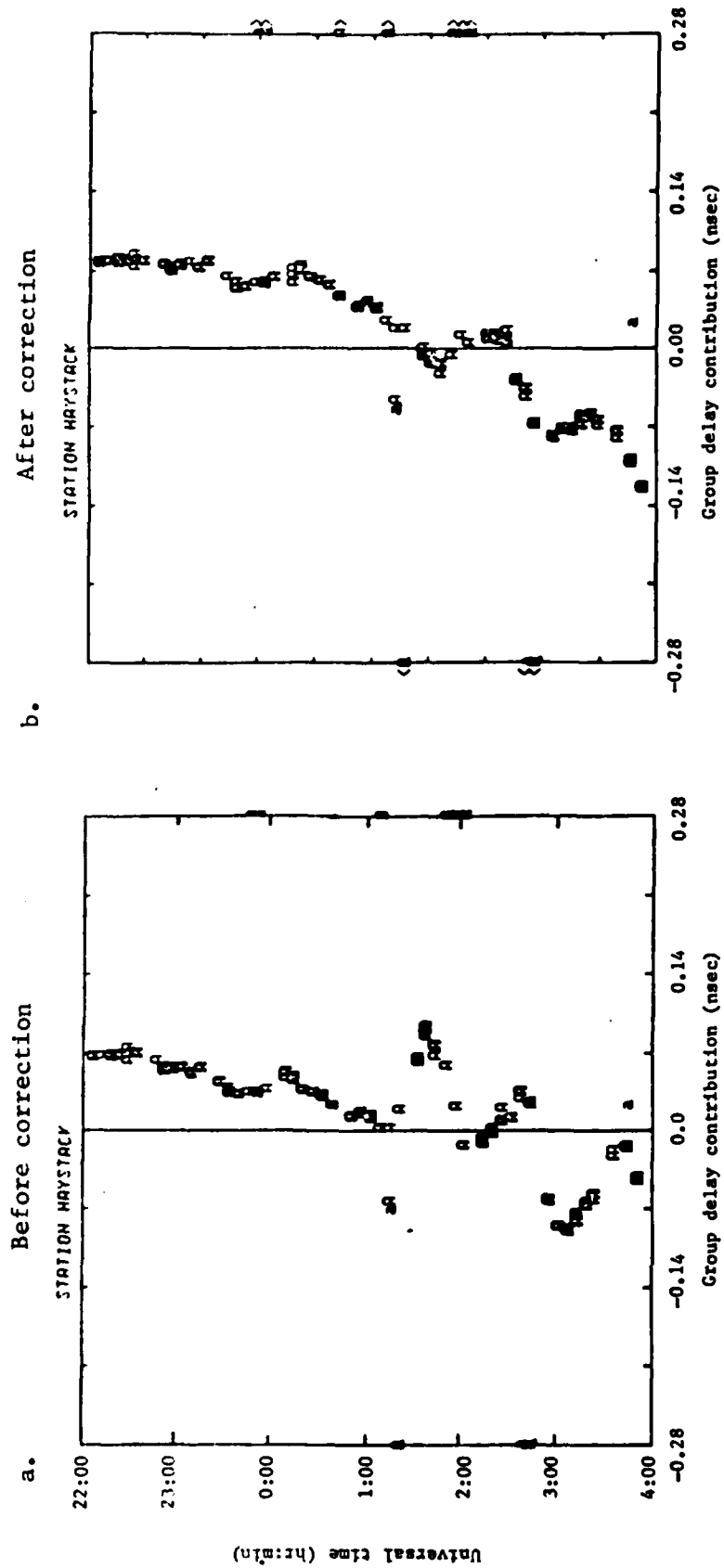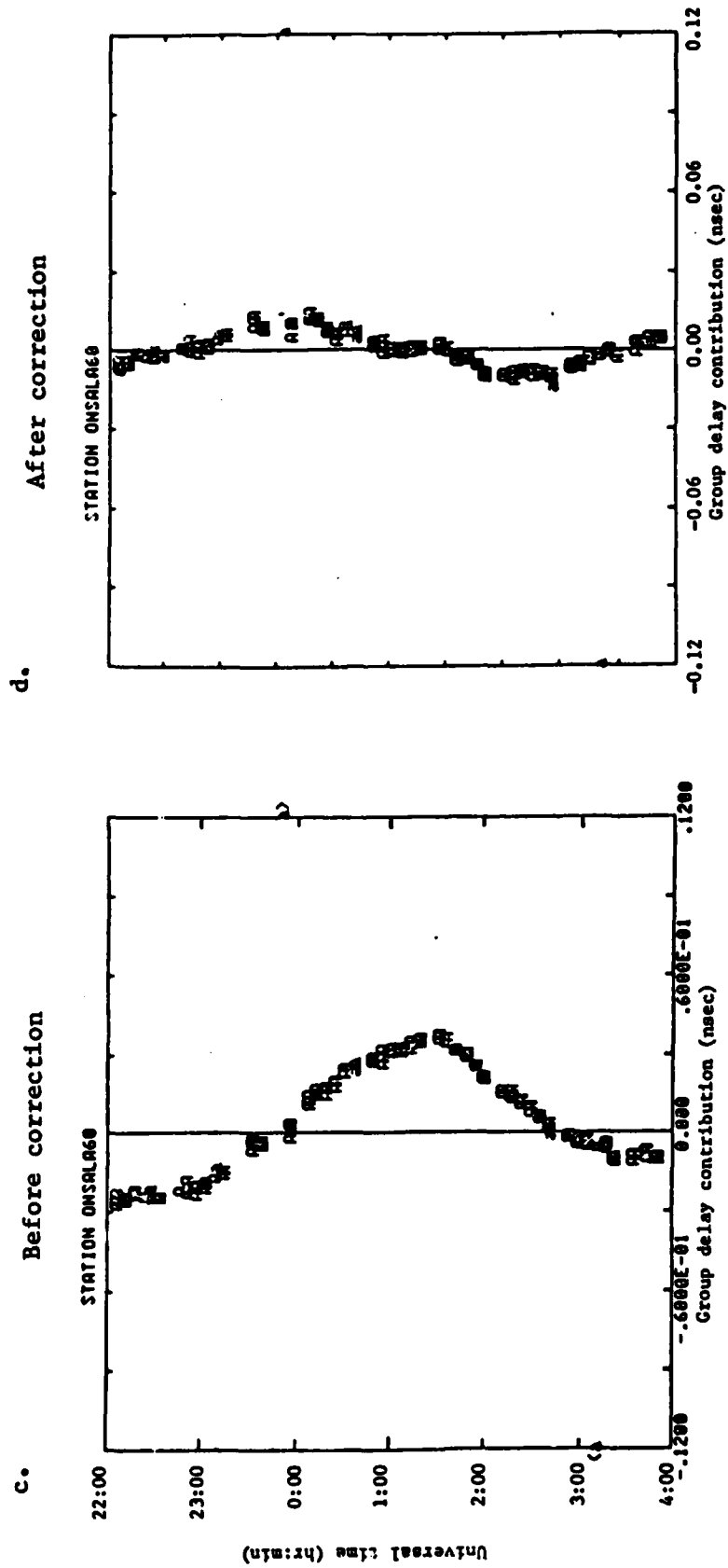
Figure 3.2.8 Continued

the spurious signal may not be constant, i.e., at times the correction seems to be too small and at other times too large, indicating that the spurious signal may be varying with time.)

The large magnitude of the spurious signal at Haystack seems to be anomalous (the effect of the spurious signal is three times larger at Haystack than at Onsala). Consequently, we have not developed any software which will handle the non-linear problems or the possibility of time variations of the spurious signal.

In Figure 3.2.9, we show the residual phases at Haystack before and after the correction for the presence of the spurious signal. Again, we see that the systematic trends in the residual phases are lessened by the correction, but not totally removed.

We have so far discussed the detection and correction of the spurious signal effects, but we have not discussed its origin. Unfortunately, the exact origin of the signal has not been determined, although there are many possibilities. The reason for the many possible origins is simple: almost every electronic component in a radio telescope is controlled by the 5 MHz signal generated by the hydrogen maser, e.g., the pulse generator, the local oscillators in the receiver and the video converters, and the time synchronization of the tape recorders. Consequently, there is a high probability that one or more of these pieces of equipment generated the spurious signal.

After the spurious signal was detected, guidelines were

-112-

Figure 3.2.9   The effects of the spurious signal on the Haystack residual phases
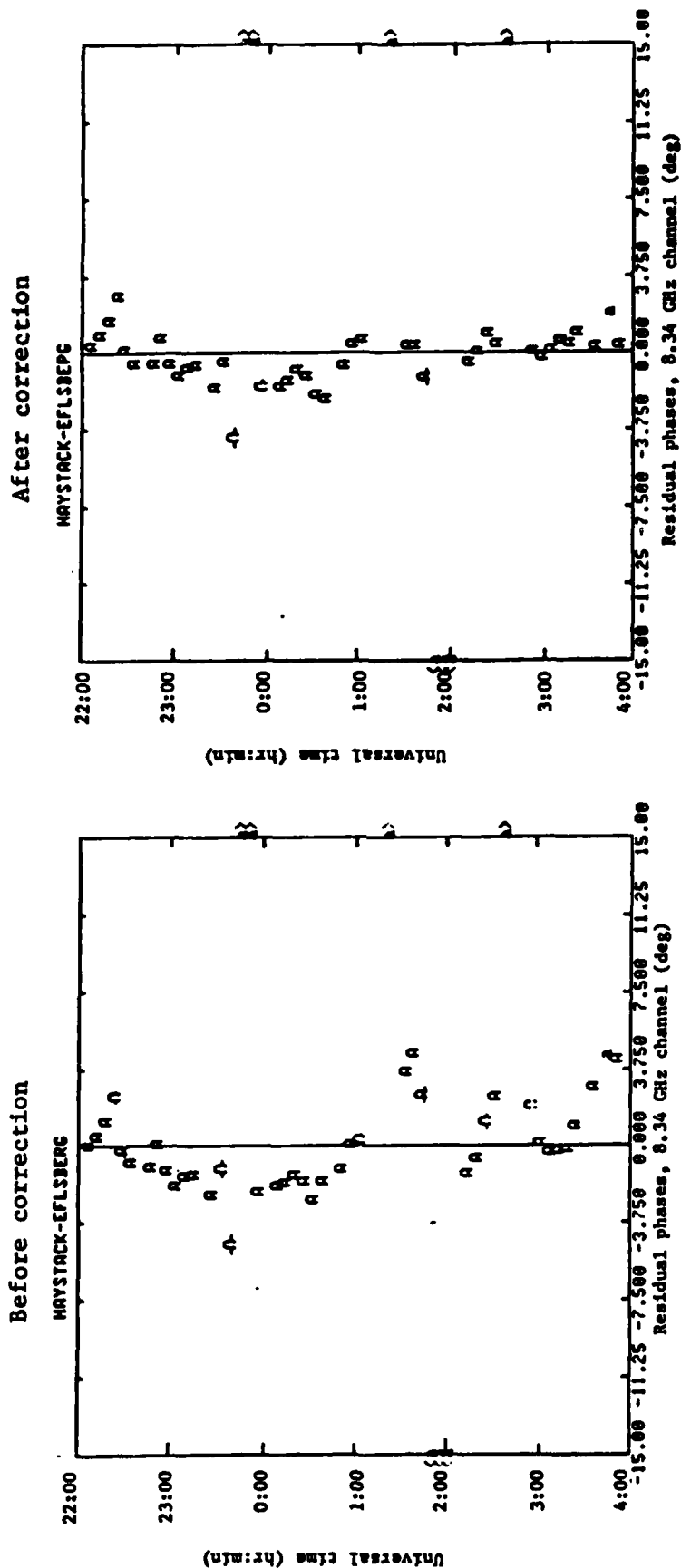(see Figure 3.2.1 for details of symbols).

Figure 3.2.9 Continued.
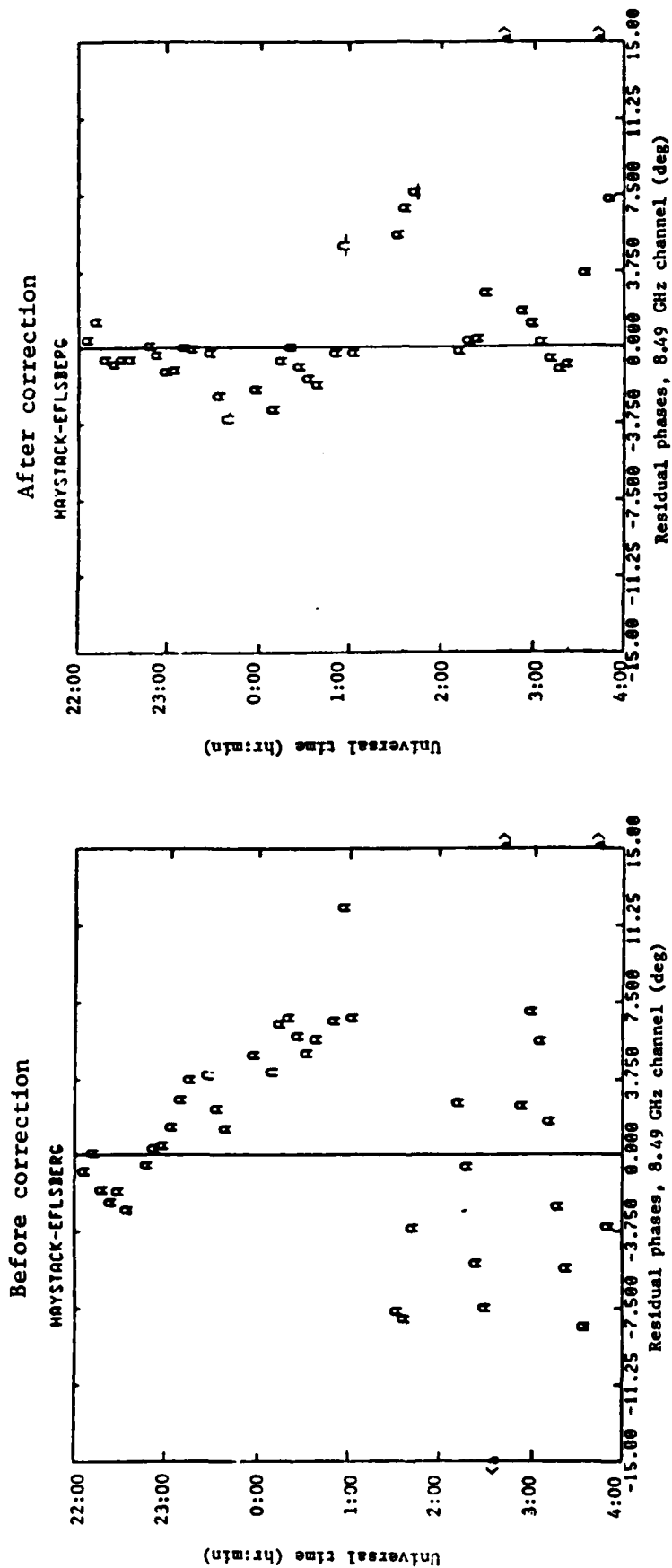
set up to check for the presence of spurious signals in the
output of the receiver and to ensure better isolation of the
different pieces of equipment (Rogers, 1981). The effective-
ness of these guidelines will be assessed when data taken
after the guidelines were issued can be fully analyzed. An
initial analysis of the two multibaseline intercontinental
experiments (June 1982 and December 1982), performed after the
introduction of the guidelines, has not detected any spurious
signals.

In this section, we have tried to demonstrate methods for
studying the Mark III performance by analyzing the system's
calibration data. The spurious signal problem indicates the
care that must be taken when accurate observations are
attempted with a complex piece of equipment. We will consider
the behavior of the phase calibration system again in Section
4.3 when we are analyzing the prediction of phase delay
ambiguities from group delays.

3.3 Group delay closures

We now discuss the final quality check which we will
investigate in this chapter -- the properties of the sum of
the delays around triplets of baselines. These sums will be
referred to as delay closures. In this section we will study
the closures of the group delays because these are the delays
which are used in geodetic data processing. As with earlier
sections in this chapter, we will study some specific examples

of this quality check, but we will not be able to present all
of the results which have been obtained thus far.

Delay closures around triplets of baselines have been
used as a quality check since the first multibaseline Mark I
VLBI experiments (Whitney, 1974, pp. 276-279). The delay
misclose, $\Delta\tau_m$, around a triplet of baselines is given by
(ibid, p. 277)

$$\Delta\tau_m = \hat{\tau}_{12} + \hat{\tau}_{23} - \hat{\tau}_{13} + \hat{\tau}_{12}\hat{\dot{\tau}}_{23} \tag{3.3.1}$$

where $\hat{\tau}_{12}$, $\hat{\tau}_{13}$, $\hat{\tau}_{23}$ are the observed delays (either phase or
group) for the three two-element interferometers formed by the
pairs of tne three sites, and $\hat{\dot{\tau}}_{23}$ is the observed phase delay
rate for the interferometer formed by Sites 2 and 3. The
final term in Equation (3.3.1) accounts for the epoch of the
delay measurement being referred to the time of arrival of a
signal at the "first" site in an interferometer (see Whitney,
1974, p. 276, and Appendix B). The method we have adopted to
evaluate Equation (3.3.1) is not direct computation. We have
instead used the property of the theoretical delays computed
by VLBI analysis software; these delays sum to zero around a
triplet of baselines (when they are referred to the same
epoch), i.e.

$$\tau^t_{12} + \tau^t_{23} - \tau^t_{13} + \tau^t_{12}\dot{\tau}^t_{23} \equiv 0 \tag{3.3.2}$$

where the superscript t refers to the theoretically calculated
delays. If we now subtract Equation (3.3.2) from Equation

(3.3.1), we obtain

$$\Delta\tau_m = (\hat{\tau}_{12}-\tau^t_{12}) + (\hat{\tau}_{23}-\tau^t_{23}) - (\hat{\tau}_{13}-\tau^t_{13}) + (\hat{\tau}_{12}-\tau^t_{12})\hat{t}_{23}$$

$$+ (\hat{t}_{23}-t^t_{23})\tau^t_{12}.$$

If we now introduce the postfit residuals, $v_{ij}\equiv(\hat{\tau}_{ij}-\tau^t_{ij})$, then the above equation reduces to

$$\Delta\tau_m = v_{12} + v_{23} - v_{13} + v_{12}\hat{t}_{23} + \dot{v}_{23}\tau^t_{12} \qquad (3.3.3)$$

The final two terms in the above equation are very small and can be neglected in the calculation of the closures. (For $v_{12}=1.0$ nsec, $\dot{v}_{23}=1.0$ psec/sec, the maximum contribution to the closure calculation of the final two terms will be 0.002 psec and 0.020 psec when $\dot{t}_{23}=2.0\times10^{-6}$ sec/sec and $\tau^t_{12}=0.02$ sec. The latter two values are upper bounds for Earth based VLBI.)

When the final two terms of Equation (3.3.3) are neglected, the calculation of the delay miscloses reduces to simply summing the postfit residuals around baseline triplets.

In Figure 3.3.1, we show the delay closures at both X- and S-band, for the radio source 3C 345, on two triplets of baselines. The error bars shown are calculated from the sum of the variances of delays from each baseline in the triplet and are shown at typical points. We show in Appendix C that the variance of the misclose calculated in this manner should be an overestimate, due to correlations between the observations on each baseline, by less than 1%.
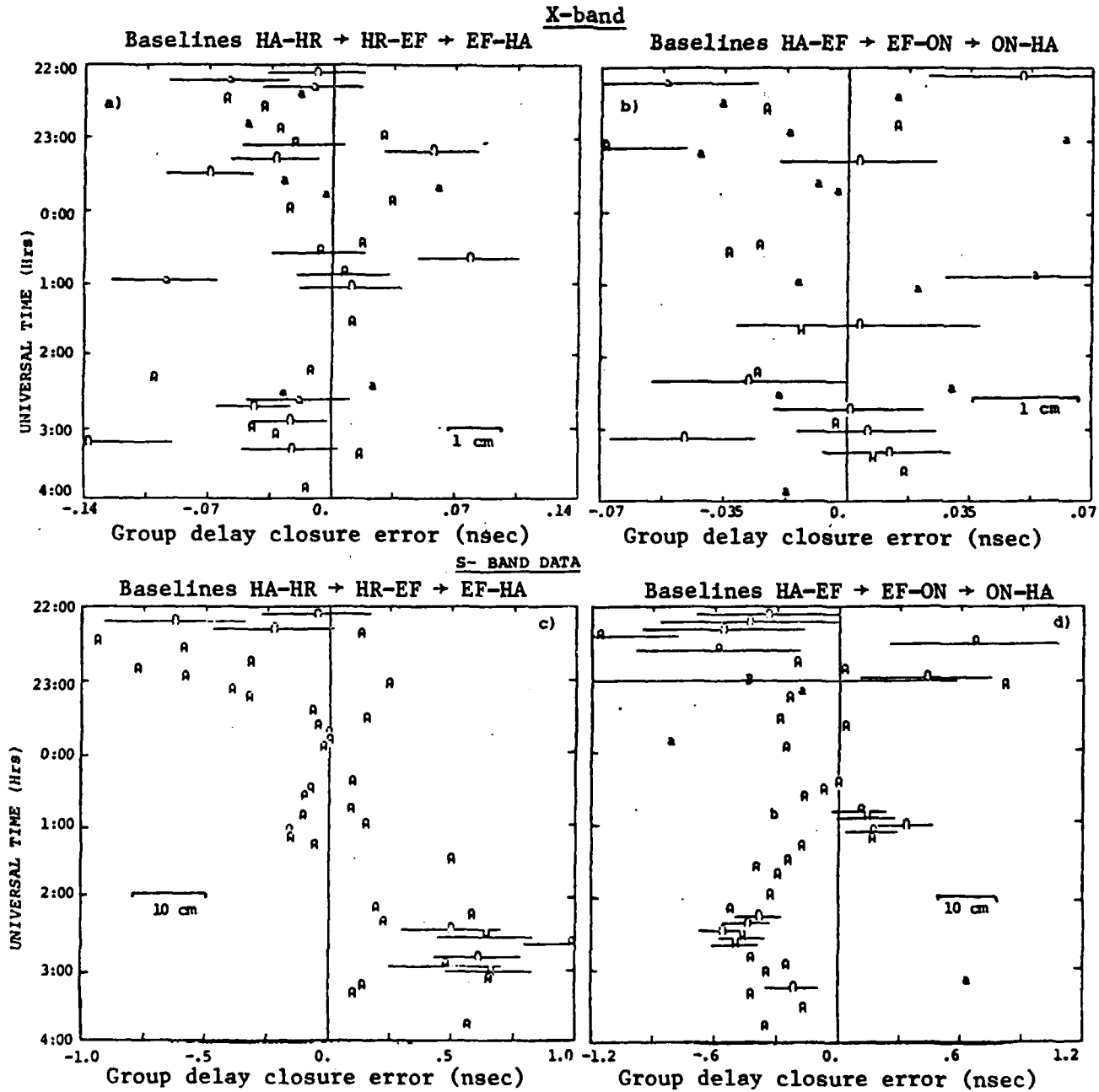
The X-band miscloses shown in Figure 3.3.1 do not show
any particular trends although their scatter is 1.3 times
greater than we would expect from the standard deviations.
The S-band results, however, do show systematic trends. These
miscloses could be due to the brightness distribution of the
radio source (see, e.g., Cotton, 1979; Shapiro et al., 1979).
However, currently available brightness distributions for this
source (Breidenthal, 1982) do not remove these trends. The
origin of these miscloses is still being investigated (see
below). In Section 4.3, we will investigate the complications
these miscloses will cause when we attempt to predict phase
delay ambiguities from group delays.

To determine a possible origin for the S-band group delay
closure errors we need to look at the Mark III hardware. In
Appendix F we develop a model for the effects of "leakage" of
left-circularly polarized radiation into the output of a feed
horn which is nominally meant to receive right-circularly
polarized radiation. (This model was originally suggested by
Dr. A.E.E. Rogers at the Haystack Observatory.) The contri-
bution of the left-circularly polarized radiation to the group
delays, $\Delta\tau_{\ell}$, is (for an unpolarized radio source; see Equation
F.5),

$$\Delta\tau_{\ell} = \Delta\tau_0^p \sin(2\Delta\psi + \xi_0^p)$$

where $\Delta\tau_0^p$ is a nominally constant delay which is function of
(i) the ratio of the power from the left-circularly polarized
to the power from right-circularly polarized radiation at the

Figure 3.3.1 Group delay closures for observations of radio source
3C 345 on July 27-28, 1980.

X-band

Baselines HA-HR → HR-EF → EF-HA   Baselines HA-EF → EF-ON → ON-HA

S- BAND DATA

Baselines HA-HR → HR-EF → EF-HA   Baselines HA-EF → EF-ON → ON-HA

Error bars are shown only at typical points.

Symbols HA Haystack, EF Effelsberg, HR HRAS 085, ON Onsala (See Table 5.1).

output of the feed horn, and (ii) the spanned bandwidth of the group delay measurement (see Appendix F for details); $\Delta\psi$ is the difference of the feed horn position angles at each site in the interferometer (see Appendix D for details); and ξ_o^p is the phase difference between the left- and right-circularly polarized radiation after these signals have propagated through the feed horns at the two sites in the interferometer.

In Figure 3.3.2 we show an approximate matching of the polarization leakage model to the group delay closure errors for the Haystack-Effelsberg-Onsala triplet from the 27-28 July 1980 experiment, assuming $\Delta\tau_o^p = 0.5$ nsec (approximately estimated) for the Haystack-Onsala interferometer. The polarization leakage model could explain many of the apparent errors in the group delays and now needs to be thoroughly investigated.

Figure 3.3.2 The effects of feed horn imperfections on the Haystack-Effelsberg-Onsala group delay closure errors for the date given in Figure 3.3.1.
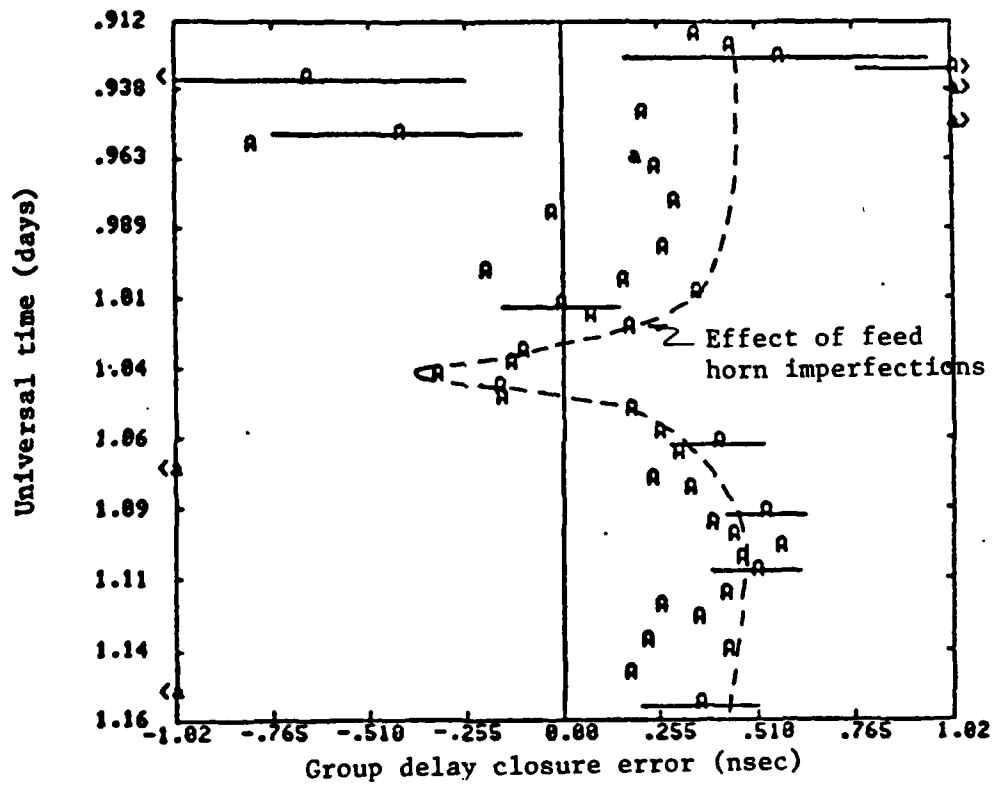
Feed horn imperfection effect computed using

$$\Delta\tau_o^P = 0.5 \text{ nsec and } \xi_o^P = 90^o$$

(see text and Appendix F for discussion)

Error bars are shown only at typical points.

-121-

4. ELIMINATION OF PHASE DELAY AMBIGUITIES

In Chapter 3 we investigated the quality of the Mark III
VLBI estimates of the group delay, phase delay and phase
delay rate using the redundant nature of the VLBI observ-
ation. In this chapter we will continue to check observation
quality by attempting to combine the group and phase delay
measurements to produce a delay measure which has the precis-
ion of the phase delays and is calibrated for the plasma
delay through the ionosphere. In Chapter 2 we discussed the
methods used to estimate both the group and phase delays from
the recordings of signals from a radio source. The phase
delay was however ambiguous by multiples of $2\pi/\omega$ seconds.
(The multiband group delay is also ambiguous, but the
ambiguity spacing is much larger for this delay measure than
for the phase delays. The group-delay ambiguities can
normally be resolved using a priori information.) In this
chapter we will investigate methods which could be used to
remove the phase-delay ambiguities.

The sensitivity of the Mark III VLBI system allows group
delay measurements to be made with uncertainties, due to
random noise, sometimes as low as a few picoseconds and quite
often lower than 30 psec. This sensitivity is achieved by
using wideband receivers (400 MHz at X-band, $\approx$8.4 GHz, and
100 MHz at S-band, $\approx$2.2 GHz) and multi-track 2 MHz-bandwidth-
per-track tape recorders. With group-delay uncertainties of
less than 30 psec at X-band and 100 psec at S-band it should

be possible to determine estimates of the phase delays at
X-band and S-band which are sufficiently accurate that we can
resolve the phase-delay ambiguities of ≈120 psec at X-band
and ≈440 psec at S-band. However, the signals which have
been recorded at each site have propagated through a disper-
sive medium, mainly the Earth's ionosphere, and therefore the
group and phase delays will not be equal. In order to be
able to predict the phase delays we will have to calibrate
the dispersive delay. In Section 4.1 we will investigate the
models which are used to calibrate the dispersive delay. We
will consider both the approximations made in using the
models and the methods which should be used to apply these
models to wideband VLBI observations.

The prediction of phase-delay ambiguities, based solely
on group delay observations, is an experimental technique.
We therefore require a method to check its validity. With
this in mind, Section 4.2 will concentrate on developing the
prediction algorithms using observations which can be "phase
connected" without the use of the group delays. These data
which usually are observations of a pair of radio sources,
closely spaced in the sky, may be "phase connected" because
of the short duration of time (<8 mins) between observations
of each of the sources. Over this short interval of time, we
can model the behavior of the phase to within much less than
a cycle allowing sequential observations to be "phase
connected" to each other, provided there are no large data
gaps. This technique has been used successfully in the past

to obtain phase delays (Wittels, 1975; Shapiro et al., 1979). In Section 4.3, we will attempt to apply the techniques developed in Section 4.2 to a standard geodetic schedule.

4.1 Ionospheric propagation delay models

In this section we will derive the frequency dependence of the delay caused by the propagation of a signal through the ionosphere. While deriving the frequency dependence we will neglect several effects which describe the interaction between the ionosphere and the propagating electromagnetic waves. In Section 4.1.2 we will investigate the possible magnitude of the errors in the dual frequency calibration system due to the neglect of these additional terms.

We model the ionosphere as a plasma of electrons with a density of N_v electrons/m^3 which are in motion around fixed positive ions. The macroscopically averaged (i.e., averaged over a volume large compared to the volume of an atom) electric dipole, $\bar{P}_t$ (see, e.g., Jackson, 1975, Section I and Chapter 4.1), for a plasma is:

$$\bar{P}_t = - N_v e \bar{x} , \qquad (4.1.1)$$

where e is the charge of an electron (1.6×10^{-19} Coul) and $\bar{x}$ is the average (vector) displacement of the electrons from their fixed ions. In the presence of an electric field $\bar{E}_t$, the equation of motion of a free electron in the ionosphere

is given by (Jackson 1975,p. 285 and p. 292):

$$m_e \ddot{\bar{x}} - (e/c)\ \bar{B}_0 \times \dot{\bar{x}} - m_e \gamma e \dot{\bar{x}} = -e\bar{E}_t \qquad (4.1.2)$$

where $\dot{\bar{x}}$ and $\ddot{\bar{x}}$ are the average velocity and the average
acceleration of the electron, m_e is the mass of the electron
$(9.11 \times 10^{-31}$ kg), c is the velocity of light, $\bar{B}_0$ is the mag-
netic induction of the Earth's magnetic field and γ is a
damping constant (we will discuss the meaning of the damping
constant in Section 4.1.2.). If we neglect the effects of
the magnetic field terms and damping terms in Equation
(4.1.2), the equation of motion reduces to

$$m_e \ddot{\bar{x}} = -e\bar{E}_t \ .$$

Incorportion of Equation (4.1.1) into the above equation
yields a relationship between the electric dipole and the
electric field given by

$$\frac{m_e}{N_v e}\ \ddot{\bar{P}}_t = e\bar{E}_t \ . \qquad (4.1.3)$$

We will now assume that $\bar{E}_t$ is a time-harmonic field, i.e.
$\bar{E}_t(\bar{x},t) = \bar{E}(\bar{x})e^{-i\omega t}$, where ω is angular frequency. For this
choice of field type, Equation (4.1.3) can be solved yielding

$$\bar{P} = -\frac{N_v e^2}{m_e \omega^2}\ \bar{E} \qquad (4.1.4)$$

where $\bar{P}$ is the positional part of the time-harmonic electric
dipole: $\bar{P}_t(\bar{x},t) = \bar{P}(\bar{x})e^{-i\omega t}$.

Equation (4.1.4) can now be used to determine the per-
mittivity tensor, $\tilde{\varepsilon}$, in region $\bar{x}$. The permittivity tensor

relates the electric field $\bar{E}$ to the electric displacement field $\bar{D}$ by $\bar{D} = \tilde{\epsilon}\bar{E}$. The electric displacement field is also given by $\bar{D} = \epsilon_0\bar{E} + \bar{P}$ (Jackson, 1975, p. 14), where ϵ_0 is the permittivity of free space. The properties of the ionosphere when modeled by Equation (4.1.3) are isotropic and the permittivity tensor reduces to $\tilde{\epsilon} = \epsilon\tilde{I}$ where $\tilde{I}$ is a unit tensor and ϵ is the dielectric constant of the ionosphere. The combination of the two expressions for the electric displacement field yields the dielectric constant for the ionosphere (remember the expression for this dielectric constant is only approximate because of the terms in Equation (4.1.1) which were neglected):

$$\epsilon = \epsilon_0\left(1 - \frac{N_v e^2}{m_e \epsilon_0 \omega^2}\right) \tag{4.1.5}$$

We may now use the dielectric constant to estimate the phase velocity of electromagnetic waves propagating through the ionosphere. For plane waves, the electric field is given by $\bar{E}_0 \exp[i\bar{k}\cdot\bar{x} - i\omega t]$, where $\bar{k}$ is the wave vector. For a wave to propagate through a medium, the dispersion relationship between $\bar{k}$ and ω must be satisfied. For plane waves propagating through an isotropic medium the dispersion relationship between $\bar{k}$ and ω is (ibid p. 271):

$$|\bar{k}|^2 = \omega^2 \mu_0 \epsilon$$

where μ_0 is the permeability of free space. The phase velocity of the wave is given by $V_p = \omega/|\bar{k}|$. The combination of Equation (4.1.5), the dispersion relationship and the

definition of phase velocity yields

$$V_p = c \left[1 - \frac{N_v e^2}{m_e \epsilon_o \omega^2} \right]^{-1/2} \tag{4.1.6}$$

where $\underline{c}$ is the free space velocity of propagation ($c^2 = 1/\mu_o \epsilon_o$). We can write the expression for the phase refractive index of the ionosphere, n_p, directly from Equation (4.1.6)

$$n_p = \left[1 - \frac{N_v e^2}{m_e \epsilon_o \omega^2} \right]^{1/2} = \left[1 - \omega_p^2/\omega^2 \right]^{1/2} \tag{4.1.7}$$

where we have substituted $\omega_p^2 = N_v e^2/m_e \epsilon_o$ in the second equality in Equation (4.1.7); ω_p is the plasma frequency.

Equation (4.1.7) gives the expression for the phase refractive index of the ionosphere. Before evaluating the expression for the propagation delay through the ionsphere we will make one further approximation in the expression for the phase refractive index; we will expand the square root in Equation (4.1.7) to first order, i.e. we will approximate n_p by

$$n_p \approx 1 - \omega_p^2/2\omega^2 . \tag{4.1.8}$$

(The effects of the neglect of the second order term in the expansion will be studied in Section 4.1.2.)

The difference in the propagation delay between the free-space propagation and the propagation through the ionosphere will be given by

$$I^p = \int_0^{s_o} (n_p(s)/c) \, ds - \int_0^{s_o} (1/c) \, ds \tag{4.1.9}$$

where the integration path is shown in Figure 4.1.1. (The variation of n_p with height, shown in this figure, is only meant to illustrate the general behavior of n_p. The values of n_p along the ray path should be used to evaluate the integrals in Equation (4.1.9).) We will refer to I^p as the "ionospheric phase delay". (Strictly the line integral given above is only valid if the the wavelength of the variations of $n_p(s)$ are large compared to the Fresnel-zone diameter at the height of the ionosphere. We will discuss the Fresnel-zone diameter in detail in Section 4.1.2.)

Substitution of Equation (4.1.8) into the above integral yields

$$I^p = \frac{-e^2}{2cm_e\epsilon_o\omega^2} \int_0^{s_o} N_v ds$$

We will define the integrated electron content N_i to be

$$N_i = \int_0^{s_o} N_v \, ds$$

which yields the expression for the ionospheric phase delay

$$I^p = -\frac{e^2 N_i}{2cm_e\epsilon_o\omega^2} \tag{4.1.10}$$

The variations of N_v with altitude, location, time of day and time of year are too erratic for N_i to be accurately determined using (non-line-of-sight) measurements or predictions of N_v. (Measurements of the integrated electron content could be used, if these measurements were made at the radio telescopes and in the direction of the radio source.)

Figure 4.1.1 Integration path for computing the ionospheric delay

from
radio
source

n(h)

0

IONOSPHERE

ds

integration
path
(the path along which the
signal propagate)

h

s_o

to radio telescope
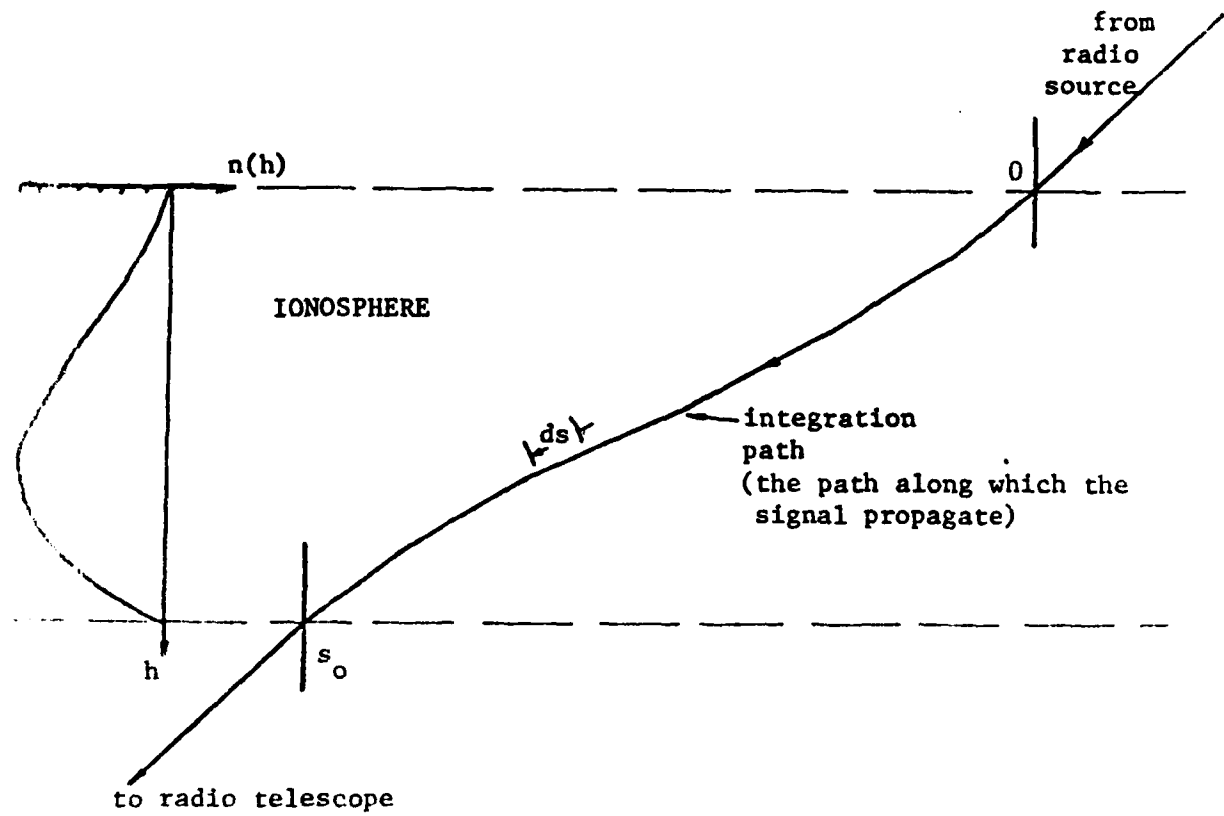
However, Equation (4.1.10) suggests a method we can use to avoid directly determining N_i. Measurements of the propagation delay through the ionosphere at two widely spaced frequencies can be used to infer N_i. This technique is used with dual frequency Mark III receivers to calibrate the differential path delay through the ionosphere.

We may derive the "ionospheric group delay" from Equation (4.1.10). If we multiply Equation (4.1.10) by ω we obtain the change in phase, ϕ^i, due to the propagation through the ionosphere:

$$\phi^i = -\frac{e^2 N_i}{2cm_e \varepsilon_o \omega} \qquad (4.1.11)$$

Differentiation of Equation (4.1.11) with respect to angular frequency will yield the ionospheric group delay, I^g,

$$I^g = +\frac{e^2 N_i}{2cm_e \varepsilon_o \omega^2} \qquad (4.1.12)$$

Strictly we should include a term involving $\partial N_i / \partial \omega$ which accounts for the difference in the integration path at different frequencies. We will investigate this path dependence in Section 4.1.2.

Equations (4.1.10) and (4.1.12) give the frequency dependence of the ionospheric phase and group delays. In the following section we will investigate the method used to calibrate the ionospheric delays using dual frequency measurements. In Section 4.1.2. we investigate the magnitude of the errors in the dual frequency calibration due to the

approximations made in the ionospheric delay derivation.

4.1.1 Application of dual-frequency correction to wideband VLBI observations

In Section 4.1 we derived the frequency dependence of both the group and phase delays due to propagation of a signal through the ionosphere. In this section we will investigate the methods used to calibrate the ionospheric delay by exploiting its frequency dependence. The algorithm to be used is not as trivial as it first appears. The ionospheric delay at X-band is usually less than 1 nsec. It can be easily calculated that if the ionospheric delay is 1.0 nsec at 8.0 GHz, then at 8.4 GHz the ionospheric delay will be 0.9 nsec. Hence, there is variation of ≈ 0.1 nsec (≈ 3 cm) in the ionospheric delay over the frequency range spanned by the synthesized band. In this section we will investigate a method for compensating for the variation in the ionospheric delay over the synthesized band. We will use the equivalence of maximizing the delay resolution function and least squares to determine the relationships between the estimates of the group delay, phase delay rate and visibility phase, and the integrated electron content and its rate of change. We will then determine an algorithm for calculating the effective frequencies for the group delay, phase delay and visibility phase measurements.

VLBI observations are difference measurements, i.e. the

group delay is the difference in the arrival times of signals from a radio source at two antennas. Consequently, VLBI observations cannot determine the total integrated electron contents along the ray paths from the radio source to each telescope. Only the difference in the integrated electron content can be determined. Hence, the contribution to the estimated cross spectral phases from the propagation of the signal through the ionosphere will be, from Equation (4.1.11),

$$\Delta\phi^i = -\frac{e^2 \Delta N_i}{2cm_e \varepsilon_o \omega}$$

where ΔN_i is the difference in the integrated electron contents along the propagation paths from the radio source to the two sites in the sense $\Delta N_i = (N_i)_2 - (N_i)_1$, where the subscripts 1 and 2 denote the two sites. For the duration of an observation (between 100 sec and 400 sec), we may model ΔN_i by

$$\Delta N_i(t) = \Delta N_i^O + \Delta \dot{N}_i(t-t_o)$$

where ΔN_i^O is the integrated electron content at the epoch, t_o, of the observation, and $\Delta \dot{N}_i$ is the rate of change of the integrated electron content at that epoch.

The effect of the ionospheric delay on the cross spectral phases from an accumulation period (see Section 2.2) as a function of time and frequency will be

$$\Delta\phi^i_{qj} = -K[\Delta N_i^O + \Delta \dot{N}_i(t_q-t_o)]/\omega_j \qquad (4.1.1.1)$$

where q is an index denoting the epoch of the accumulation

period, j is an index denoting frequency, and $K = e^2/(2cm_e\epsilon_o)$ $= 5.2 \times 10^{-6} \ m^2 s^{-1}$.

In Appendix A, we see that the model used to relate the cross spectral phases to the estimates of the group delay, the phase delay rate and the visibility phase is (Equation A.3):

$$\phi_\omega = (\omega - \omega_o)\tau_g + \omega \dot{t}_p (t - t_o) + \phi_t$$

In the discrete form of Equation (4.1.1.1) the above equation becomes

$$\phi_{qj} = (\omega_j - \omega_o)\tau_g + \omega_j \dot{t}_p (t_q - t_o) + \phi_t \qquad (4.1.1.2)$$

We now will investigate the relationship between the estimates of τ_g, t_p and ϕ_t obtained from the model given by Equation (4.1.1.2) and the integrated electron content, ΔN_i^o, and the time derivative of the integrated electron content, $\Delta \dot{N}_i$.

From a single VLBI observation the cross-spectral phases will be estimated at Q epochs and at J frequencies. For a typical observation, Q will of the order of 50 (for a 100 sec observation with 2 sec accumulation periods) and J will be approximately four. For these QJ estimates of the cross spectral phases, Equation (4.1.1.2) may be written as a matrix equation of the form

$$\underline{\phi} = A \ \underline{p} \qquad (4.1.1.3)$$

where $\underline{p}^T = (\phi_t, \ \tau_g, \ \dot{t}_p)$ and the QJ rows of A will be given by $[1, \ \omega_j - \omega_o, \ \omega_j (t_q - t_o)]$, $j=1 \rightarrow J$, $q=1 \rightarrow Q$. The weighted least

squares estimates of $\underline{p}$ given $\underline{\phi}$ will be

$$\hat{\underline{p}} = (A^T V^{-1} A)^{-1} A^T V^{-1} \underline{\phi} \qquad (4.1.1.4)$$

where V is the covariance matrix of $\underline{\phi}$ i.e. $V = \langle \underline{\epsilon}\ \underline{\epsilon}^T \rangle$ where $\underline{\epsilon}$ is the vector of noise contributions to $\underline{\phi}$. (For those readers unfamiliar with weighted least squares, this technique is reviewed in Appendix E.)

We will assume that all of the cross spectral phase estimates have the same variance and that the covariances between the estimates are zero. With these assumptions the covariance matrix cancels in Equation (4.1.1.4) and the equation reduces to

$$\hat{\underline{p}} = (A^T A)^{-1} A^T \underline{\phi} \qquad (4.1.1.5)$$

We may evaluate Equation (4.1.1.5) analytically when the observed phases are given by Equation (4.1.1.1). (The linear properties of the estimator, Equation 4.1.1.5, allow us to study the ionospheric delay independently of the other delays which may be present in the observations.) We will first evaluate $A^T A$ which will is by

$$A^T A = \begin{bmatrix} QJ & Q \sum\limits_{j=1}^{J} (\omega_j - \omega_0) & \sum\limits_{j=1}^{J} \sum\limits_{q=1}^{Q} (t_q - t_0) \omega_j \\[2ex] & Q \sum\limits_{j=1}^{J} (\omega_j - \omega_0)^2 & \sum\limits_{j=1}^{J} \sum\limits_{q=1}^{Q} (t_q - t_0)(\omega_j - \omega_0) \omega_j \\[2ex] & \text{symmetric} & \sum\limits_{j=1}^{J} \sum\limits_{q=1}^{Q} (t_q - t_0)^2 \omega_j^2 \end{bmatrix}$$

We may simplify the above matrix by choosing t_0 such that it is at the center of the observation. The epochs of the

accumulation periods, t_q, are equally spaced and $t_q - t_o$ may be written as $t_q - t_o = (q - q_c)\Delta t_a$, where Δt_a is the duration of an accumulation period (typically 2 sec) and q_c is the index corresponding to t_o; q_c is not necessarily an integer. If the epoch of the observation is taken as the center of the observation duration, q_c will equal $(Q-1)/2$. For this choice of epoch, $A^T A$ reduces to

$$A^T A = \begin{bmatrix} QJ & Q\sum_{j=1}^{J} \Delta\omega_j & 0 \\[2ex] & Q\sum_{j=1}^{J} \Delta\omega_j^2 & 0 \\[2ex] \text{symmetric} & & \sum_{j=1}^{J} \omega_j Q(Q^2-1)/12 \end{bmatrix}$$

where we have substituted $\Delta\omega_j = \omega_j - \omega_o$. This matrix may be easily inverted yielding

$$(A^T A)^{-1} = \begin{bmatrix} Q\sum_{j=1}^{J} \Delta\omega_j^2/\text{DET} & -Q\sum_{j=1}^{J} \Delta\omega_j/\text{DET} & 0 \\[2ex] & QJ/\text{DET} & 0 \\[2ex] \text{symmetric} & & [\sum_{j=1}^{J} \omega_j Q(Q^2-1)/12]^{-1} \end{bmatrix}$$

where $\text{DET} = Q^2 J \sum_{j=1}^{J} \Delta\omega_j^2 - Q^2 (\sum_{j=1}^{J} \Delta\omega_j)^2$. (Note: DET is the determinant of the upper 2X2 portion of $A^T A$, not the determinant of the complete matrix.)

We may now evaluate $(A^T A)^{-1} A^T$. The columns of $(A^T A)^{-1} A^T$

will be given by

$$
[(A^T A)^{-1} A^T]_{col_m} = \begin{bmatrix} Q \sum_{j=1}^{J} \Delta\omega_j^2 / DET - Q \sum_{j=1}^{J} \left(\Delta\omega_j / DET \right) \Delta\omega_k \\[2ex] -Q \sum_{j=1}^{J} \left(\Delta\omega_j / DET \right) + (QJ/DET) \Delta\omega_k \\[2ex] [\sum_{j=1}^{J} \omega_j Q (Q^2 - 1)/12]^{-1} \omega_k (m - q_c) \Delta t_a \end{bmatrix}
$$

$$(4.1.1.6)$$

for the phase from the accumulation period with epoch $(m-q_c) \Delta t_a$ relative to the center of the observation and at frequency ω_k. (m denotes the sequential number of the accumulation period, i.e. m is a specific value of the index q; see Equation (4.1.1.2) and Appendix B for more details.)

We should note that the expression (4.1.1.6) is only dependent on the form of the mathematical model used to express the cross spectral phases as functions of the group delay, phase delay rate and visibility phase. We may now find the relationship between the group delay, phase delay rate and visibility phase estimates for the ionospheric signature in the cross spectral phases given by Equation (4.1.1.1).

The ionospheric delay contributions to the values of the VLBI observables may be obtained by performing the multiplication given in Equation (4.1.1.5) using Equations (4.1.1.1)

and (4.1.1.6). The product is

$$\hat{\phi}_t^i = \frac{-K\Delta N_i^0 \left[\left(\sum_{j=1}^{J}\Delta\omega_j^2\right)\left(\sum_{k=1}^{J}1/\omega_k\right) - \left(\sum_{j=1}^{J}\Delta\omega_j\right)\left(\sum_{k=1}^{J}\Delta\omega_k/\omega_k\right)\right]}{J\sum_{j=1}^{J}\Delta\omega_j^2 - \left(\sum_{j=1}^{J}\Delta\omega_j\right)^2}$$

$$(4.1.1.7.a)$$

$$\hat{\tau}_g^i = \frac{-K\Delta N_i^0 \left[-\left(\sum_{j=1}^{J}\Delta\omega_j\right)\left(\sum_{k=1}^{J}1/\omega_k\right) + J\sum_{k=1}^{J}\Delta\omega_k/\omega_k\right]}{J\sum_{j=1}^{J}\Delta\omega_j^2 - \left(\sum_{j=1}^{J}\Delta\omega_j\right)^2}$$

$$(4.1.1.7.b)$$

$$\hat{\tau}_p^i = \frac{-K\Delta\dot{N}_i J}{\sum_{j=1}^{J}\omega_j^2}$$

$$(4.1.1.7.c)$$

where $\hat{\phi}_t^i$, $\hat{\tau}_g^i$ and $\hat{\tau}_p^i$ are the contributions to the visibility phase, the group delay and the phase delay rate, respectively, from the ionospheric delay and rate.

We should notice that the delay estimates $\hat{\phi}_t^i$ and $\hat{\tau}_g^i$ depend only on the integrated electron content at the epoch of the observation and not on the rate of change of the integrated electron content. Similarly the phase delay rate estimate, $\hat{\tau}_p^i$, depends only on the rate of change of the integrated electron content. This decoupling of the delay and rate estimates follows from the choice of the epoch of the observation to be at the center of the data duration.

We may rewrite Equations (4.1.1.7) as

$$\hat{\phi}_t^i = -K\Delta N_i^0/\omega_{eff}^\phi$$

$$(4.1.1.8.a)$$

$$\hat{\tau}_g^i = +K\Delta N_i^o/(\omega_{eff}^g)^2 \qquad\qquad (4.1.1.8.b)$$

$$\hat{\tau}_p^i = -K\Delta\dot{N}_i/(\omega_{eff}^r)^2 \qquad\qquad (4.1.1.8.c)$$

where ω_{eff}^ϕ, ω_{eff}^g and ω_{eff}^r are, respectivley, the effective
frequencies of the visibility phase, the group delay and the
phase-delay rate estimates for calibrating the ionospheric
delay. By equating Equations (4.1.1.7) and (4.1.1.8) we may
determine analytic expressions for the effective frequencies

$$\omega_{eff}^\phi = \frac{J\sum_{j=1}^{J}\Delta\omega_j^2 - (\sum_{j=1}^{J}\Delta\omega_j)^2}{\sum_{j=1}^{J}\Delta\omega_j^2\sum_{k=1}^{J}(1/\omega_k) - \sum_{j=1}^{J}\Delta\omega_j\sum_{k=1}^{J}\Delta\omega_k/\omega_k} \qquad (4.1.1.9.a)$$

$$\omega_{eff}^g = \left[\frac{J\sum_{j=1}^{J}\Delta\omega_j^2 - (\sum_{j=1}^{J}\Delta\omega_j)^2}{\sum_{j=1}^{J}\Delta\omega_j\sum_{k=1}^{J}(1/\omega_k) - J\sum_{k=1}^{J}\Delta\omega_k/\omega_k}\right]^{1/2} \qquad (4.1.1.9.b)$$

$$\omega_{eff}^r = \left(\sum_{j=1}^{J}\omega_j^2/J\right)^{1/2} \qquad\qquad (4.1.1.9.c)$$

In Section 4.2 when we are developing the phase delay
prediction techniques, we will be using phase delays rather
than the visibility phases. We may easily obtain the effec-
tive frequency of this delay measure from the above
equations. The contribution of the ionospheric delay to the
phase-delay estimate will be

$$\hat{\tau}_p^i = \hat{\phi}_t^i/\omega_o$$

and hence from Equation (4.1.1.7.a) we have that

$$\hat{\tau}_p^i = \frac{-K\Delta N_i^o \left[\sum\limits_{j=1}^{J} \Delta\omega_j^2 \sum\limits_{k=1}^{J} 1/\omega_k - \sum\limits_{j=1}^{J} \Delta\omega_j \sum\limits_{k=1}^{J} \Delta\omega_k/\omega_k \right]}{\omega_o \left[J \sum\limits_{j=1}^{J} \Delta\omega_j^2 - \left(\sum\limits_{j=1}^{J} \Delta\omega_j \right)^2 \right]}$$

In terms of effective frequencies, the equivalent expression to the expression given above is

$$\hat{\tau}_p^i = -K\Delta N_i^o/(\omega_{eff}^p)^2 \tag{4.1.1.10}$$

and therefore by comparing the above two equations we have the effective frequency of the phase delays:

$$\omega_{eff}^p = \left[\frac{\omega_o \left[J \sum\limits_{j=1}^{J} \Delta\omega_j^2 - \left(\sum\limits_{j=1}^{J} \Delta\omega_j \right)^2 \right]}{\sum\limits_{j=1}^{J} \Delta\omega_j^2 \sum\limits_{k=1}^{J} (1/\omega_k) - \sum\limits_{j=1}^{J} \Delta\omega_j \sum\limits_{k=1}^{J} \Delta\omega_k/\omega_k} \right]^{1/2} \tag{4.1.1.11}$$

The expressions for effective frequencies do have a simple interpretation which is shown in Figure 4.1.1.1. The group delay effective frequency is the frequency at which the tangent to the phase versus frequency curve of the dispersive medium is parallel to the straight line determined by the least squares fit to the discretely sampled points on the phase versus frequency curve. (Each of the points shown represents an average of the phase across the 4 MHz recorded bandwidth. It can be easily shown that, for an integrated electron content of 10^{18} electrons/m^2, the curvature of the phase verus frequency curve, at 2.3 GHz, deviates from a straight line by a maximum of 0.016. For our purposes we can treat each channel as a discrete frequency. The slope of the

Figure 4.1.1.1 Graphical representation of the phase and group delay effective frequencies.
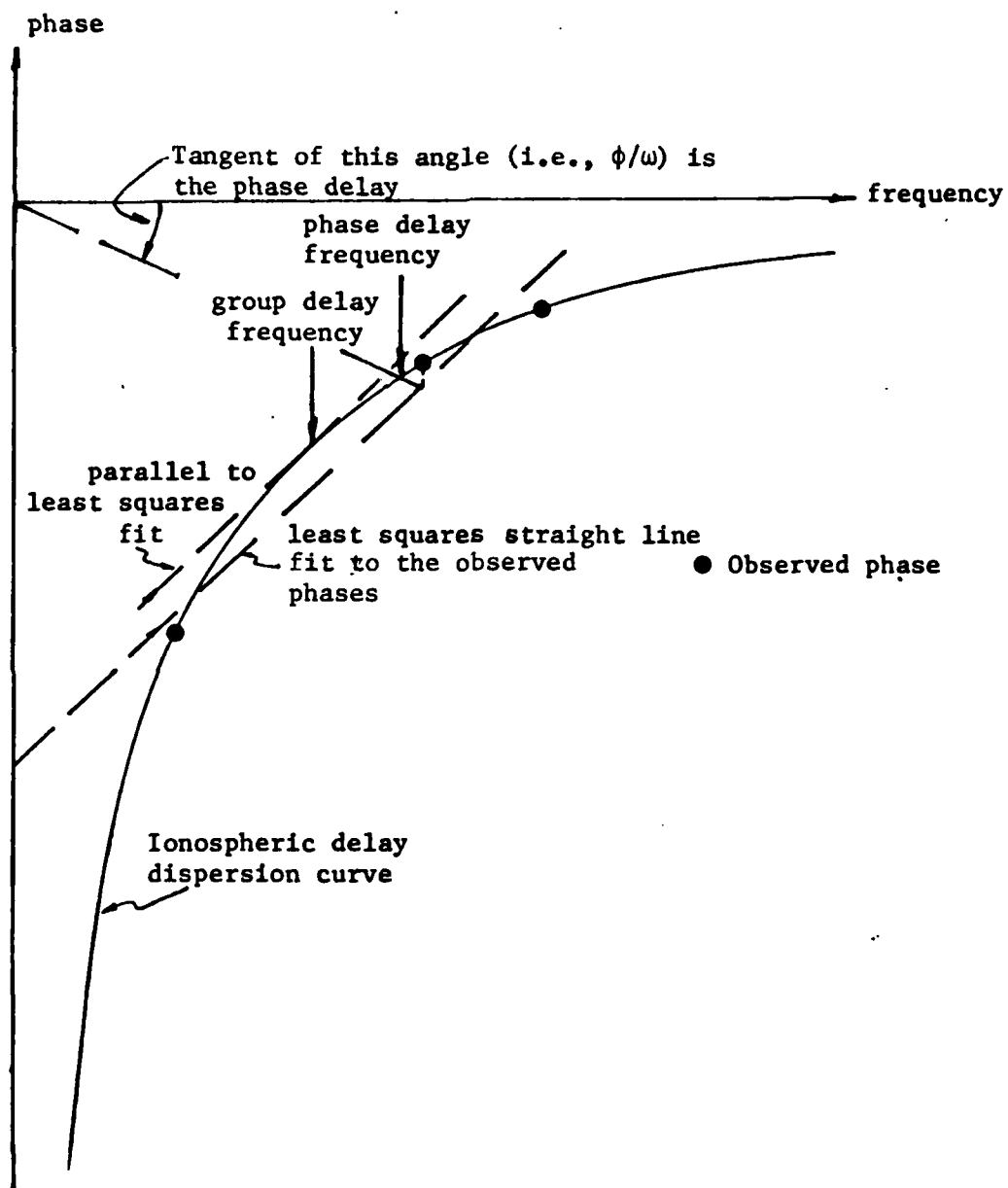
phase across the recorded bandwidth is absorbed into the single band delay estimate; see Appendix B, Equation (B.3.3), for discussion. If we wished to account for the curvature of the phase across the recorded bandwidth, we could extend Equations (4.1.1.2) and (4.1.1.3) to include the singleband delay. At this time, such an extension does not seem warranted.) The phase delay effective frequency is the frequency at which a straight line from the origin to the least squares estimate of the visibility phase intersects the phase versus frequency curve for the ionospheric delay dispersion. (We should note that the curvature of the phase versus frequency curve is **very** exaggerated in Figure 4.1.1.1. The deviation of the curve from a straight line, at 2.3 GHz, over the spanned bandwidth of 75 MHz, is less than 1°, for an integrated electron content of 10^{18} electrons/m^2. We could attempt to estimate the dispersive delay and the non-dispersive delay from a single frequency band, but these estimates would be highly correlated and have very low precision. (For the frequency sequence 2.295 GHz, 2.270 GHz, and 2.220 GHz, and phase measurements of with standard deviations of 1° in each channel, the non-dispersive delay estimate would have a standard deviation of 6.2 nsec. If the dispersive delay is not estimated, the standard deviation of the group delay would be >0.05 nsec. Clearly, we can not separate the dispersive and the non-dispersive delays with observations from only one frequency band.)

We are now able to develop an algorithm which will allow

us to estimate the contribution of the dispersive delay to the group-delay observations by using observations of the delays at two widely separated frequency bands. Equations (4.1.1.8 a-c) and Equation (4.1.1.10) give the contributions to the observed delay and rate values of a plasma with a difference integrated electron density of ΔN_i^0 and rate of change of the integrated electron density of $\Delta \dot{N}_i$. We will now use these relationships to develop an algorithm which will allow us to eliminate the contributions of the plasma delay and rate.

We will assume that observations of the group delay, phase delay and phase delay rate are available at two widely separated frequencies. For the Mark III system these frequencies are X-band (≈ 8.4 GHz) and S-band (≈ 2.3 GHz). We will denote these observations by τ_g^x, τ_p^x and $\dot{\tau}_p^x$ for the X-band group delay, phase delay and phase delay rate, respectively, and τ_g^s, τ_p^s and $\dot{\tau}_p^s$ for the same quantities at S-band. Normally the phase delay observations will be ambiguous and would not be used in the analysis. In Section 4.3 we will discuss the possibilities for the elimination of their ambiguities. For completeness we will develop the algorithms for the elimination of the plasma contributions to these delays at this time.

We now wish to decompose each observed delay and rate into two parts - a contribution from all of the non-dispersive delays (i.e, the geometry, clock epoch offsets and the neutral atmosphere), τ_n, and a contribution from the dispers-

ive plasma delay. For the delay observations, we will also need to include a constant additive offset between each delay measure to account for (the nominally constant) instrumental dispersions which are not calibrated by the phase calibration system. (See Chapter 3 for details.) These offsets allow for the delays at X-band and S-band to be different even in the absence of a dispersive propagation medium. An offset between the group and the phase delays must also be included because of instrumental dispersions (see Section 3.1 which discusses some of the dispersive properties of feed horns and waveguides).

We will now develop the relationships between the delay measurements, the non-dispersive delay and the dispersive delay. Equation (4.1.1.8.b) gives the contribution of the ionospheric propagation delay to the observed group delay values. If we include the non-dispersive delay, the observed group delay value at X-band, τ_g^x, will be

$$\tau_g^x = \tau_n + K\Delta N_i^o/(\omega_g^x)^2 \qquad\qquad (4.1.1.12)$$

where ω_g^x is the effective frequency of the X-band group delays (i.e., $\omega_g^x = \omega_{eff}^g$ computed for the X-band frequencies using Equation (4.1.1.19.b). We will denote $K\Delta N_i^o/(\omega_g^x)^2$ by I_x -- the ionospheric contribution to the X-band group delays. Equation (4.1.1.12) can be written as

$$\tau_g^x = \tau_n + I_x \qquad\qquad (4.1.1.13)$$

The S-band group delay measurements, τ_g^s, will be given by

$$\tau_g^s = \tau_n + (\omega_g^x/\omega_g^s)^2 I_x + O_g^s$$

where ω_g^s is the effective frequency of the S-band group delay observations and O_g^s is a delay offset which accounts for the difference in propagation delay of the X-band and S-band signals through the antenna, feed horns and waveguides up to the injection point of the phase calibration pulses. We will denote $(\omega_g^x/\omega_g^s)^2$ by β which yields

$$\tau_g^s = \tau_n + \beta I_x + O_g^s \qquad (4.1.1.14)$$

The relationship for the X-band phase delays, $\tau_p^{x\prime}$, will be (assuming right circular polarization (RCP))

$$\tau_p^{x\prime} = \tau_n - \alpha I_x + O_p^x + (\psi_1 - \psi_2)/\omega_x$$

where α is $(\omega_g^x/\omega_p^x)^2$, O_p^x is an offset between X-band group delays and the X-band phase delays, ψ_1 and ψ_2 are the position angles of the feed horns at each site of the baseline (see Appendix D), and ω_x is the radio frequency of the visibility phase at X-band. (ω_x is equivalent to ω_0 in previous derivations in this thesis. We change its symbol now to differentiate between X-band and S-band observations.) The analytic expressions for ψ_1 and ψ_2 are derived in Appendix D.

We will redefine the phase delays such that they will be referred to a reference orientation which is fixed with respect to inertial space. This new phase delay, τ_p^x, will be given by

$$\tau_p^x = \tau_p^{x\prime} - (\psi_2 - \psi_1)/\omega_x$$

The expression for these new phase delays will be

$$\tau_p^x = \tau_n - \alpha I_x + O_p^x \qquad (4.1.1.15)$$

The S-band phase delays (again, referred to a constant reference orientation) will be

$$\tau_p^s = \tau_n - \gamma I_x + O_p^s \qquad (4.1.1.16)$$

where $\gamma = (\omega_g^x/\omega_g^s)^2$, ω_p^s is the effective frequency of the S-band phase delays and O_p^s is the offset between the S-band phase delays and the X-band group delays in the absence of propagation through a dispersive medium, aside from the antenna, etc.

If we examine Equations (4.1.1.12) through (4.1.1.15), we see that for each epoch there are four measured quantities (if we include the phase delays), and five unknowns. If only the group delays are considered there will be two measurements and 3 unknowns. With each additional pair of group delays there are two new unknown quantities introduced, τ_n and I_x, for this new epoch. Hence when only the group delay measurements are used, there is no redundancy. If both the group- and phase-delay measurements are used, at each additional epoch, four new measurements are made, but only two new unknowns are introduced. With observations from more than one epoch, redundant information is obtained and we may use this additional information to check the quality of the group-delay measurements.

We will now derive the algorithms which are used to

estimate τ_n and I_x from the various delay measures. If only the group delays are available we have two equations, Equations (4.1.1.12) and (4.1.1.13), which can be used to estimate a linear combination of τ_n, I_x and O_g^S. Examination of these equations yields

$$I_x = (\tau_g^x - \tau_g^S)/(1-\beta) + O_g^S/(1-\beta) \qquad (4.1.1.16)$$

If we assume that O_g^S is constant, but unknown, we see that Equation (4.1.1.16) yields an estimate of the ionospheric delay plus a constant. Without an independent estimate of either I_x or O_g^S we can not uniquely separate these two quantities. This constant additive contribution to the ionospheric delay estimate (if it is indeed constant) is not a severe problem because it will be absorbed into the estimated clock offsets.

An estimate of the non-dispersive delay may be obtained by

$$\tau_n = (\beta\tau_g^x - \tau_g^S)/(\beta-1) - O_g^S/(1-\beta) \qquad (4.1.1.17)$$

Again we see that we cannot separate uniquely the neutral delay from the offset $O_g^S/(1-\beta)$.

When only the group delays are used we have no means of assessing the constancy of $O_g{}^S$ without assuming some properties of either τ_n and I_x.

However, if the phase delays are available we can make some statements concerning the constancy of the offsets given in Equations (4.1.1.13)-(4.1.1.15). Using the same approach which led to Equation (4.1.1.16), we have, from Equations

(4.1.1.14) and (4.1.1.15) that

$$I_x = (\tau_p^X - \tau_p^S)/(\gamma - \alpha) - (0_p^X - 0_p^S)/(\gamma - \alpha) \qquad (4.1.1.18)$$

and

$$\tau_n = (\gamma \tau_p^X - \alpha \tau_p^S)/(\gamma - \alpha) + (\gamma 0_p^X - \alpha 0_p^S)/(\gamma - \alpha) \qquad (4.1.1.19)$$

If we now compare Equations (4.1.1.16) and (4.1.1.18), we see that these equations provide two estimates of the ionospheric delay which should differ by only a constant $[-0_g^S/(1-\beta) - (0_p^X - 0_p^S)/(\gamma - \alpha)]$. In Section 4.3 we will examine the behavior of this difference from several VLBI experiments to ascertain the constancy of these offsets. We will resume this discussion in Section 4.3 and investigate methods which can be used to determine which of the offsets are varying if we do see a non-constant difference between the group-delay and the phase-delay estimates of the ionospheric delay.

The estimates of the non-dispersive delay will be used in the weighted least squares analyses to be discussed in Chapter 5. The weight given to each measurement in these analyses will be inversely proportional to the variance of the measurement (see Appendix E for more details). We may easily calculate the variance of the non-dispersive delay estimates, given the the variances of the group-delay measurements, at X- and S-band, from Equation (4.1.1.7). The variance of τ_n, $(\sigma_n^\tau)^2$, will be (assuming that 0_g^S is constant)

$$(\sigma_n^\tau)^2 = (\beta^2 (\sigma_g^X)^2 + (\sigma_g^S)^2)/(\beta-1)^2 \qquad (4.1.1.20)$$

where $(\sigma_g^X)^2$ and $(\sigma_g^S)^2$ are the variances of the X- and S-band

group delays, and we have assumed that the X- and S-band group delay measurements are uncorrelated (see Section 2.1). Similarly, the variance of the ionospheric-group-delay estimate, $(\sigma_{gg}^{xs})^2$, will be

$$(\sigma_{gg}^{xs})^2 = ((\sigma_g^x)^2 + (\sigma_g^s)^2)/(\beta-1)^2 . \qquad (4.1.1.21)$$

We will investigate these variances in more detail in Section 4.2, when we will need to know the precision of the phase-delay predictions.

We have now derived all of the theoretical relationships which are necessary for us to be able to predict the phase delays from the group delays. While deriving these formulas we seemed to make assumptions in every step of the derivations. Therefore before proceeding any further we should evaluate the possible magnitudes of the neglected terms.

4.1.2 Accuracy of the ionospheric correction

In the preceding sections of this chapter we have developed a model for the frequency dependence of the ionospheric delay for both the group delays and the phase delays. While deriving these results (Equations (4.1.8) and (4.1.11)), we made a number of approximations and neglected several terms. In this section we will investigate the magnitude of the effects of these approximations and neglected terms.

In Table 4.1.2.1 we summarize the terms which were

Table 4.1.2.1 Terms neglected in the derivation of the dual frequency ionospheric delay correction

Neglected term	Symbol used in text[*]	
A. Magnetic field term	ΔI_m^p	ΔI_m^g
B. Damping term (due to collisions)	ΔI_c^p	ΔI_c^g
C. Higher order terms in the the refractive index expansion	ΔI_e^p	ΔI_e^g
D. Ray integral through the ionosphere - bending	ΔI_b^p	ΔI_b^g
E. Ray integral through the ionosphere - scintillation	σ_s^p	σ_s^g

---

[*] These symbols are used to denote the contribution of each of these terms to the observed delays. The superscripts p and g denote contributions to the phase delay and group delay, respectively.

neglected in the derivations in Section 4.1. We will now analyze each of these terms and try to place upper bounds on their contributions to the ionospheric delay corrections. Each term has been given a letter code which will be used when the magnitude of each term's contribution is listed in a summary.

We will study each of the neglected terms in the order in which they are listed in Table 4.1.2.1. At the end of this section we will tabulate the effects of each of the neglected terms for ease of analysis.

The first approximations we made were in Equation (4.1.2) when we neglected the effects of the magnetic and damping terms. We will firstly consider the magnetic term. When an external magnetic field is applied to a plasma, the plasma becomes birefringent, i.e. the phase velocity of an electromagnetic wave propagating through the plasma depends on the wave's polarization. For circularly polarized radiation the dielectric constant of the plasma can be written as (Jackson 1975, p. 293)

$$\epsilon_{\pm} \simeq \epsilon_0 \left(1 - \omega_p^2 / (\omega^2 \pm \omega\omega_b)\right) \tag{4.1.2.1}$$

where the upper sign refer to right circular polarization (IEEE definition) and the lower sign to left circular polarization and ω_b, the precession frequency, is given by (MKS units)

$$\omega_b = |e| \, \bar{B}_0 \cdot \hat{k} / m_e \simeq 8.8 \times 10^6 \, s^{-1} \text{ for } |\bar{B}_0 \cdot \hat{k}| = 0.5 \times 10^{-4} \text{Wbm}^{-2}$$

where $\hat{k}$ is a unit vector in the direction of propagation.

Equation (4.1.2.1) is an aproximation because terms of the order of ω_b^2 have been neglected during its derivation. This approximation is adequate for our purposes because the ω_b^2 terms are small compared to the $\omega\omega_b$ terms which are small compared to the primary term.

We will restrict the following discussion to right circular polarization. Equation (4.1.2.1) may be expanded in Taylor series, which yields,

$$\varepsilon_+ \simeq \varepsilon_0 (1 - (\omega_p^2/\omega^2) + (\omega_p^2\omega_b/\omega^3))$$

The two leading terms in the expression for the dielectric constant are the terms which we have already considered (Equation (4.1.5)). The contribution to the dielectric constant from the magnetic field, $\Delta\varepsilon_+^m$, will be

$$\Delta\varepsilon_+^m = \varepsilon_0\omega_p^2\omega_b/\omega^3$$

and the resultant error in the phase refractive index, Δn_+^m, will be

$$\Delta n_+^m = \omega_p^2\omega_b/(2\omega^3) \qquad (4.1.2.2)$$

To determine the additional phase delay, ΔI_m^p, due to the magnetic field we integrate $\Delta n_+^m/c$ through the ionosphere. Hence

$$\Delta I_m^p = \int_0^{s_o} \omega_p^2\omega_b/(2\omega^3) \ ds$$

If we now substitute for ω_p we obtain

$$\Delta I_m^p = \frac{e^2}{2cm_e\varepsilon_0\omega^3} \int_0^{s_o} \omega_b N_v \ ds \qquad (4.1.2.3)$$

The precession frequency, ω_b, will in general be a function of s because of the variation of $\bar{B}_o$ with altitude and location. However, we are interested in only bounding the magnitude of this term and we will assume that $\bar{B}_o \cdot \hat{R} = |\bar{B}_o|$ at the Earth's surface. If we assume ω_b is constant at its maximum value, ω_b^m, the contribution of the magnetic field term to the phase delay will be (from Equation (4.1.2.3))

$$|\Delta I_m^p| < \frac{e^2 \omega_b^m N_i}{2 c m_e \epsilon_o \omega^3} \qquad (4.1.2.4)$$

We may now determine the corresponding bound on the group delay error, ΔI_m^g, by multiplying Equation (4.1.2.4) by ω to find the phase error and then by differentiating the phase error with respect to angular frequency,

$$|\Delta I_m^g| < \frac{e^2 \omega_b^m N_i}{c m_e \epsilon_o \omega^3} \qquad (4.1.2.4)$$

where the ΔI_m^p and ΔI_m^g will have opposite signs.

To evaluate ω_b^m, we will use the dipole approximation of the Earth's magnetic field which accounts for 95% of the total field strength. The dipole moment of the Earth's magnetic field is 7.94×10^{22} Am^2 which yields a magnetic induction at the antenns site of (Stacy 1977, pp. 211-212)

$$|\bar{B}_o| = 3.07 \times 10^{-5} (1 + 3\cos^2\theta)^{1/2} \quad \text{Wb m}^{-2}$$

where θ is the angle between the dipole axis and the radius vector to the site.

At mid-latitude sites $|\bar{B}_o|$ will be approximately

0.5×10^{-4} Wb m^{-2} which yields a value of 8.8×10^6 s^{-1} for ω_b^m. If we take an extreme value for the integrated electron content of 10^{18} electrons/m^2, the magnitude of the contribution of of magnetic field term would be

$$|\Delta I_m^p| = 0.3 \text{ psec}, \qquad |\Delta I_m^g| = 0.6 \text{ psec}; \qquad \text{at } 8.4 \text{ GHz}$$

$$|\Delta I_m^p| = 15.4 \text{ psec}, \qquad |\Delta I_m^g| = -30.9 \text{ psec}; \qquad \text{at } 2.3 \text{ GHz}$$

These errors will lead to errors in ionospheric delay-corrected group and phase delays of magnitude 1.9 psec and 0.9 psec, respectively. These errors are small compared to other inadequacies in the geodetic models for the group delays.

The second term which was neglected in Equation (4.1.2) was the damping term. When the damping term is considered the dielectric constant becomes complex (in the sense of real and imaginary components) and is given by (Jackson 1975, p. 287)

$$\varepsilon = \varepsilon_0 [1 + i\omega_p^2/\omega(\gamma - i\omega)]$$

which when separated into real and imaginary parts becomes

$$\varepsilon = \varepsilon_0 \left(1 - \frac{\omega_p^2}{(\omega^2 + \gamma^2)} + \frac{i\gamma\omega_p^2}{\omega(\gamma^2 + \omega^2)}\right) \qquad (4.1.2.6)$$

The imaginary part of Equation (4.1.2.6) will cause a loss in the energy of the wave as the wave propagates through the plasma. The real part will affect the propagation velocity of the wave. The damping constant, γ, is a phenomenological model of damping, i.e. a parameterization which seems to model damping in materials. The damping appears to be due to collisions between electrons and molecules and γ is inter-

preted as the collision frequency between electrons and molecules (see Jackson 1975, p. 287-288).

In order to bound the magnitude of the effect of the damping term we will use a value of the collision frequency in the lower part of the ionosphere (≈85 km altitude). The collision frequency of electrons and neutral atoms has been calculated by Nicolet (1959). At 85 km altitude Nicolet gives a value of approximately 10^6 sec^{-1}. The theoretical values of Nicolet for different altitudes are higher (by up to a factor of two), than the measured values given by Kane (1959). (Kane used the relative absorption of two polarizations of 7.75 MHz continuous wave (CW) signals transmitted from a rocket ascending between 60 and 85 km.) More recent measurement of the collision frequency between ions and neutral atoms based on incoherent radar scatter observations (Schlegel et al., 1980; Murdin 1981) can also be used to deduce electron-neutral atom collision frequncies. At 85 km altitude, the maximum measured value of ion-neutral atom collision frequency was 6×10^4 sec^{-1}. This value may be converted to the electron-neutral collision frequency by the relationship (Schlegel et al., 1980)

$$\gamma = \gamma_{in} \, 0.357\sqrt{(m_i/m_e)(T_e/T_i)}$$

where γ and γ_{in} are the electron-neutral and ion-neutral collision frequencies, respectively, m_e and m_i are the electron and ion masses, T_e and T_i are the electron and ion temperatures and the factor 0.357 accounts for the difference

in effective collisional cross section of the electron-neutral and ion-neutral collisions. The mass and temperature ratios account for the difference in velocity of the electrons and ions. If we take the ratio of the mass of the ions (mainly NO^+ and O_2^+) to electrons to be $\approx 6 \times 10^4$ and T_e/T_i to be unity, the electron-neutral collision frequency corresponding to $\gamma_{in} = 6 \times 10^4$ sec^{-1} would be $\approx 5 \times 10^6$ sec^{-1}.

The collisional frequency is highly variable and the difference by a factor of five between this value and Nicolet's value is not unexpected. Since we are only interested in bounding the magnitude of the effects of collisions we will use the higher value of $\gamma = 5 \times 10^6$ sec^{-1}.

Now that we have some knowledge of the magnitude of γ we may return to the analysis of Equation (4.1.2.6). Since γ is small compared to ω we will again expand the real part of dielectric constant in series which yields

$$\epsilon \approx \epsilon_o \left(1 - (\omega_p^2/\omega^2) + (\omega_p^2 \gamma^2/\omega^4) \right)$$

The contribution of the collision frequency will lead to an error in the phase refractive index, Δn^c, of

$$\Delta n^c = \omega_p^2 \gamma^2/(2\omega^4) \qquad\qquad (4.1.2.7)$$

We will now assume that the collision frequency remains constant throughout the ionosphere (this assumption will lead to a large over-estimation of its effect). Integration of $\Delta n^c/c$ through the ionosphere leads to a phase delay error of

$$\Delta I_c^p = \frac{e^2}{2cm_e \epsilon_o} N_i \frac{\gamma^2}{\omega^4} \qquad\qquad (4.1.2.8)$$

and a group delay error of

$$\Delta I_c^g = - \frac{3e^2}{2cm_e\epsilon_o} N_i \frac{\gamma^2}{\omega^4} \tag{4.1.2.9}$$

Substitution into Equations (4.1.2.8) and (4.1.2.9) leads to (assuming $N_i = 10^{18}$ electrons/m^2)

$$\Delta I_c^p = 1.7 \times 10^{-5} \text{ psec}, \quad \Delta I_c^g = -5.1 \times 10^{-5} \text{ psec}; \quad \text{at 8.4 GHz}$$

$$\Delta I_c^p = 0.003 \text{ psec}, \quad \Delta I_c^g = -0.009 \text{ psec}; \quad \text{at 2.3 GHz}$$

These errors are negligible.

The next term we neglected was the second order terms in the expansion of the square root of $(1-\omega_p^2/\omega^2)$ in Equation (4.1.7). When the second order terms are included the phase refractive index becomes

$$n_p = 1 - (\omega_p^2/2\omega^2) + (\omega_p^2/\omega^2)^2/8 \tag{4.1.2.10}$$

The error in the phase delay due to the neglect of the second order term will be

$$\Delta I_e^p = \left| \frac{q^2}{m_e\epsilon_o} \right|^2 \frac{1}{8c\omega^4} \int_0^{s_o} N_v^2 \, ds \tag{4.1.2.11}$$

To compute an approximate value of ΔI_e^p, we will adopt the simplified profile of the electron density content shown in Figure 4.1.2.1 . We assume that the electron density increases linearly with decreasing altitude from h_m to h_1 (or equivalently with increasing distance along the ray path from 0 to s_1) until a maximum volumetric density of N_v^{max} is reached. The electron density then linearly decreases until altitude h_o is reached (equivalently, distance s_o). The

Figure 4.1.2.1 Electron density profile used to evaluate the effects
of neglected terms
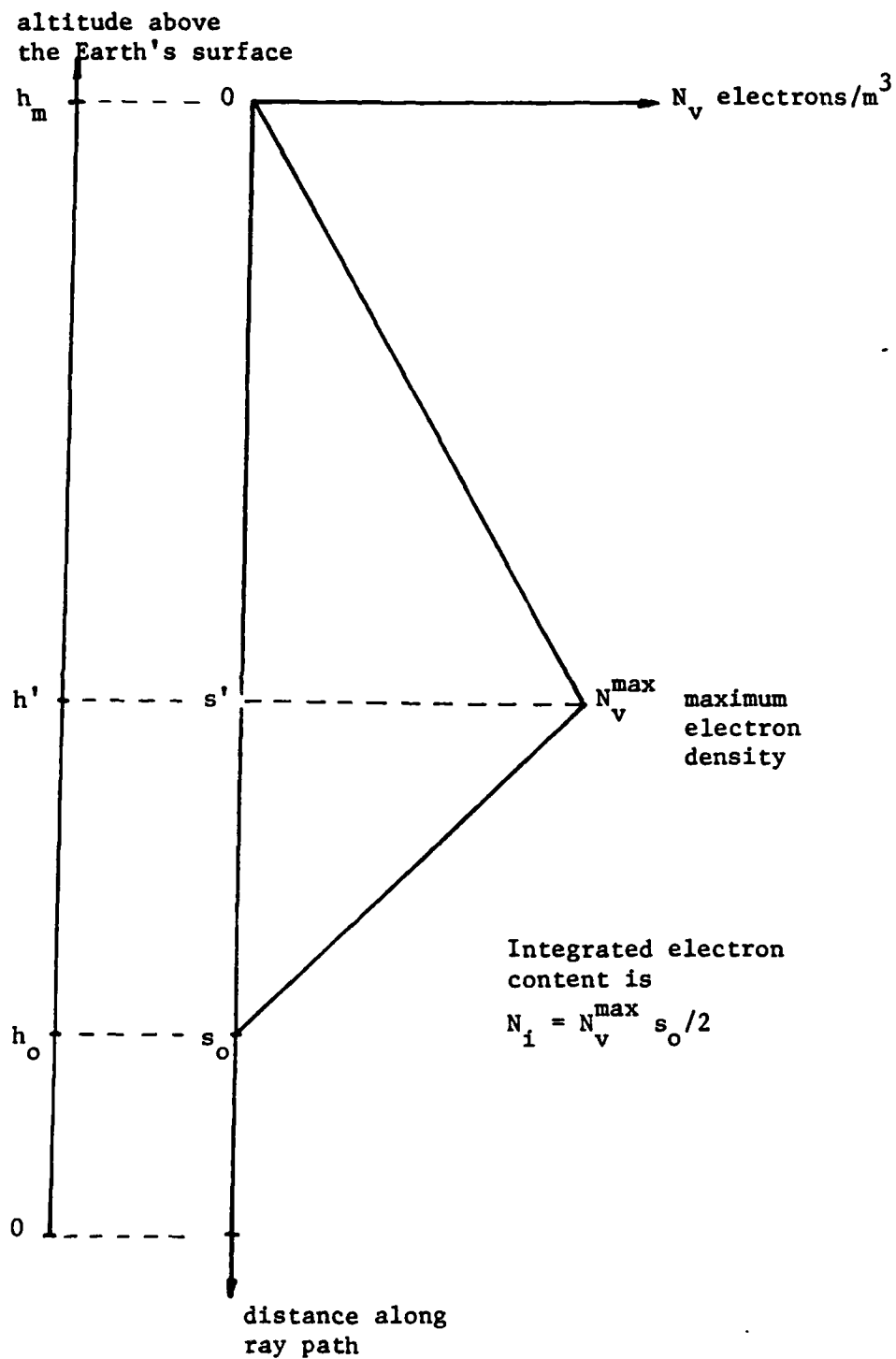
total integrated electron content for this model will be

$$N_i = N_v^{max} s_o / 2 \ .$$

In terms of the total integrated electron content, we find

$$\int_0^{s_o} N_v^2 \ ds = 4N_i^2 / 3 s_o$$

Substitution of this result into Equation (4.1.2.11) yields for the phase delay error

$$\Delta I_e^p = \frac{q^4 N_i^2}{6 m_e^2 \epsilon_o^2 c s_o \omega^4} \qquad (4.1.2.12)$$

The group delay error will be

$$\Delta I_e^g = - \frac{q^4 N_i^2}{2 m_e^2 \epsilon_o^2 c s_o \omega^4} \qquad (4.1.2.13)$$

If we take $N_i = 10^{18}$ electrons/m^2 and the thickness of the ionosphere (s_o) to be 800 km, the contribution of the second order expansion term to the X- and S-band phase and group delays will be

$$\Delta I_e^p = 9. \times 10^{-4} \text{ psec}, \quad \Delta I_e^g = -3. \times 10^{-3} \text{ psec}; \quad \text{at 8.4 GHz}$$

$$\Delta I_e^p = 0.2 \quad \text{psec}, \quad \Delta I_e^g = -0.5 \quad \text{psec}; \quad \text{at 2.3 GHz}$$

These errors are insignificant. Even if the thickness of the ionosphere were reduced by a factor of 10 (to 80 km) the contribution of this neglected term would still be small.

The next inadequacy in our model for the ionosphere delay was the ray integrals given in Equation (4.1.9). There are several inadequacies in this equation which we will

discuss. These inadequacies are 1) the line integrals are along the ray paths through the ionosphere. Since the ionosphere will refract the radiation, the paths of integrations will be different for the two frequencies used. Hence for the two frequencies, the integrated electron contents could be different; 2) the concept of a ray-path integration will be inadequate if there are inhomogeneities in the ionosphere on scales smaller than the Fresnel zone diameter of radiation from the height of the ionosphere. We will now discuss these errors in the order listed.

The difference in the integration paths can be treated within the framework which we have already developed. We will assume that the ionosphere is a plane stratified medium (this assumption should be adequate since we are studying an error which should be small; if the error is large then we should re-evaluate the plane stratified assumption). Figure 4.1.2.2 shows our idealized model of the layered ionosphere. For a plane stratified medium the angle between ray direction before entering the medium and the ray direction in a layer with refractive index $n_p(s)$ will be given by Snell's law

$$\sin \theta_i / \sin \theta_r(s) = n_p(s)/n_o \qquad (4.1.2.14)$$

where n_o is the refractive index in the region outside the medium (for our purposes we can take n_o to be unity). As we have seen already $n_p(s)$ deviates from n_o by only a small amount, $(\approx \omega_p^2/2\omega^2)$, and hence we will expand Snell's law in series. We will define $\Delta\theta(s)$ to be $\theta_r(s)-\theta_i$. $\Delta\theta(s)$ will be

Figure 4.1.2.2   Definitions of the refraction angles for a plane stratified
ionosphere
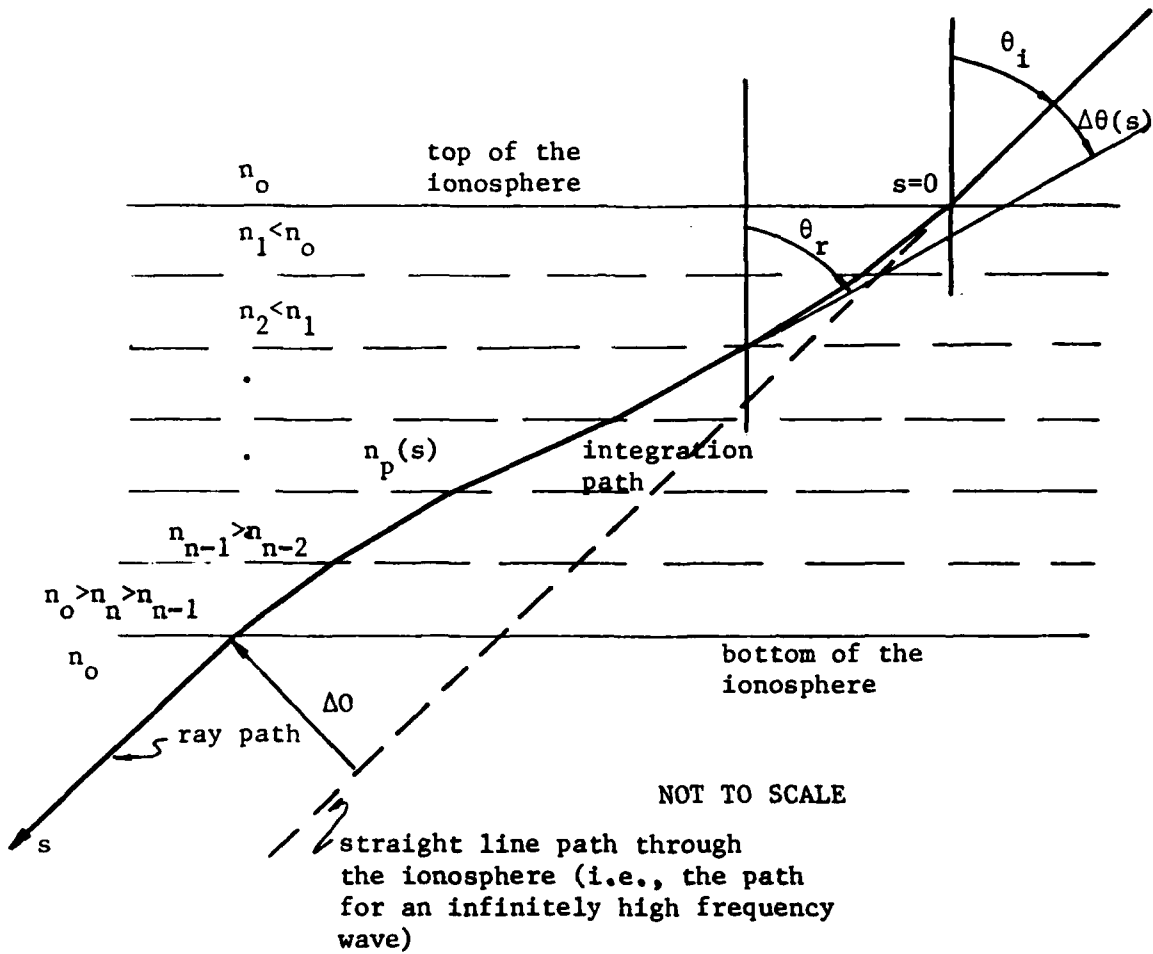
approximately

$$\Delta\theta(s) \simeq -\tan\theta_i \ \Delta n_p(s) \qquad (4.1.2.15)$$

where $\Delta n_p(s) = n_p(s) - n_o$. Equation (4.1.2.15) allows us to deduce both the deviation of the path through the ionosphere from a straight line and the additional path travelled by the radiation because of the ray curvature. The offset, ΔO, between the actual path travelled by the radiation and a straight line path through the ionosphere, will be

$$\Delta O \simeq \int_0^{s_o} \Delta\theta \ ds \simeq -\tan\theta_i \int_0^{s_o} \Delta n_p \ ds \qquad (4.1.2.16)$$
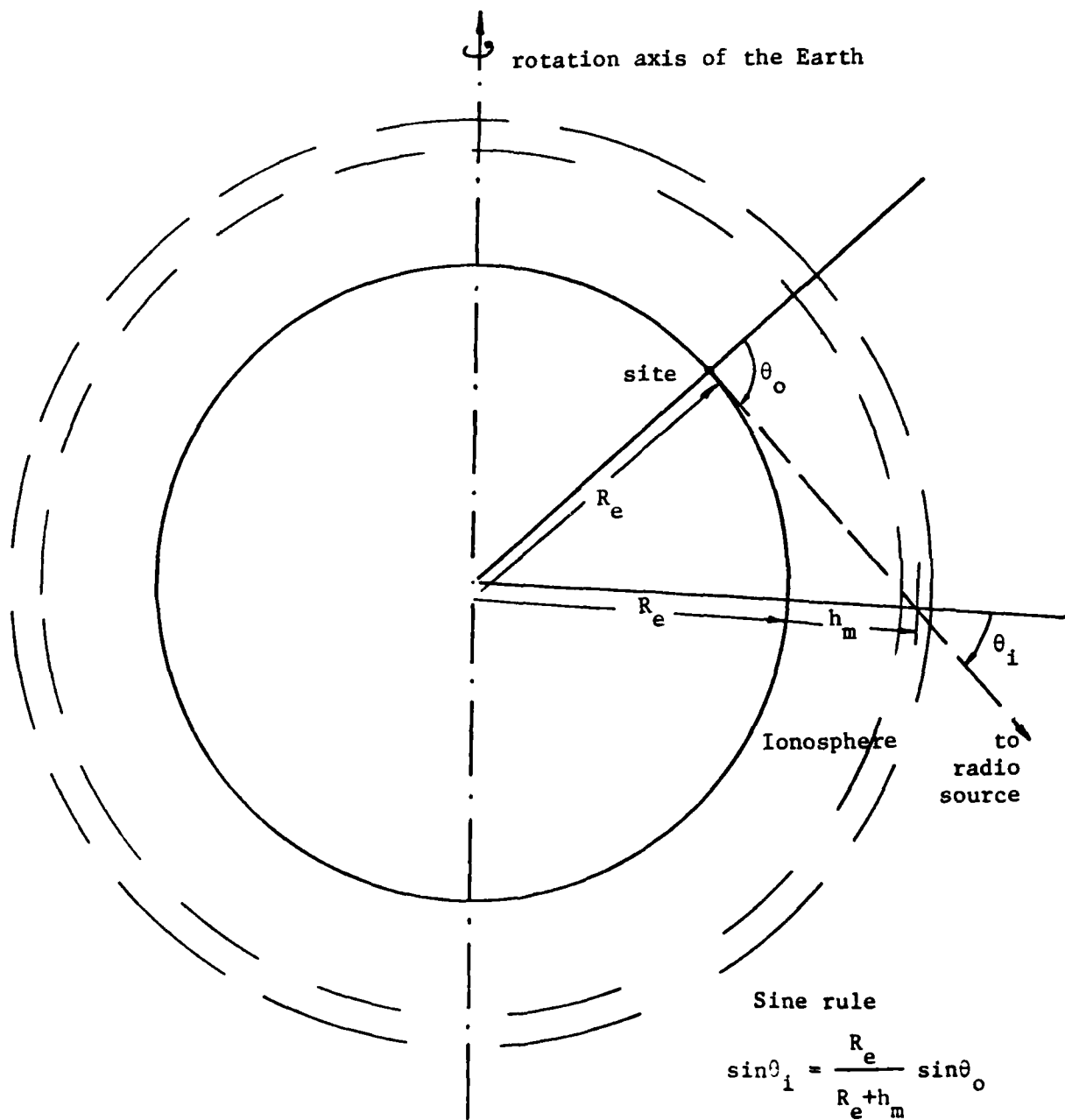
The integral in Equation (4.1.2.16) has already been evaluated. This integral is simply the ionospheric phase delay multiplied by the speed of light. For an integrated electron content of 10^{18} electrons/m^2, the deviations of the S-band signal from a straight line would be $\Delta O \simeq 7.5 \tan\theta_i$ m.

This offset would appear to grow very large for low elevation observations (i.e., $\theta_i \simeq 90^\circ$). This behavior is, however, an artifact of the plane approximation which was used to derive Equation (4.1.2.16). When we consider a curved Earth (Figure 4.1.2.3) an observation at 0° elevation angle at the surface of the earth (neglecting tropospheric refraction) would impinge on a spherical layer h_m above the Earth's surface at a zenith angle of

$$\theta_i^o = \sin^{-1}\left(R_e/(R_e+h_m)\right)$$

where R_e is the radius of the Earth. If we take h_m to be 300 kms, θ_i^o would be 73 deg. The maximum deviation of an S-band path from a straight line would then be $\simeq 25$ m in this case.

Figure 4.1.2.3 Geometry of impinging angle for a spherical ionospheric
layer

rotation axis of the Earth

site

θ_o

R_e

R_e

h_m

θ_i

Ionosphere

to
radio
source

Sine rule

$$\sin\theta_i = \frac{R_e}{R_e+h_m} \sin\theta_o$$

If large variations in the electron content occur on scales of 25 meters, then the integrated electron content would be different for the two frequencies. However if this variation were to occur, then the ray approximations we have been using thus far would no longer be adequate. We will continue this discussion when we discuss the adequacy of the ray trace method.

A second consequence of the path of the radiation being deflected as it propagates through the ionosphere is that the path length is increased. The contribution to the phase delay from the additional path length due to ray bending will be approximately

$$\Delta I_b^p = (1/c) \int_0^{s_o} \Delta\theta^2/2 \ ds$$

which, by substituting Equation (4.1.2.14) into the above equation, becomes

$$\Delta I_b^p = \left((\tan^2\theta_i)/2c\right) \int_0^{s_o} \Delta n_p^2 ds = \left((\tan^2\theta_i)/8c\right) \int_0^{s_o} \left(\omega_p^2/\omega^2\right)^2 ds$$

If we refer back to Equations (4.1.2.10) and (4.1.2.11) we see that

$$\Delta I_b^p = \tan^2\theta_i \ \Delta I_e^p \qquad\qquad (4.1.2.17)$$

where ΔI_e^p is the phase delay error due to the second order terms in the expansion of n_p (see Equation (4.1.2.12)). The group-delay contribution of the bending term will be

$$\Delta I_b^g = \tan^2\theta_i \ |\Delta I_e^g| \qquad\qquad (4.1.2.18)$$

For an observation at 0° elevation on the surface of the

earth and for an integrated electron content of 10^{18} electron/m^2, the contribution of the additional path length will be

$$\Delta I_b^p = 0.003 \quad \text{psec}, \qquad \Delta I_b^g = 0.009 \text{ psec}; \qquad \text{at 8.4 GHz}$$

$$\Delta I_b^p = 1.7 \quad \text{psec}, \qquad \Delta I_b^g = 5.0 \quad \text{psec}; \qquad \text{at 2.3 GHz}$$

Again these contributions are small.

We will now study the second aspect of the ray path integration -- scintillation. We calculated in an earlier part of this section that the ray path of the S-band signals could deviate from a straight line by up to 25 m. If large inhomogeneities in the plasma density are present on scales of 25 m, then the use of a ray to represent the propagation of radio wave is invalid. We will now investigate the effects of small scale ionospheric inhomogeneities on the propagation of radio waves.

There are two methods we could use to study the effects of scintillation on VLBI observations. One is to use observations of scintillation on satellite signals and the other is a theoretical approach. Because we are interested in bounding the effects of scintillation we will adopt the former aproach. (The reader is referred to the works of Booker, 1975; Costa and Kelly, 1976; Rino, 1979a; Rino, 1979b; and Bogusch et al., 1981, for discussions of the theoretical methods of studying scintillation. These works are primarily concerned with studying the ionosphere using scintillation observations which is the inverse of our aim in

this study.) The most extensive study of scintillation of satellite radio transmissions has been made by Fremouw et al. (1978). This work studied the phase and amplitude fluctuations of 10 mutually coherent continuous wave (CW) signals transmitted from Defense Nuclear Agency (DNA) satellite P76 -5. The ten CW signals were distributed in frequency between VHF and S-band. These experiments confirm that the frequency dependence of the root-mean-square (RMS) phase fluctuations obey the inverse frequency law which is predicted from the theoretical models (see references above).

Figure 4.1.2.4 reproduces the phase RMS versus frequency plots from Fremouw et al. (1978) for Poker Flat (geomagnetic dip latitude $65°.4$ N) and Ancon (geomanetic dip latitude $0°.4$ N). We believe that these are the only results from the DNA Wideband experiment which have been analyzed. The Poker Flat results are most applicable to the mid-latitude sites which have been used in the VLBI experiments we will be studying. The results bound the scintillation effects at 25 psec for the S-band phase delays (corresponding to $20°$ phase RMS error), and 2 psec for the X-band phase delays (corresponding to $5°$ RMS phase error).

Observations of group delay scintillations are non-existent although Bogusch et al. (1981) address this problem from a theoretical aspect. We can place upper and lower bounds on the scintillation errors based on the phase-delay results. The lower bound for the group-delay scintillation will be the same as the phase delay scintillation, assuming

Figure 4.1.2.4   Phase scintillation as a function of frequency
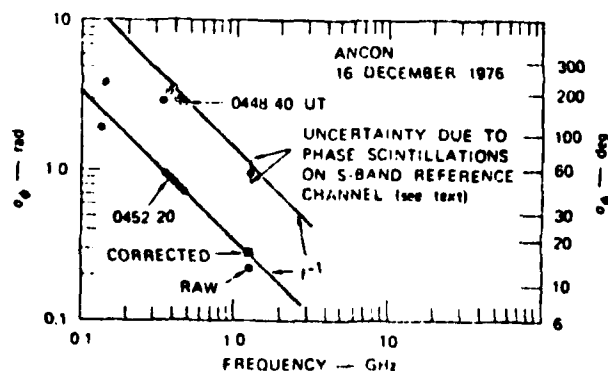(reproduced from Fremouw et al., 1978)
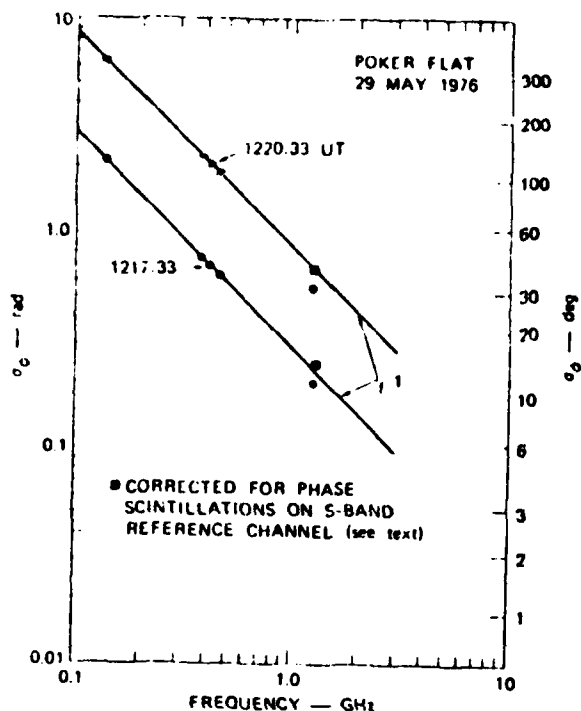
Poker Flat, Alaska

latitude 65° 7' N

longitude 147° 29' W

geomagnetic dip
latitude   65°.4 N

Ancon, Peru

latitude 11° 46' S

longitude 77° 9' W

geomagnetic dip
latitude   0°.4 N

that the phase scintillations are perfectly correlated across
the bandwidth used to measure the group delay. The upper
bound can be calculated assuming the phase errors due to
scintillation are uncorrelated across the observing band-
width. At S-band, the frequency range used to estimate the
group delay is $\approx$75 MHz and 20° phase errors in three channels
(at 2.295 GHz, 2.270 GHz and 2.220 GHz) spanning this band-
width would lead to group-delay errors of $\approx$1 nsec. At
X-band, the spanned bandwidth is $\approx$300 MHz and 5° RMS phase
errors in 4 channels (at 8.34 GHz, 8.49 GHz, 8.54 GHz, and
8.24 GHz) would lead to group-delay errors of $\approx$0.06 nsec.
These upper bounds are most likely far too large.

We can obtain much better estimates of both group-delay
and phase-delay scintillation effects from VLBI data.
Scintillations are due to small scale irregularites in iono-
spheric plasma density. The Fresnel-zone diameter at a
height of 350 kms, for a radio source an infinite distance
from the observer and an observing frequency of 2.3 GHz, is
$\approx$430 m. (The Fresnel-zone diameter at a distance d_1 from the
observer with the source at infinity is $2\sqrt{\lambda d_1}$ where λ is the
wavelength of the observed radiation.) Most of the scintil-
lation will be caused by ionospheric density fluctuations
which have spatial wavelengths of less than the Fresnel-zone
diameter. By looking at VLBI measurements of ionospheric
delay on a short baseline we should be able to observe
directly the scintillation, or at least place bounds on its
magnitude. In Figures 4.1.2.5a and b, we show the iono-

Figure 4.1.2.5.a Ionospheric delay correction at X-band computed from
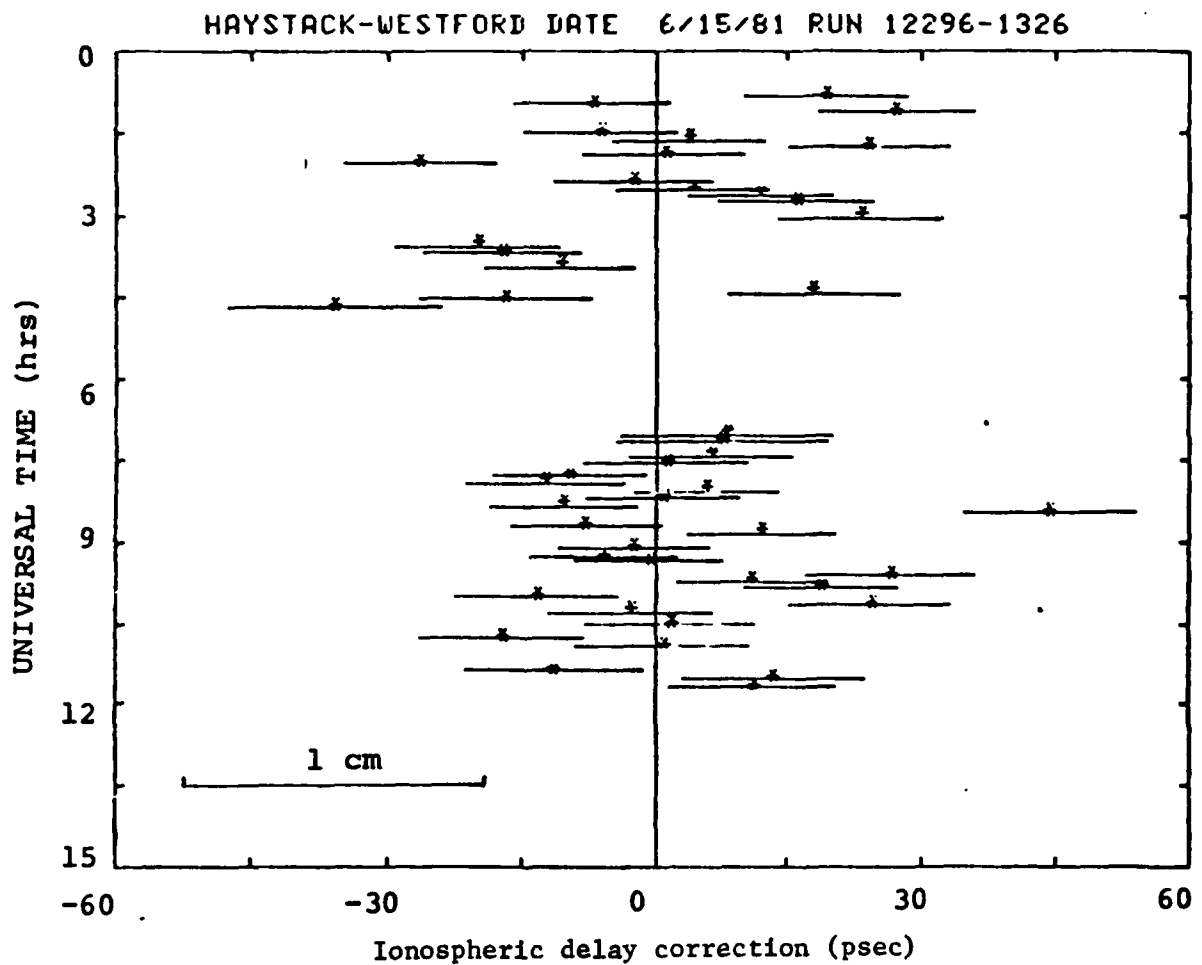X- and S-band group delays.

Figure 4.1.2.5.b Ionospheric delay correction at X-band computed from
X-band and S-band phase delays.

HAYSTACK-WESTFORD DATE 6/15/81 RUN 12296-1316

UNIVERSAL TIME (hrs)

1 cm

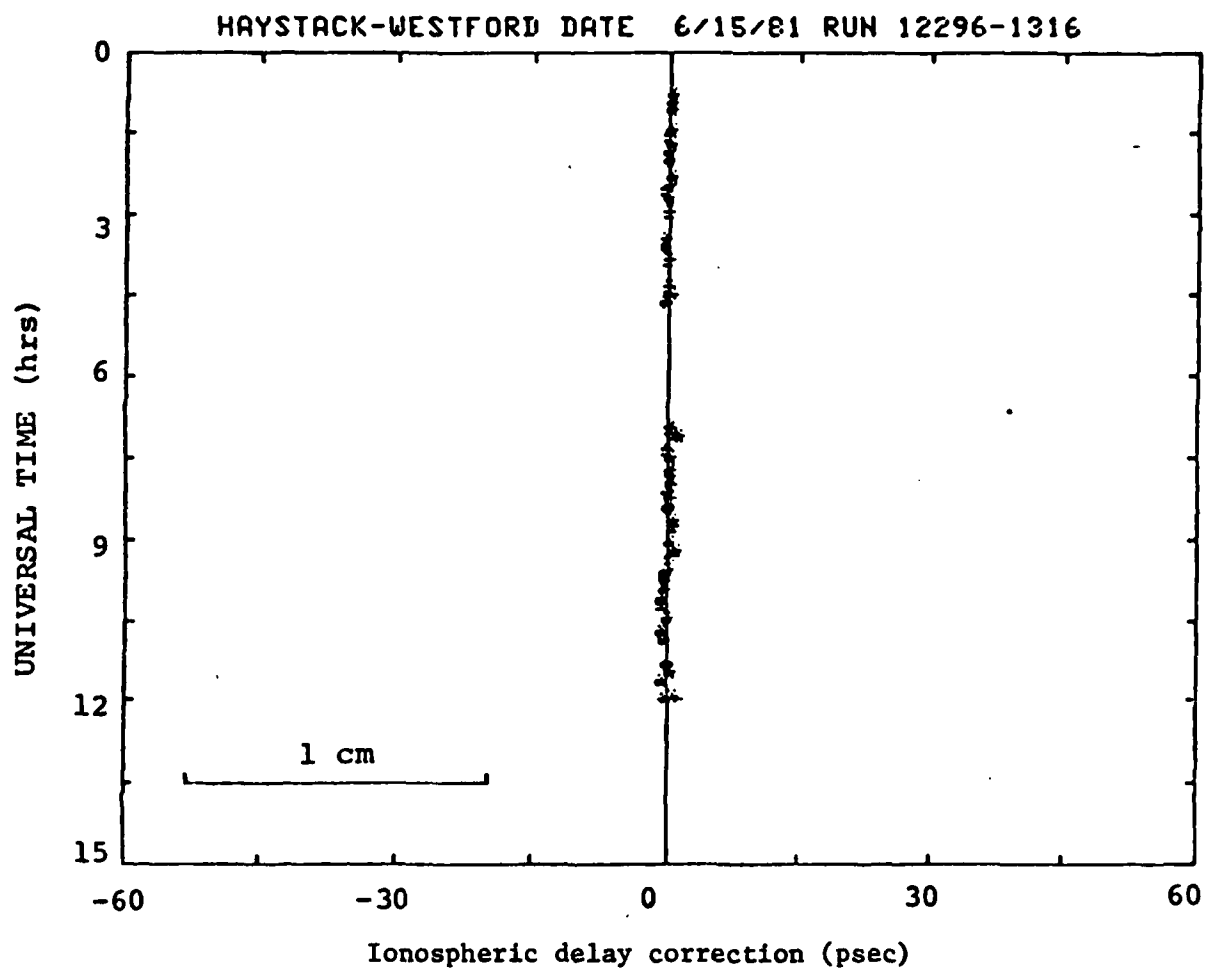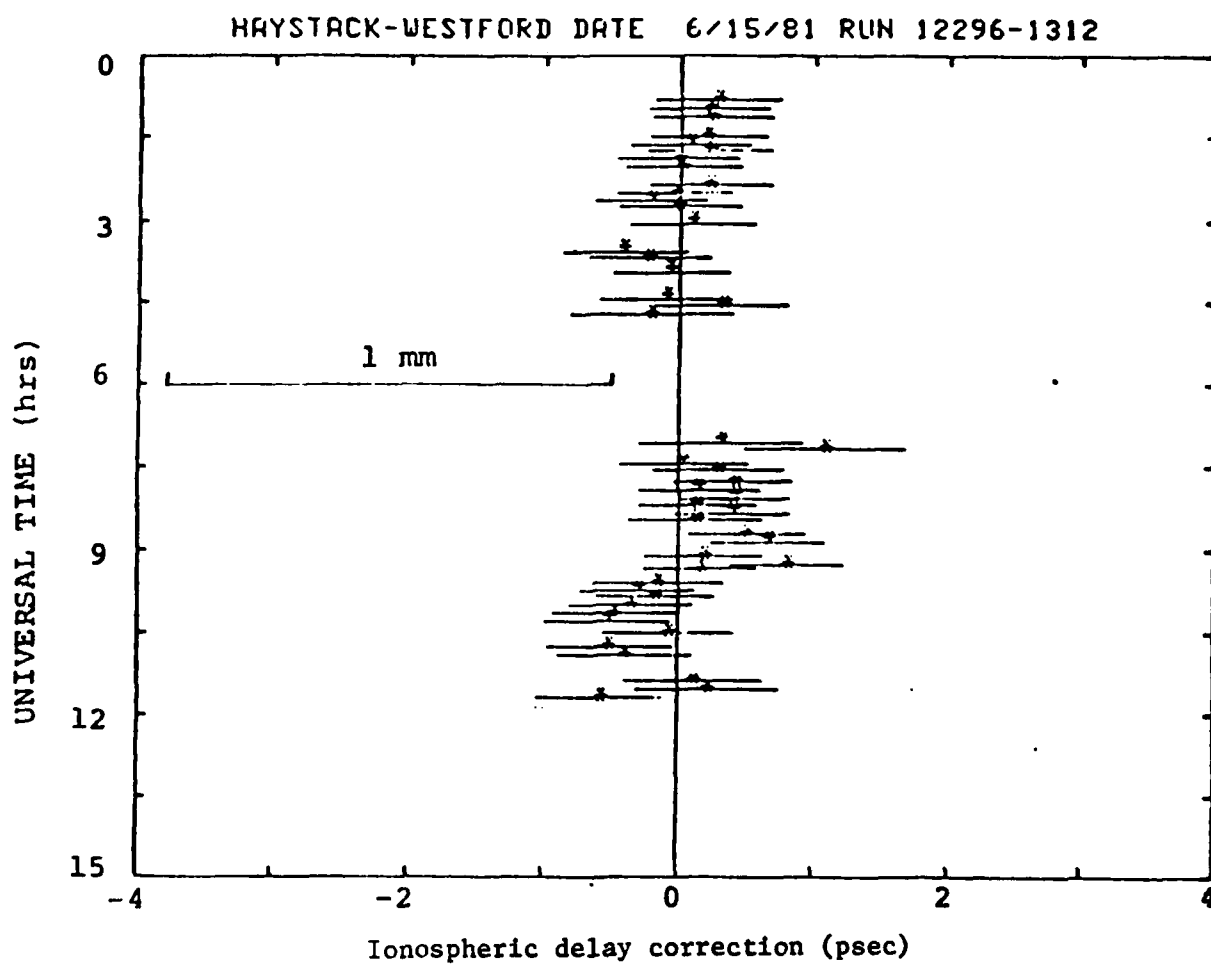Ionospheric delay correction (psec)

Figure 4.1.2.5.c   Ionospheric delay correction at X-band computed from
X- and S-band phase delays (same as Figure 4.1.2.5.b
except plotted on an enlarged scale)

HAYSTACK-WESTFORD DATE   6/15/81 RUN 12296-1312

spheric delay at X-band for the Haystack-Westford baseline
(the length of this baseline ≈1.24 km), measured from both
group and phase delays. Figure (5.1.2.5c) is the same as
Figure (4.1.2.5b), but plotted on a much larger scale. The
ionospheric delay measurements on this baseline should only
be sensitive to small scale ionospheric density variations.
The RMS scatter of the group-delay ionospheric measurements
is 15.3 psec. The expected scatter of the results, assuming
that the ionospheric delay is zero is 9.0 psec. The differ-
ence between the observed scatter and the expected could be
accounted for by ionospheric scintillations with 12.4 psec
RMS at X-band. This RMS places a bound on the contribution of
scintillation to the X-band group delay measurements of the
ionospheric delay. We believe that a large portion of the
12.4 psec additional scatter is due, not to scintillation,
but to instrumental errors which are not accounted for in
computing the group-delay uncertainties (see discussions in
Chapter 3).

The phase-delay measurements of the ionospheric delay on
the Haystack-Westford baseline show a scatter which is
consistent with the uncertainties of the phase delays. We
have shown these results in two figures, the first plotted at
the same scale as the group-delay measurements of the iono-
spheric delay (Figure 4.1.1.5b) and the other at a much
larger scale (Figure 4.1.2.5c).

The RMS scatter of these phase-delay results is 0.34
psec. We see no evidence of scintillation at all in these

results because the scatter of the results is 0.8 times the expected scatter of these measurements, given the uncertainities of the phase delays used in the calculations which indicates that the presence of additional noise sources in unlikely.

These results are of course not definitive but they do indicate that scintillation should not be a major source of error (at least in the June, 1981 experiment).

In Table 4.1.2.2 we summarize the expected error in the dual-frequency ionospheric-delay calibration due to each of the contributions discussed in this section. Even the largest term which would be systematic in nature is less than 1 mm, i.e. the term due to magnetic field. The magnetic-field error source could be reduced to a very small amount by the adoption of tri-frequency receivers, i.e. to first order the magnetic field contribution to the group delays is inversely proportional to frequency cubed. Hence by observing in three frequency bands we could separate this contribution from the inverse frequency squared term. An alternative approach would be to combine the group and the phase delays in a dual-frequency system. At present it seems that the instrumental errors in the group delays are too large for the magnetic field term to be successfully recovered using this technique (see Section 4.2).

Scintillation could potentially cause large random errors in the ionospheric delay calibration, but the results from the Haystack-Westford baseline seem to bound this error

Table 4.1.2.2 Summary of the magnitudes of the errors in the dual frequency ionospheric correction

Term	Error in X-band ionospheric delay correction		Error in estimated non-dispersive delay	
	phase (psec)	group (psec)	phase (psec)	group (psec)
A. Magnetic field	1.2	2.4	0.9	1.8
B. Damping	0.0002	0.0007	0.0002	0.0007
C. Higher order terms	0.02	0.04	0.02	0.04
D. Bending	0.1	0.4	0.1	0.04
E. Scintillation[*]	RMS <1.8	RMS <12.0[‡]	RMS <2.7	RMS <18.0[‡]

The values in this table are appropriate for observations at 8.4 GHz and 2.3 GHz.

---

[*] The contribution of the scintillation term is expressed as an RMS because its contribution will be random.

[‡] These values are based on the VLBI measurements of the ionospheric delay on a short baseline (Haystack-Westford ≈1.24 km length) in June, 1981.

at <12 psec and at <0.34 psec for the group and phase delays, respectively.

We have now discussed the theoretical limitations of the dual-frequency ionospheric correction. In Section 4.2 we will investigate the actual performance of this system by trying to eliminate the phase-delay ambiguities using the group-delay observations.

4.2 The phase delay ambiguity elimination algorithms

The technique of eliminating the phase delay ambiguities is essentially an application of Equations (4.1.1.16) through (4.1.1.19). The procedure we have adopted is to firstly calculate the ionospheric delay correction and the "non-dispersive" delay from the group delay observations at X- and S-band. These values are then used to predict the X- and S-band phase delays. The number of phase delay ambiguities are then calculated such that the phase delays (calculated from the visibility phases) will be within half an ambiguity of the predicted phase delays. We then reverse the process and predict the expected values of the group delays based on the phase delays. This final stage is useful in analyzing errors in the group delays.

Prediction of phase delay ambiguities based on group delays is a new technique. In order to check these procedures we will first analyze experiments which allow the phase delay ambiguities to be removed independently of the group delays.

These experiments, carried out in July 1980 and June 1981,
consisted of observing two sources, close in the sky, which
could be repeatably observed in a cyclical manner for many
hours (six hours in July 1980 and twelve hours in June 1981
for one pair, $0^\circ.5$ apart in the sky, and three hours for a
second pair $10^\circ.0$ apart). During the short interval between
successive observations on each source, we can model the
change in phase to much better than half an ambiguity, in most
cases, thus allowing each phase delay observation to be
"connected" to the ones before and after it.

Before discussing the results of these experiments we
will first study the problem of phase delay ambiguity predic-
tion. In particular we will need to know the effects of
random errors in the group delays on the phase delay predic-
tions. We will also need to know the effects of phase delay
ambiguities on the prediction of the group delays.

The phase-delay-ambiguity prediction algorithms are
derived from Equations (4.1.1.16) through (4.1.1.19). We
reproduce these equations here, with an additional term added
to each equation to account for observation noise (either
random or systematic). The relationships between our various
observations are:

$$\tau_g^x = \tau_n + I_x + \epsilon_g^x \tag{4.2.1}$$

$$\tau_g^s = \tau_n + \beta I_x + O_g^s + \epsilon_g^s \tag{4.2.2}$$

$$\tau_p^x = \tau_n - \alpha I_x + O_p^x + \epsilon_p^x \tag{4.2.3}$$

$$\tau_p^S = \tau_n - \gamma I_x + O_p^S + \varepsilon_p^S \tag{4.2.4}$$

where ε_g^x, ε_g^s, ε_p^x, and ε_p^s are additive noise contributions to the X- and S-band group delays, and the X- and S-band phase delays, respectively. The phase delays given in Equations (4.2.3) and (4.2.4) are not normally known. The quantities which we do know are the visibility phases, ϕ_x and ϕ_S (corrected for feed rotation; see Section 4.1.2), which are related to the phase delays by (see Appendix A)

$$\tau_p^x = (2\pi N_x + \phi_x)/\omega_x \tag{4.2.5}$$

and

$$\tau_p^s = (2\pi N_s + \phi_s)/\omega_s \tag{4.2.6}$$

where ω_x and ω_s are the (angular) RF's in each band. The (integral) numbers N_x and N_s are the phase delay ambiguities. These are the numbers which we wish to predict based on the group delay measurements. Normally, we make crude estimates of N_x and N_s, and then calculate corrections, ΔN_x and ΔN_s, to these estimates in order to find the values which satisfy Equations (4.2.1)-(4.2.4). The phase delays calculated from the crude estimates of the ambiguities and the visibility phases will be called the a priori phase delays.

An obvious approach to estimating N_x and N_s is to manipulate Equations (4.2.1) and (4.2.2), and form estimates of the phase delays based on Equations (4.2.3) and (4.2.4). A simple rearrangement of Equations (4.2.1) and (4.2.2) yields,

$$I_x = (\tau_g^x - \tau_g^s)/(1-\beta) + (O_g^s + \varepsilon_g^s - \varepsilon_g^x)/(1-\beta) \tag{4.2.7}$$

and

$$\tau_n = (\beta\tau_g^x - \tau_g^s)/(\beta-1) + (O_g^s + \varepsilon_g^s - \beta\varepsilon_g^x)/(\beta-1) \quad . \qquad (4.2.8)$$

Since τ_g^x and τ_g^s are the only measured quantities in Equations (4.2.7) and (4.2.8), we can form estimates of I_x and τ_n using the first terms on the right hand sides of Equations (4.2.7) and (4.2.8). These estimates will be given by

$$(\hat{I}_x)_{gg}^{xs} = (\tau_g^x - \tau_g^s)/(1-\beta) \qquad (4.2.9)$$

and

$$(\hat{\tau}_n)_{gg}^{xs} = (\beta\tau_g^x - \tau_g^s)/(\beta-1) \qquad (4.2.10)$$

and they will be related to I_x and τ_n by

$$(\hat{I}_x)_{gg}^{xs} = I_x - (O_g^s + \varepsilon_g^s - \varepsilon_g^x)/(1-\beta) \qquad (4.2.11)$$

and

$$(\hat{\tau}_n)_{gg}^{xs} = \tau_n - (O_g^s + \varepsilon_g^s - \beta\varepsilon_g^x)/(\beta-1) \qquad (4.2.12)$$

where the superscripts xs and corresponding subscripts gg denote the observation types used to make the estimate. (In practice, an estimate of O_g^s is made before Equations (4.2.9) and (4.2.10) are evaluated. The O_g^s given in Equations (4.2.11) and (4.2.12) can then be treated as a correction to the a priori value.)

We may now make predictions of the phase delays based on Equations (4.2.11) and (4.2.12), and Equations (4.2.3) and

-177-

(4.2.4). The predictions are

$$(\hat{\tilde{\tau}}_p^x)_{gg}^{xs} = (\hat{\tilde{\tau}}_n)_{gg}^{xs} - \alpha(\hat{I}_x)_{gg}^{xs} \qquad (4.2.13)$$

and

$$(\hat{\tilde{\tau}}_p^s)_{gg}^{xs} = (\hat{\tilde{\tau}}_n)_{gg}^{xs} - \gamma(\hat{I}_x)_{gg}^{xs} \quad . \qquad (4.2.14)$$

Some simple algebra shows that the difference between the
a priori phase delays and the predictions will be

$$\tilde{\tau}_p^x - (\hat{\tilde{\tau}}_p^x)_{gg}^{xs} = -2\pi\Delta N_x/\omega_x + O_p^x + \varepsilon_p^x + [(1+\alpha)(O_g^s+\varepsilon_g^s) - (\alpha+\beta)\varepsilon_g^x]/(\beta-1),$$

$$(4.2.15)$$

and

$$\tilde{\tau}_p^s - (\hat{\tilde{\tau}}_p^s)_{gg}^{xs} = -2\pi\Delta N_s/\omega_s + O_p^s + \varepsilon_p^s + [(1+\gamma)(O_g^s+\varepsilon_g^s) - (\gamma+\beta)\varepsilon_g^x]/(\beta-1)$$

$$(4.2.16)$$

where $\tilde{\tau}_p^x$ and $\tilde{\tau}_p^s$ are the a priori estimates of τ_p^x and τ_p^s based
on assumed values of, N_x^o and N_s^o, N_x and N_s (i.e., $\tilde{\tau}_p^x = (2\pi N_x^o+\phi_x)/\omega_x$); and the actual number of ambiguities will be
given by $N_x=N_x^o+\Delta N_x$. Similar relationships hold for the S band
ambiguities.

Equations (4.2.15) and (4.2.16) provide us with a method
of calculating ΔN_x and ΔN_s, but we should notice that these
estimates will be corrupted by the offsets O_g^s, O_p^x and O_p^s, and
the noise contributions to each of the delay measurements ε_g^x,
ε_g^s, ε_p^x, and ε_p^s.

It will be instructive to consider each of these corrup-
tions (the offsets and the noise), separately. The offsets,
if constant, are more of an annoyance than a problem. (In our
formulation any variations in the offsets, by definition, are

absorbed into the noise contributions.) If we rewrite Equations (4.2.15) and (4.2.16) excluding the noise terms, we have

$$\tilde{\tau}_p^X - (\hat{\tau}_p^X)_{gg}^{XS} = -2\pi\Delta N_X/\omega_X + O_p^X + (1+\alpha)O_g^S/(\beta-1) \qquad (4.2.17.a)$$

and

$$\tilde{\tau}_p^S - (\hat{\tau}_p^S)_{gg}^{XS} = -2\pi\Delta N_S/\omega_S + O_p^S + (1+\gamma)O_g^S/(\beta-1) \quad . \qquad (4.2.17.b)$$

We have two equations above, and 5 unknowns, although two of these unknowns (ΔN_X and ΔN_S) must be integral. Therefore, from the first pair of prediction differences, we can set the non-integral parts of the prediction differences equal to the offsets. Subsequent prediction differences could then be used to check the constancy of the offsets. We should note that there is no unique separation of the ambiguity corrections and the offsets, i.e. ΔN_X and ΔN_S can be changed by any integral values (but the same ones for all observations) and these changes can be absorbed into the offsets. These arbitrary shifts of the phase delays, by constant amounts, pose no serious problem because they are indistinguishable from an epoch offset between the hydrogen-maser frequency standards at each site.

We now turn our attention to the noise contribution to the prediction differences. These contributions do pose a serious problem for ambiguity prediction, if the noise is large. Equations (4.2.15) and (4.2.16), omitting the "O"

terms, are

$$\tilde{\tau}_p^X - (\hat{\tau}_p^X)_{gg}^{XS} = -2\pi \Delta N_x/\omega_x + \epsilon_p^X + [(1+\alpha)\epsilon_g^S - (\alpha+\beta)\epsilon_g^X]/(\beta-1)$$

$$(4.2.18.a)$$

and

$$\tilde{\tau}_p^S - (\hat{\tau}_p^S)_{gg}^{XS} = -2\pi \Delta N_s/\omega_s + \epsilon_p^S + [(1+\gamma)\epsilon_g^S - (\gamma+\beta)\epsilon_g^X]/(\beta-1) \quad .$$

$$(4.2.18.b)$$

In the following discussion we will consider the effects of several different forms of noise on the prediction process. Before doing so we will compress the notation of Equations (4.2.18). We will write the equations as

$$\Delta P_p^X = -2\pi \Delta N_x/\omega_x + \epsilon_p^X + (C_x)_g^S \epsilon_g^S - (C_x)_g^X \epsilon_g^X \qquad (4.2.19.a)$$

and

$$\Delta P_p^S = -2\pi \Delta N_s/\omega_s + \epsilon_p^S + (C_s)_g^S \epsilon_g^S - (C_s)_g^X \epsilon_g^X \qquad (4.2.19.b)$$

where ΔP_p^X and ΔP_p^S are the prediction differences at X- and S-band, respectively, and $(C_x)_g^S$, $(C_x)_g^X$, $(C_s)_g^S$, and $(C_s)_g^X$ are the coefficients which multiply the group delay errors in Equations (4.2.18). The values of these coefficients for $\omega_x/(2\pi) \approx 8.3$ GHz and $\omega_s/(2\pi) \approx 2.3$ GHz are $(C_x)_g^S \approx 0.16$, $(C_x)_g^X \approx 1.16$, $(C_s)_g^S \approx 1.16$, and $(C_s)_g^X \approx 2.16$, for $\alpha \approx 1.0$, and $\beta \approx \gamma \approx 13.5$. ($\alpha$, β and γ were the ratio of the respective effective frequencies squared; see Equations (4.1.1.12) to (4.1.1.16).)

Clearly from Equations (4.2.19) we can see a coupling of the effects on the prediction errors, of the errors in the group delays at X- and S-bands, i.e., an error in the S-band

group delay will affect the prediction of both the S-band and the X-band phase-delay ambiguities. We may statistically calculate this coupling by computing the variance-covariance matrix (or simply covariance matrix) of the prediction errors. If the group-delay errors are uncorrelated, and if we neglect the error contribution of the phase delays (these are approximately 40-fold smaller than the group delay errors), the covariance matrix of the prediction errors will be given by

$$
V^{\Delta P} = \begin{bmatrix} \langle (\Delta P_p^X)^2 \rangle & \langle (\Delta P_p^X)(\Delta P_p^S) \rangle \\ \langle (\Delta P_p^X)(\Delta P_p^S) \rangle & \langle (\Delta P_p^S)^2 \rangle \end{bmatrix}
$$

$$
= \begin{bmatrix} 1.35(\sigma_g^X)^2 + 0.026(\sigma_g^S)^2 & -2.51(\sigma_g^X)^2 + 0.019(\sigma_g^S)^2 \\ \text{symmetric} & 4.67(\sigma_g^X)^2 + 1.35(\sigma_g^S)^2 \end{bmatrix}
$$

$$(4.2.20)$$

where $(\sigma_g^X)^2$ and $(\sigma_g^S)^2$ are the variances of the X- and S-band group delays, respectively, and where we have substituted the numerical values of the coefficients in the covariance matrix.

In order to resolve correctly the phase delay ambiguities, we need the prediction errors to be much less than one cycle at each frequency. In Figure 4.2.1, we plot the bounds on the group delay uncertainties such that the phase-delay prediction differences will have uncertainties less than 0.030 nsec (≈ 0.25 cycles) and 0.110 nsec (≈ 0.25 cycles) at X- and S-band, respectively. (These uncertainties represent 1/4 of a

Figure 4.2.1   X- and S-band group delay standard deviation requirements necessary for phase-ambiguity elimination
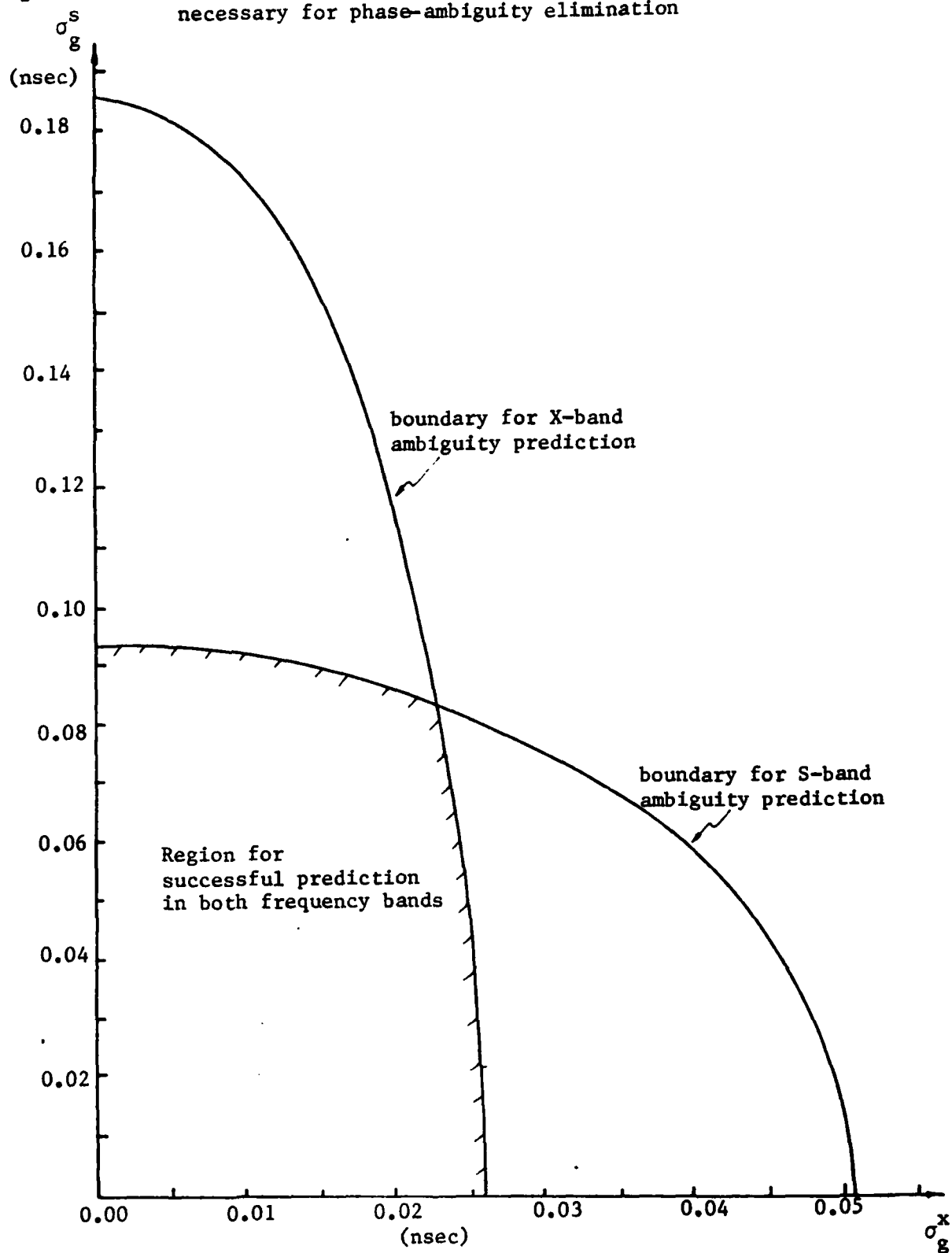
cycle at each frequency.) From the figure we see that in order to simultaneously predict both the X- and S-band ambiguities, we need to know the group delays with uncertainties of <0.025 nsec and <0.090 nsec, respectively, at these two frequencies. These bounds put very stringent accuracy requirements on the Mark III system if we want to predict phase-delay ambiguities reliably. In Section 4.3, we will investigate whether or not the Mark III system can meet these accuracy requirements.

After the estimates of the phase delay ambiguities have been obtained, the resultant phase delays are used to obtain estimates of the group delays. The differences between the observed values of the group delays and their predicted values can be used to study the errors in the group delays and to detect any remaining ambiguity errors. We will refer to these differences as "apparent group delay errors."

The relationship between the observed group delays and the predictions based on the phase delays may be derived in a manner similar to that used to derive Equations (4.2.15) and (4.2.16). The predicted group delays will be given by

$$(\hat{\tau}_g^x)_{pp}^{xs} = ((\gamma+1)\tilde{\tau}_p^x - (1+\alpha)\tilde{\tau}_p^s)/(\gamma-\alpha) \qquad (4.2.21)$$

and

$$(\hat{\tau}_g^s)_{pp}^{xs} = ((\gamma+\beta)\tilde{\tau}_p^x - (\beta+\alpha)\tilde{\tau}_p^s)/(\gamma-\alpha) \qquad (4.2.22)$$

and the errors in the predictions will be

$$\tau_g^x - (\hat{\tau}_g^x)_{pp}^{xs} = \varepsilon_g^x - (\gamma+1)(O_p^x + 2\pi\Delta N_x/\omega_x + \varepsilon_p^x)/(\gamma-\alpha)$$

$$+ (1+\alpha)(O_p^s + 2\pi\Delta N_s/\omega_s + \varepsilon_p^s)/(\gamma-\alpha)$$

$$(4.2.23.a)$$

and

$$\tau_g^s - (\hat{\tau}_g^s)_{pp}^{xs} = \varepsilon_g^s - (\gamma+\beta)(O_p^x + 2\pi\Delta N_x/\omega_x + \varepsilon_p^x)/(\gamma-\alpha)$$

$$+ (\alpha+\beta)(O_p^s + 2\pi\Delta N_s/\omega_s + \varepsilon_p^s)/(\gamma-\alpha)$$

$$(4.2.23.b)$$

As we did earlier in this section, we will consider parts of the above equations separately. We consider firstly the effects of the phase-delay ambiguities. In this case,

$$\tau_g^x - (\hat{\tau}_g^x)_{pp}^{xs} = -0.139\Delta N_x + 0.070\Delta N_s \qquad (nsec) \quad (4.2.24.a)$$

and

$$\tau_g^s - (\hat{\tau}_g^s)_{pp}^{xs} = -0.259\Delta N_x + 0.510\Delta N_s \qquad (nsec) \quad (4.2.24.b)$$

where we have substituted the approximate numerical values for the constants given in Equations (4.2.23). The above equations highlight one of the disadvantages of the choice of X- and S-band for the Mark III system when phase-delay ambiguity prediction is attempted. If $\Delta N_x=2$ and $\Delta N_s=1$, then the difference between the observed X-band group delay and the value predicted from the phase delays would differ by only 0.001 nsec. Of course, in this case the difference between the observed S-band group delay and its prediction would be 0.761 nsec. However, there are cases where the noise contribution

to the S-band group delays could be this large (see Section 4.3). In this case, these ambiguity errors could not easily be detected.

In general, we see from Equations (4.2.24) that ambiguities in the phase delays produce differences between the observed group delays and the predictions of multiples of ≈ 0.070 nsec and ≈ 0.26 nsec at X- and S-band, respectively. The process of phase-delay-ambiguity prediction is somewhat harder than may have been initially thought, given the ambiguity spacings of ≈ 0.12 nsec and ≈ 0.44 nsec at X- and S-band, respectively, because of the coupling of the X- and S-band errors. We should also notice that changing the ambiguities at one frequency will affect the implied group delay errors at both X- and S-band.

The noise contributions to the apparent group-delay errors may also be calculated from Equations (4.2.23). In this case,

$$\tau_g^X - (\hat{\tau}_g^X)_{pp}^{XS} = \epsilon_g^X - 1.16\epsilon_p^X + 0.16\epsilon_p^S \qquad (4.2.25.a)$$

and

$$\tau_g^S - (\hat{\tau}_g^S)_{pp}^{XS} = \epsilon_g^S - 2.16\epsilon_p^X + 1.16\epsilon_p^S \qquad (4.2.25.b)$$

where again we have substituted the numerical values of the coefficients. If we firstly consider random error only, then $\epsilon_p^X \ll \epsilon_g^X$, and $\epsilon_p^S \ll \epsilon_g^S$, and the major contribution to the differences given above will be from the group delays. If the errors in the group delays are too large then the phase delay ambiguities will be resolved incorrectly (if no other means

are available for determining the phase delay ambiguities),
and the differences between the observed group delays and
their estimates based on phase delays will be smaller than the
actual errors.

Figure 4.2.2 allows us to study the "folding" of group
delay errors in more detail using a graphical technique.   In
this figure we have shown, as a function of ϵ_g^x and ϵ_g^s, the
boundary of those errors that will lead to the correct
ambiguity predictions. (This boundary is the central diamond
shaped region.)   If the errors in the group delays fall out-
side of this region (see, for example, point labeled A), then
the phase delay ambiguities will be incorrectly estimated.
When the errors in the group delays are computed from predict-
ed phase delay ambiguities, the apparent errors will fall
within the central region (see point A'). Consequently,
independent of the original errors in the group delays, the
apparent errors will always be inside the small central
region.   Figure 4.2.2 will be useful when we are analyzing
data because it will give us a convenient means for determin-
ing the relationship between the actual errors in the group
delays and the apparent errors calculated from the phase
delays (if the phase-delay ambiguities are determined solely
from the group delays).

So far we have been discussing the X- and S-band phase
delays separately.   However, if we want to benefit from the
precision of the phase delays, we should compute the non-dis-
persive delay from the phase delays.   Again, we will be inter-

Figure 4.2.2 Transformation of actual group delay errors to apparent group delay errors
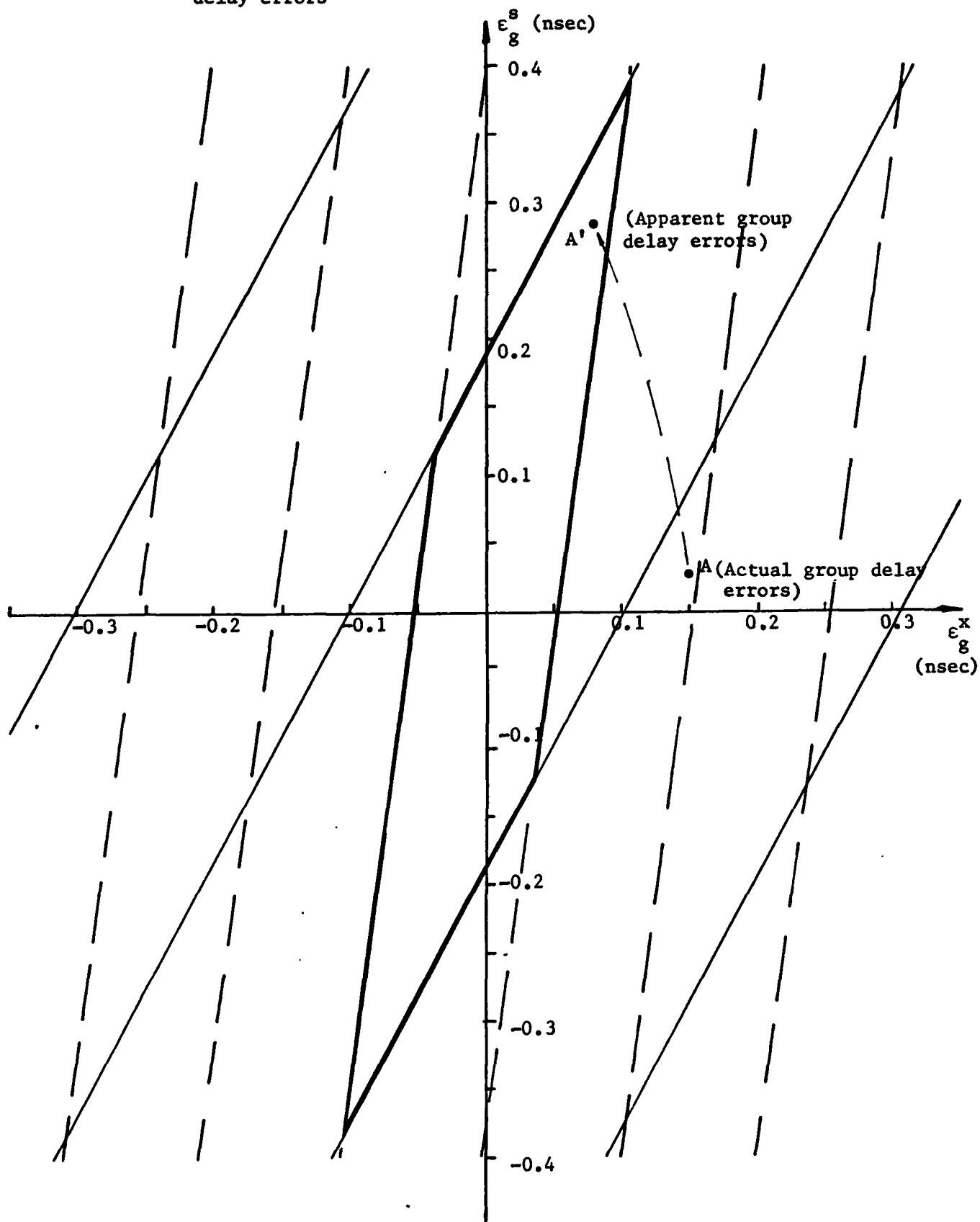
ested in the effects of ambiguities in the phase delays on the estimate of the non-dispersive delay. Firstly, we will examine the expressions for the non-dispersive delay estimates computed from the group delays. From Equations (4.2.12), we have for the group delays

$$(\hat{\tau}_n)^{xs}_{gg} = \tau_n + 1.08\,\epsilon^x_g - 0.08\,\epsilon^s_g \ . \tag{4.2.26}$$

The expression for the non-dispersive delay estimated from the phase delays may be derived in a manner similar to that used to derive Equation (4.2.12). The result is

$$(\hat{\tau}_n)^{xs}_{pp} = \tau_n + 1.08\,(2\pi\Delta N_x/\omega_x + \epsilon^x_p) - 0.08\,(2\pi\Delta N_s/\omega_s + \epsilon^s_p). \tag{4.2.27}$$

We should notice in Equation (4.2.27) that $(\hat{\tau}_n)^{xs}_{pp}$ has two ambiguity spacings, one of 0.13 nsec due to the X-band phase delay ambiguity, and the other of 0.035 nsec due to the S-band phase delay ambiguity. This latter ambiguity spacing clearly indicates the need for a reliable prediction of the S-band phase delay ambiguities. The effect of an S-band ambiguity (0.035 nsec) would be difficult to detect in the postfit residuals from a geodetic solution using $(\hat{\tau}_n)^{xs}_{pp}$ because the errors in the geodetic models for the delays have RMS scatters of ≈ 0.090 nsec (see Chapter 5).

In anticipation of the results which will be presented in Section 4.3, in the remainder of this section we discuss the effects of a number of error types on the differences between the various delay measurements and their predicted values.

We now investigate the detection of errors in both the group and phase delays, by comparing the differences between Equations (4.2.25.a) and (4.2.25.b) and Equations (4.2.26) and (4.2.27). In view of the results which will be presented in the next section, we will discuss two classes of systematic errors in the observations. Firstly, we consider the case when the S-band group delay measurements have a systematic, time dependent, error, and all of the other measurement errors are assumed to be zero. We see from Equation (4.2.25.a) that the differences between observed X-band group delays and their predicted values from the phase delays will be unaffected by such an error (i.e., $\epsilon_g^X = \epsilon_p^X = \epsilon_p^S = 0$). The S-band group delay differences will be non-zero, and will equal the errors in the S-band group delays. The non-dispersive delay estimates will differ by -0.08 times the errors in the S-band group delays. We can easily detect this error because the X-band group delay predictions will agree with the observed values. Similarly, if the X-band group delays are in error (but not the S-band group delays), then the X-band group delay predicted and observed values would differ. In this case, the S-band group delay predictions would agree with the predicted values, hence allowing the X-band group delays to be isolated as the delay measure with an error.

Thus far, we have considered the phase delays to be error free because of the ≈40-fold smaller uncertainties, and their insensitivity to instrumental errors (see Section 3.1). However, there are certain errors which will affect the group

and phase delays equally. Such an error can arise in several different circumstances. For example, all dual-band observations made at Onsala, Sweden, use two radio telescopes, one for receiving X-band signals and the other, ≈600 m away from the X-band antenna, for receiving S-band signals. In this case, the S-band observations are corrected to the X-band radio telescope's position, using the coordinate differences between the two radio telescopes. Any errors in the relative positions of the S- and X-band antennas will introduce an error in the S-band delays referred to the X-band telescope's position. In addition, any differences in the atmospheric delay between the two antennas will introduce a similar type error. If we consider Equations (4.2.25) we see that if the S-band group and phase delays have equal errors then the differences between the observed and predicted values of the group delays at X- and S-band will be non-zero and equal. However, if we examine the expressions for the non-dispersive delays (Equations (4.2.26) and (4.2.27)) we see that this error will cause no difference between the estimates of the non-dispersive delay from the group and phase delays. Again, based on these properties we should be able to detect this type of error.

Of course, with actual measurements, all of the delay measurements will have some amount of error, and this will complicate the isolation of errors when we compare predictions and observed values of the various delay measures. We will investigate the magnitude of these errors by testing our

ambiguity elimination techniques on data which can be phase connected without the use of group delays.

4.3 Results from "phase-connection" experiments

When the studies of predicting phase delay ambiguities were started, it was hoped that a technique could be developed which would allow phase delays to be used in geodetic experiments. The initial studies soon indicated that, although the Mark III system could theoretically measure sufficiently accurate group delays to allow reliable ambiguity prediction, the actual measurements were of insufficient accuracy. Consequently, the studies on phase-delay-ambiguity prediction, have been used instead to study the quality of the group delays. In this section, we will therefore concentrate on errors in the group delays which have been estimated by computing estimates of the group delays from the phase delays.

The techniques used to carry out the ambiguity predictions were explained in Section 4.2. We demonstrate the application of these techniques in Figure 4.3.1. This figure shows the differences between the predicted and the phase-connected phase delays at both X- and S-band for the Effelsberg-Onsala interferometer. One aspect of this figure is immediately apparent. The differences between the predicted and the phase-connected phase delays at X- and S-band are not randomly distributed. They show systematic trends at both X- and S-band. We should also notice that the drift between
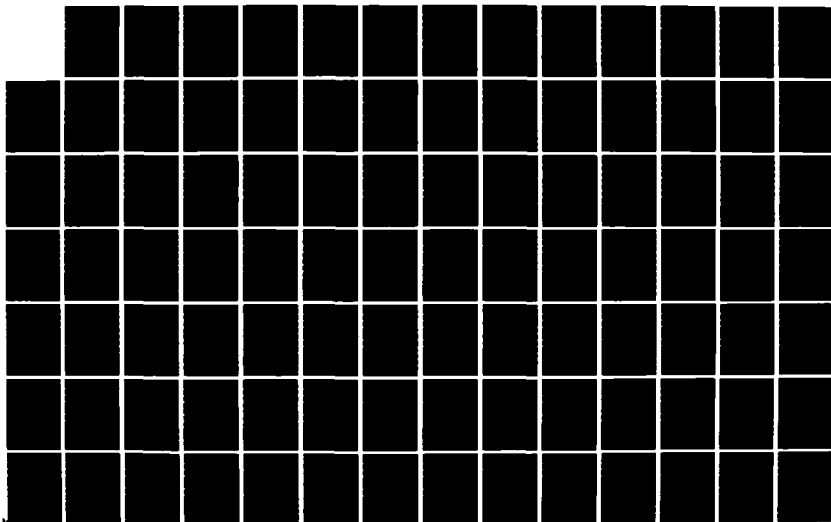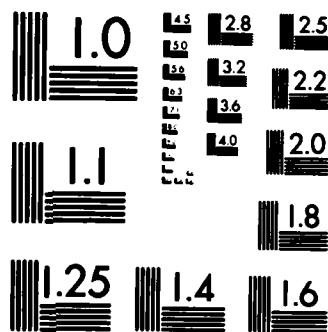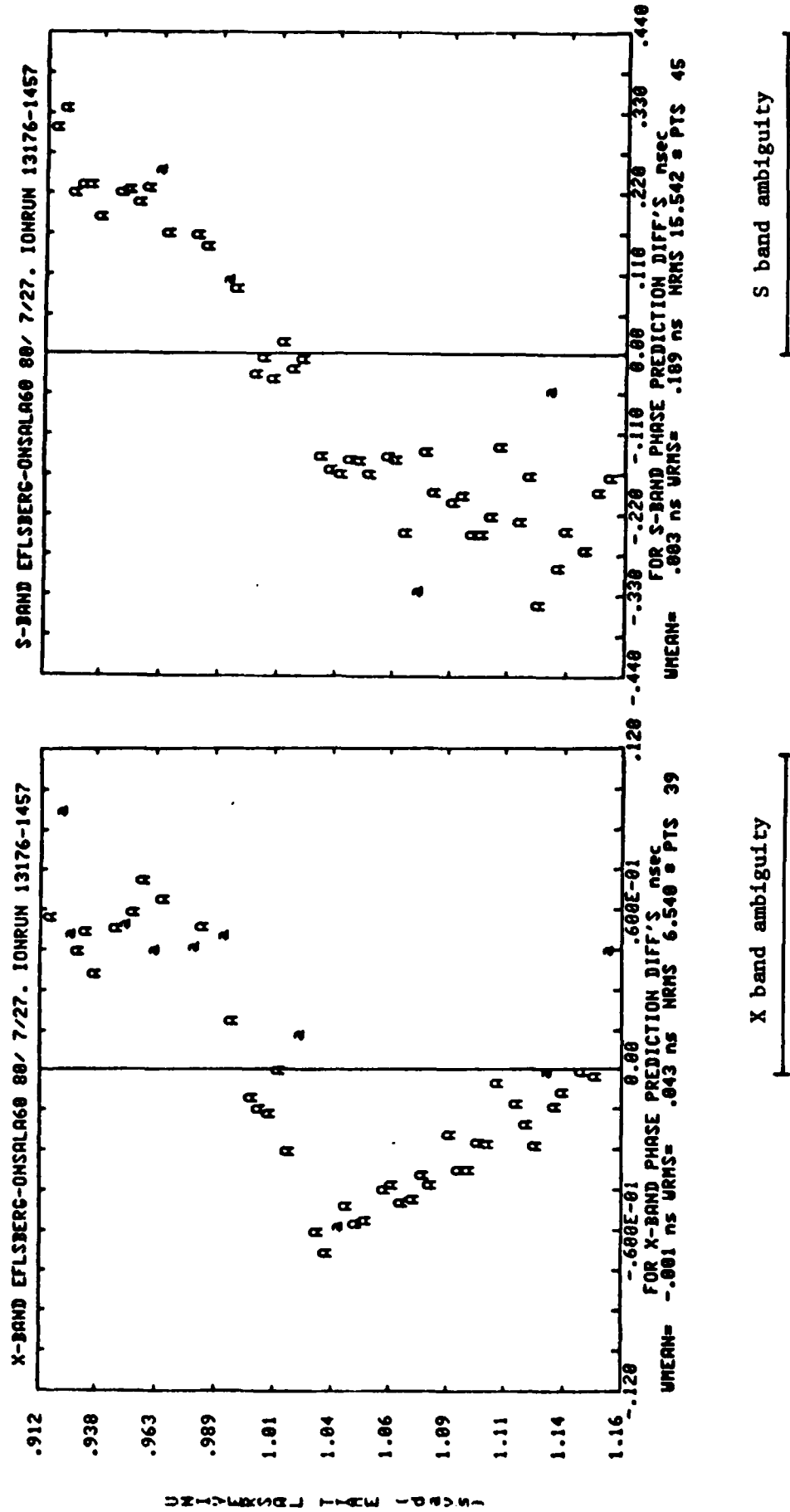
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

Figure 4.3.1 Differences between the predicted and the phase connected phase delays for observations of the radio source 3C 345 on the Effelsberg-Onsala baseline

the predictions and the phase-connected phase delays is suffi-
ciently large that, without the frequent observations of this
source which were available in these experiments, the phase
delay ambiguities would have been incorrectly eliminated. If
we now compare the observed group delays with their predicted
values (from the phase delays) we see the reason for the
systematic differences of the phase delay predictions (Figure
4.3.2). These differences will be called the "apparent group
delay errors." (We use the word "apparent" because these
differences could be due to remaining phase delay ambiguities
or errors in the phase delays, in addition to group delay
errors. However, any ambiguity errors will cause discontinu-
ities in the systematic trends.)

Most of the variations in the X-band group delays can be
explained by the effects of the spurious signal in the Onsala
X-band receiver (see Section 3.2). We show in Figure 4.3.3
the X-band apparent group delay errors after correcting for
the presense of the spurious signal. This correction reduces
the apparent errors, but there are still some small (≈0.03
nsec) differences. The apparent S-band group delay errors can
not be explained by the effects of a spurious signal.

In an attempt to determine the origin of these differ-
ences, we show in Figure 4.3.4, the phase calibration contri-
butions to the X- and S-band group delays at Effelsberg and
Onsala. Clearly, we cannot account for the S-band differences
with phase calibration errors. Nor can the difference be due
to an error in the relative positions of the two radio tele-

Figure 4.3.2 Differences between the observed and predicted group delays for the data shown in Figure 4.3.1
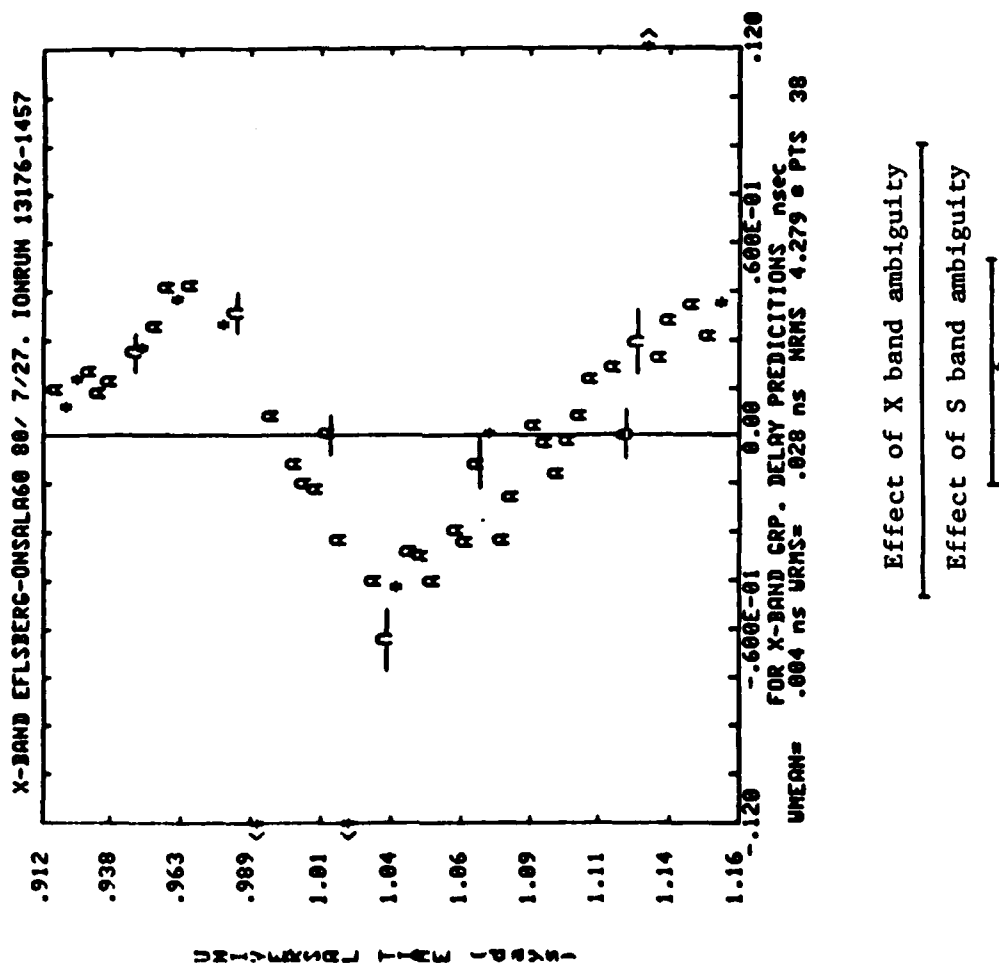
Figure 4.3.3 Differences between the X band observed and predicted
group delays for the Effelsberg-Onsala baseline after
correcting for the phase calibration error

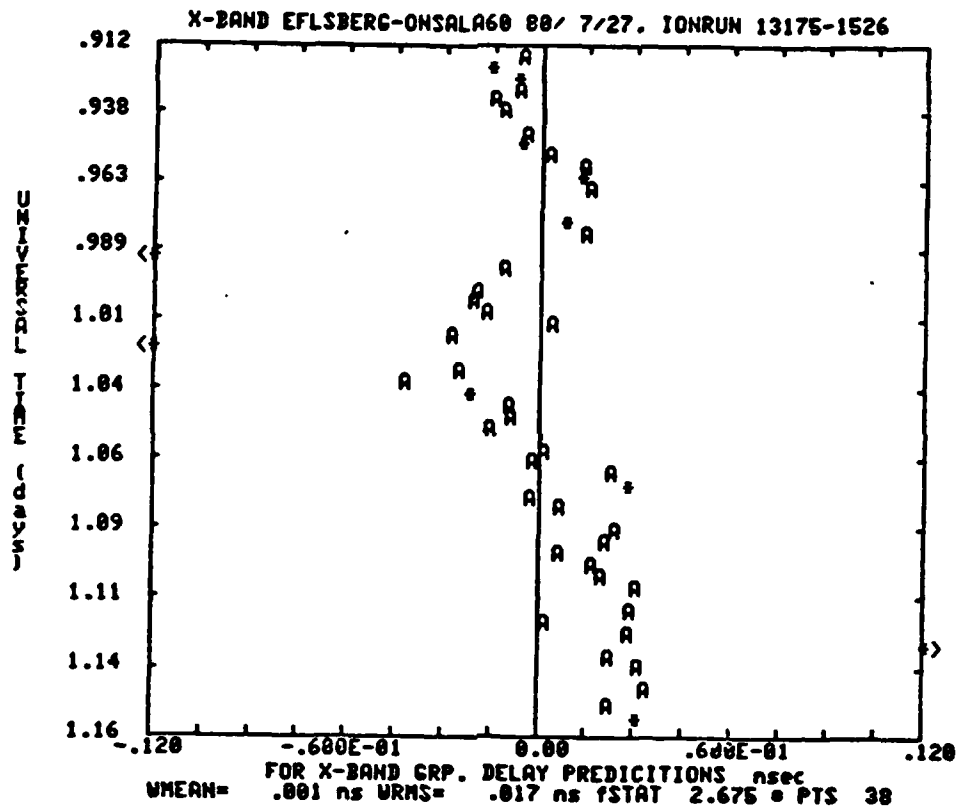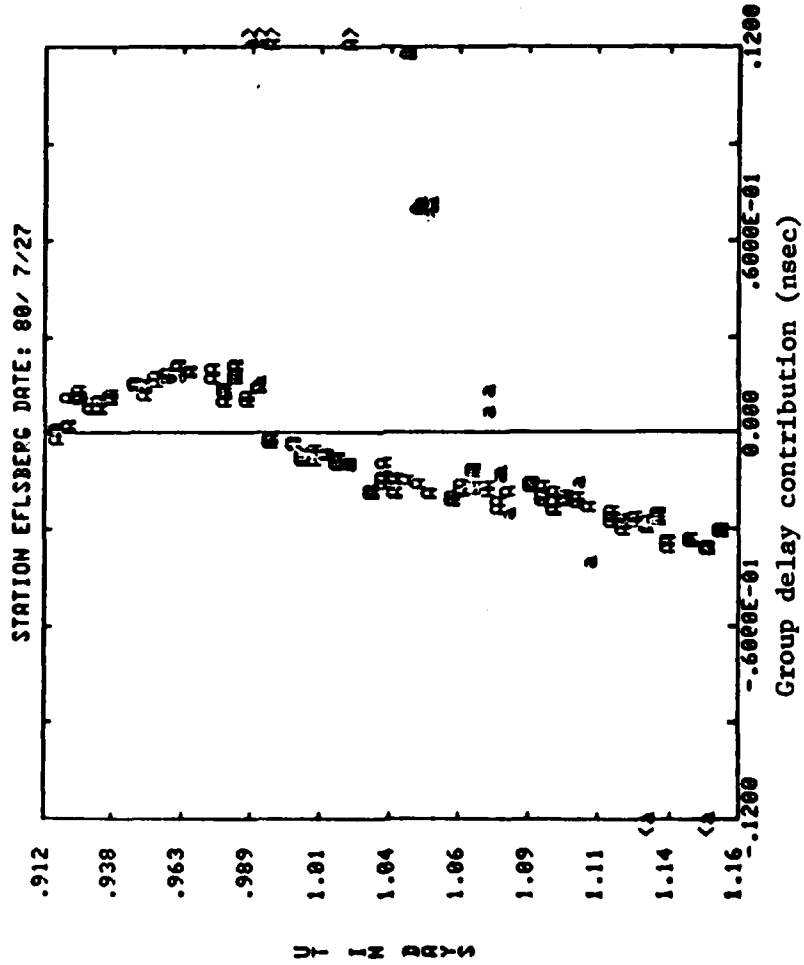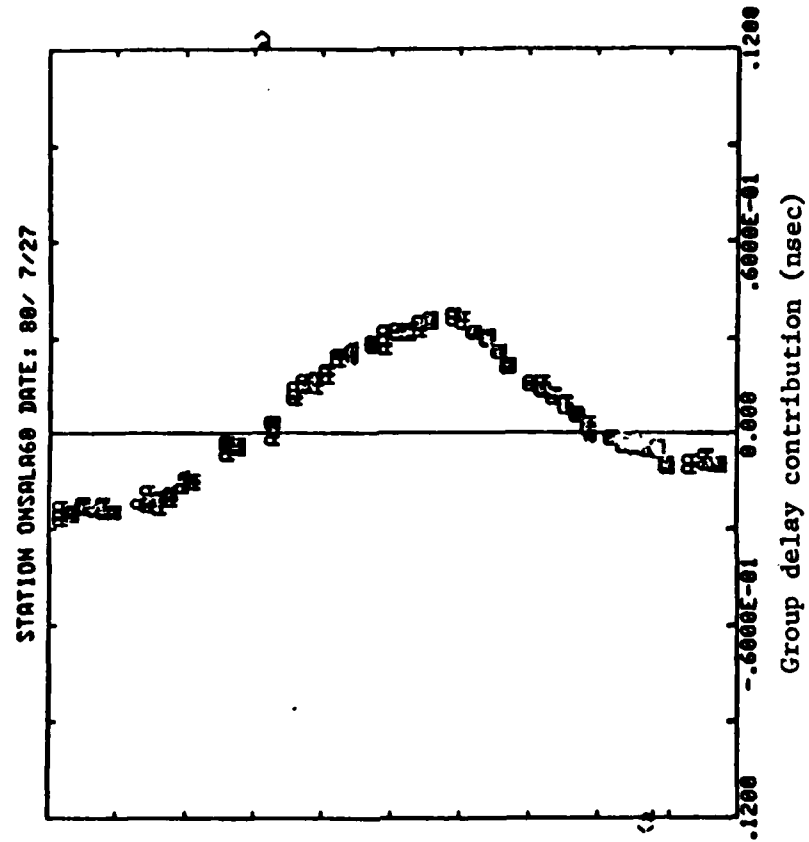X-BAND EFLSBERG-ONSALA60 80/ 7/27. IONRUN 13175-1526

Figure 4.3.4 Contributions of the phase calibration signal to group delays for the Effelsberg-Onsala baseline

a. Contribution at X band

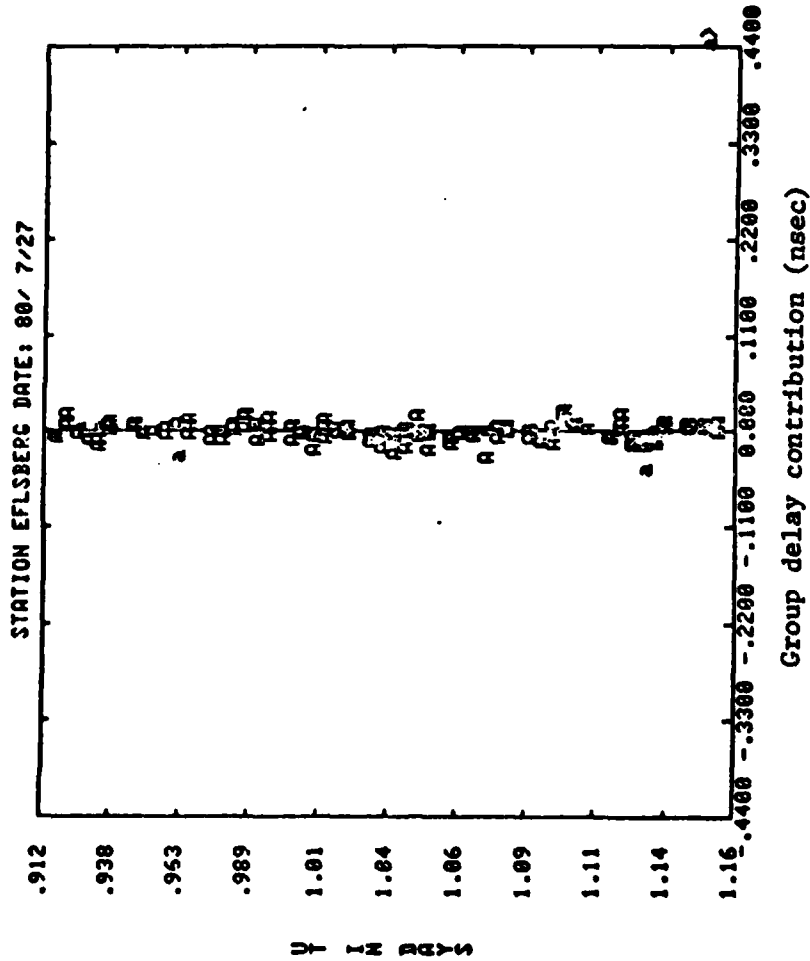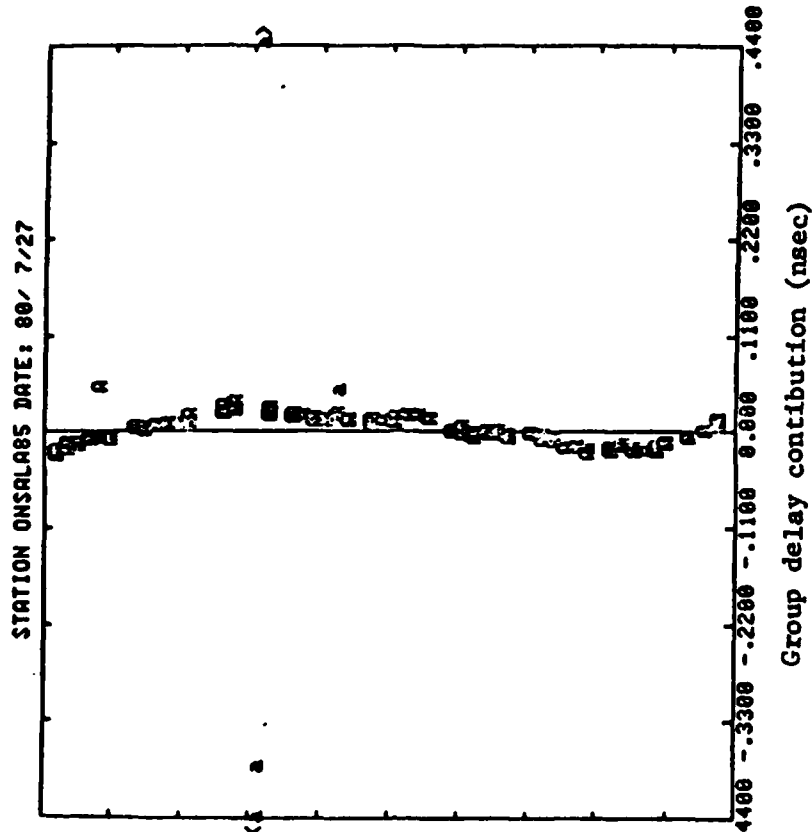b. Contribution at X band

Figure 4.3.4 Continued

c. Contribution at S-band

d. Contribution at S-band

scopes at Onsala (see discussions in Section 4.2), because if the differences were due solely to an error in the relative positions of the telescopes, then the apparent X-band group delay errors would equal the apparent S-band group delay errors. At this time the origin of the S-band differences is unknown. We will discuss these differences further in Section 4.4.

The Effelsberg-Onsala interferometer is not the only one to show systematic differences between the group delays and their predicted values from the phase delays. In Figure 4.3.5, we show the apparent group delay errors at X- and S-band for the several other interferometesr involved in the July 1980 experiment. (Note that the differences are plotted on different scales for X- and S-band, and within each band we use two different scales depending on the size of the differences.)

For all baselines the weighted-root-mean-sqaure (WRMS) scatter of the apparent group delay errors are greater than those expected from the signal-to-noise ratio of the observations. Of course, there will be apparent errors in group delays if there are phase-connection errors. However, we must remember that changing the phase delay ambiguities at one frequency will change the apparent errors in the group delays at both X- and S-band (see Section 4.2). Consequently, it becomes necessary to check the two group-delay differences when considering a phase-delay-ambiguity error as a possible cause for an apparent group delay error. In addition, the

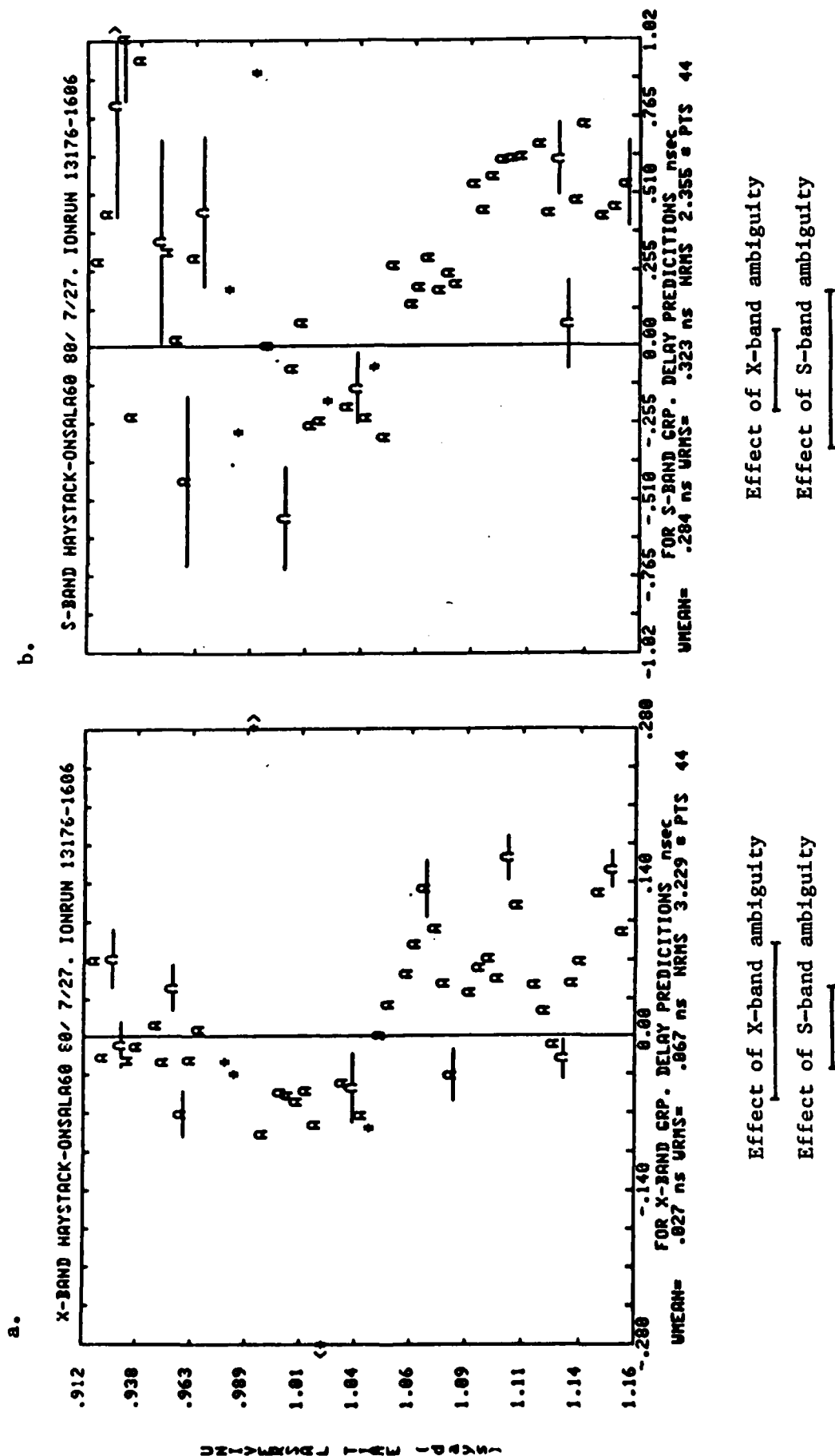Figure 4.3.5   Apparent group delay errors in two bands for several interferometers for the July 1980 experiment
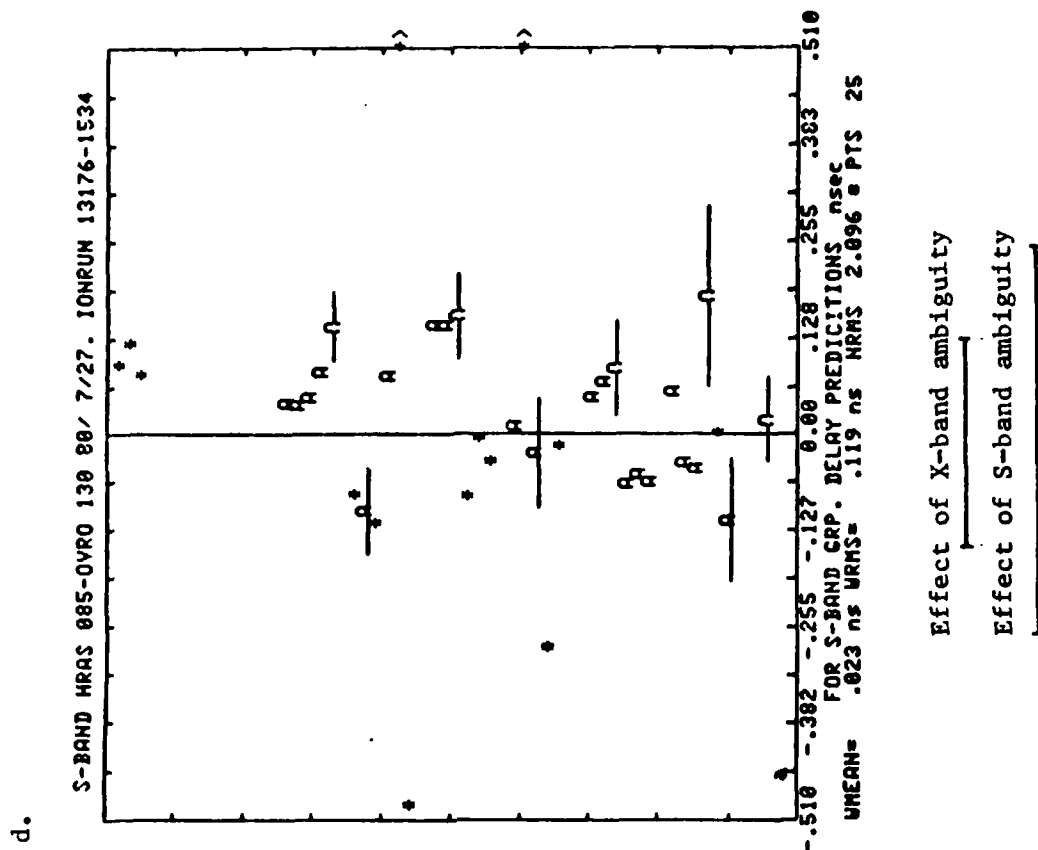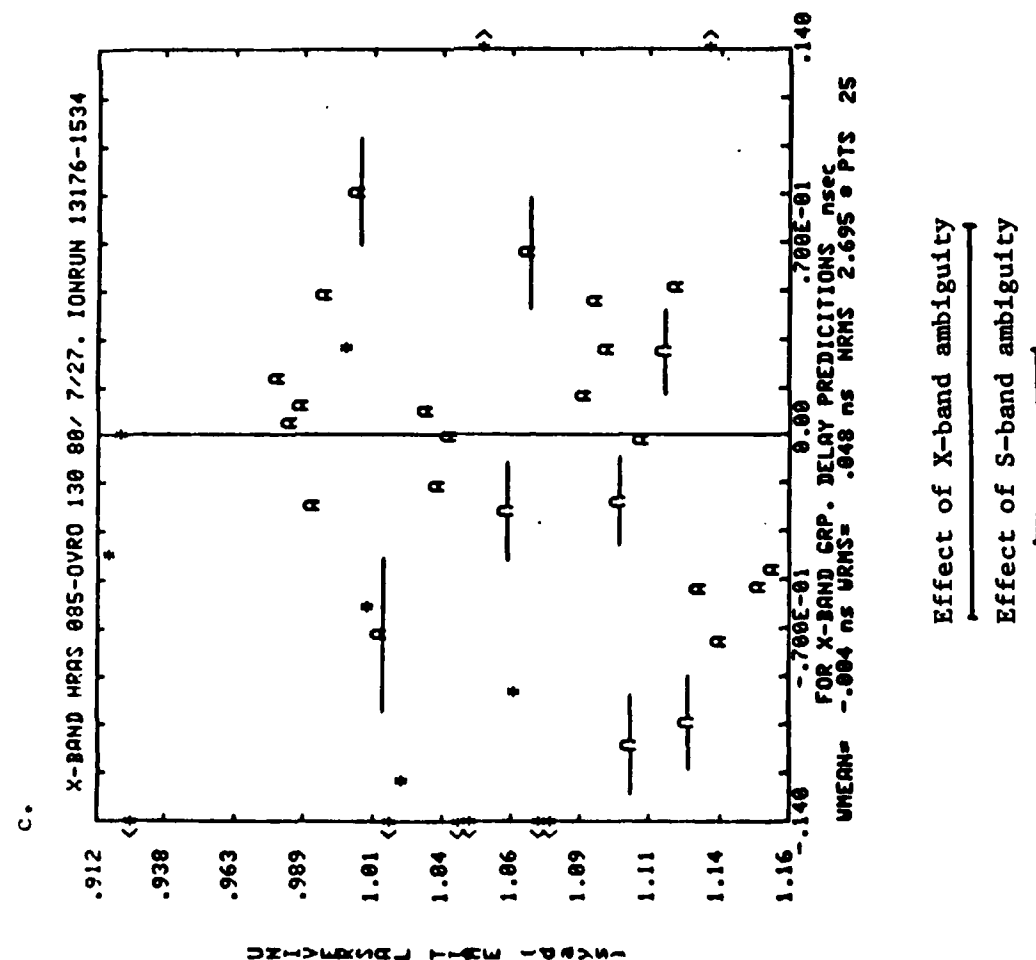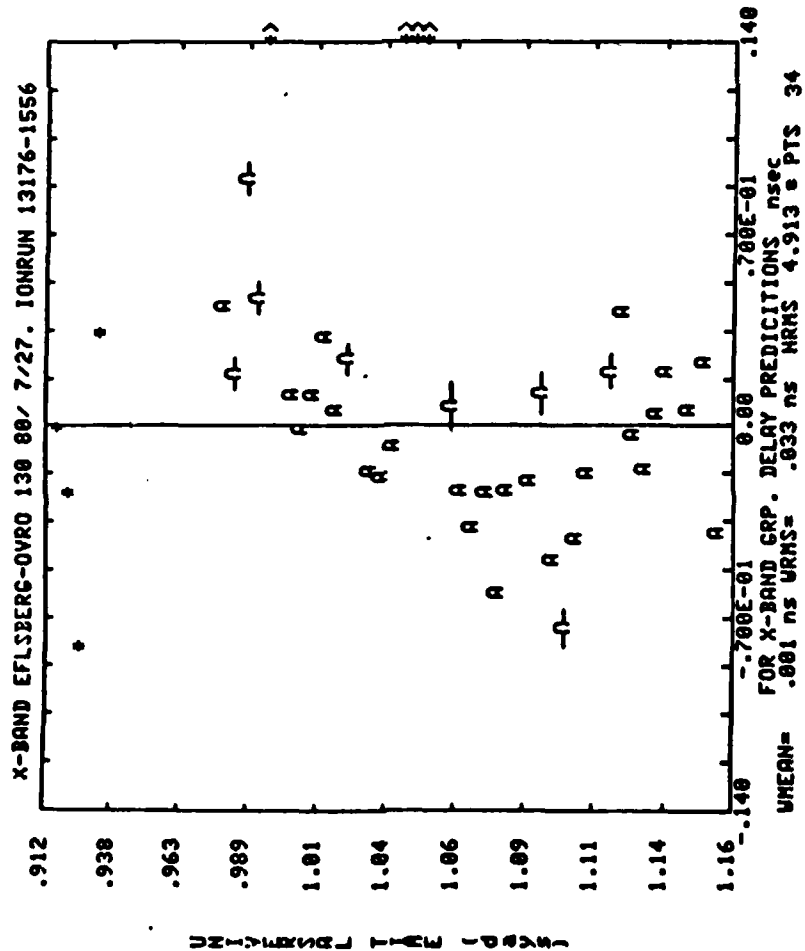
Figure 4.3.5 Continued

Figure 4.3.5 Continued

e.

f.
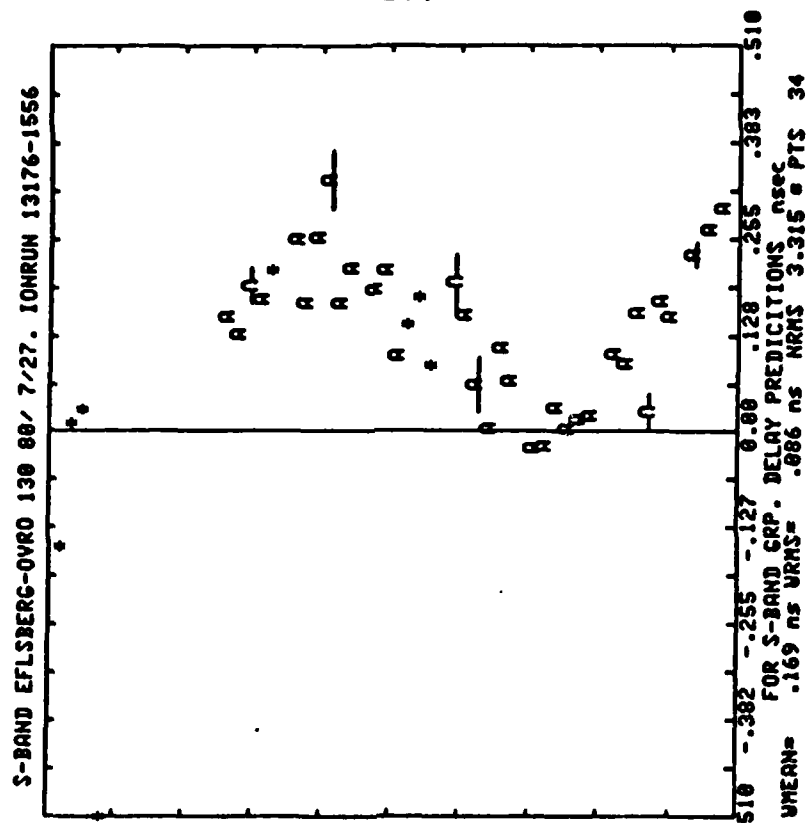
phase delays used to predict the group delays, satisfy a closure condition, i.e., they sum to approximately zero (and much less than an ambiguity) around closed triplets of base-lines when the phase delays are referred to the same epoch (see Section 3.3). Hence, if the phase-delay ambiguities are changed on one baseline, they must be changed on at least one other baseline to ensure that the closure conditions remain satisfied.

We now consider the differences for several of the base-lines shown in Figure 4.3.5. We firstly consider the apparent S-band group delay errors for the Haystack-Onsala interfer-ometer. We see in this case that for the last 3 hours of the experiment the group delays seem to be in error by 0.510 nsec which suggests an S-band phase connection error. However, when we plot the group and phase delay closure errors for a triplet of baselines involving Haystack and Onsala (Figure 4.3.6), we see immediately that the group delays have closure errors which are remarkably similar to the Haystack-Onsala apparent group delay errors. The phase delays, however, show very small closure errors. Hence, even before we attempted to predict group delays from phase delays, we should have expected large differences based on the group-delay closure errors.

One other case from Figure 4.3.5 which we will examine is the apparent group-delay errors for the Ft. Davis-Owens Valley interferometer. We notice that in this case the apparent group delay errors seem to have a bimodal distribution,

Figure 4.3.6 S-band group and phase delay closure errors for the Haystack-Effelsberg-Onsala triplet of baselines

Phase delay closure error

closure error (nsec)

Group delay closure error

closure error (nsec)

Universal time (days)
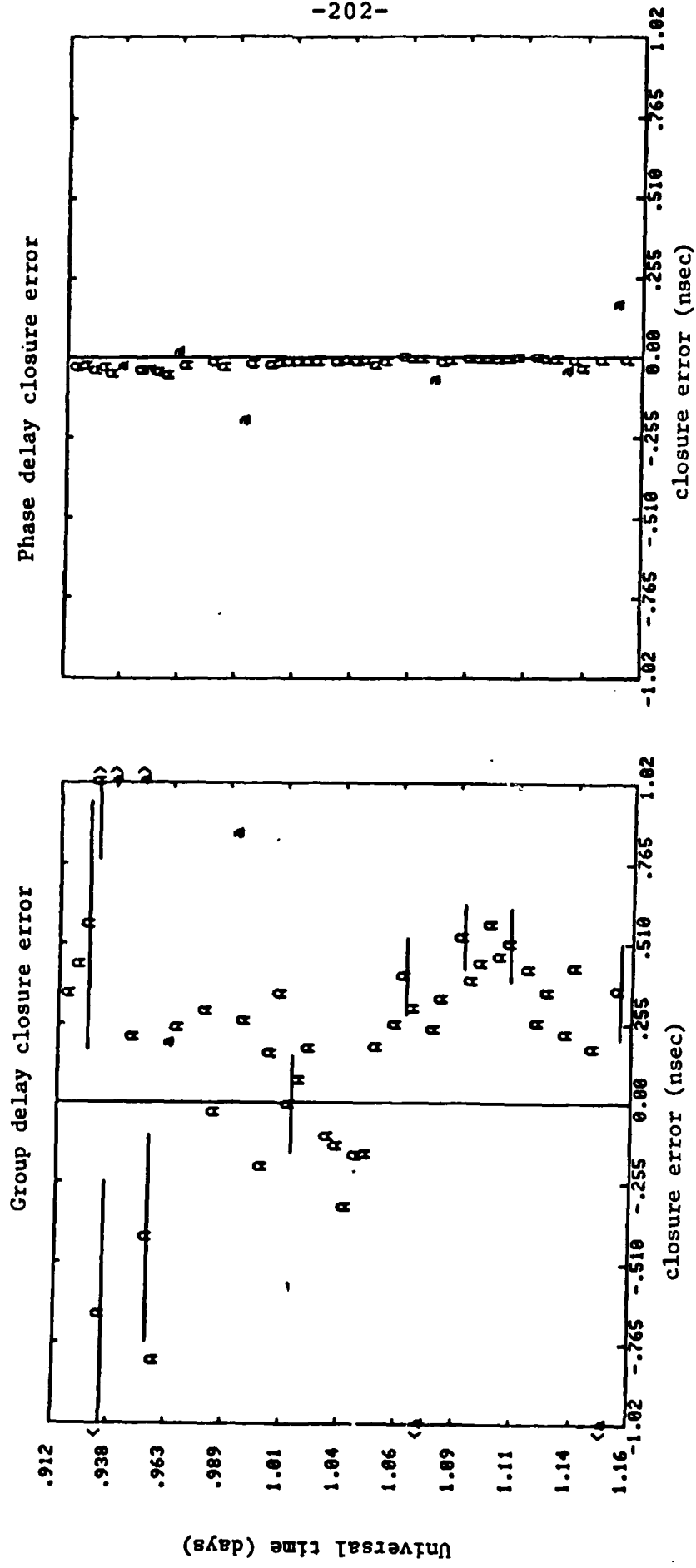
especially at S-band. This type of pattern would be expected if there were remaining phase delay ambiguities. However, the phase delay postfit residuals at X- and S-band for this baseline do not show any apparent phase connection errors (Figure 4.3.7). (In this case, the group delay closure errors have too much scatter to allow the detection of any bimodal errors.)

We have compiled the statistics of the apparent group delay errors (in a manner similar to that used in Chapter 3), and these results are shown in Figure 4.3.8. We have plotted as a function of the expected scatter of the group delay errors, the normalized RMS scatter. We have shown these plots for both the July 1980 and June 1981 experiments. The normalized RMS plot shows trends which are similar to those for the residual phases (see Figures 3.1.12 and 3.1.13), i.e., there appears to be an overall scaling factor of approximately 1.1 at X-band and 1.2 at S-band, and performance limits of 40 psec and 14 psec at X-band for the July 1980 and June 1981 experiments, respectively, and 150 psec at S-band for both July 1980 and June 1981.

There appears to be a large improvement in the X-band system performance between these two experiments which is probably due to the smaller spurious signals during the June experiment. (Remember that at Haystack even after correcting for the presence of the spurious signal, there were residual trends in the contribution of the calibration phases to the group delays. See Section 3.2 for discussion.) The S-band

Figure 4.3.7 Ft. Davis-Owens Valley interferometer phase delay postfit residual at X- and S-band

a.   X-band post-fit residuals

b.   S-band post-fit residuals
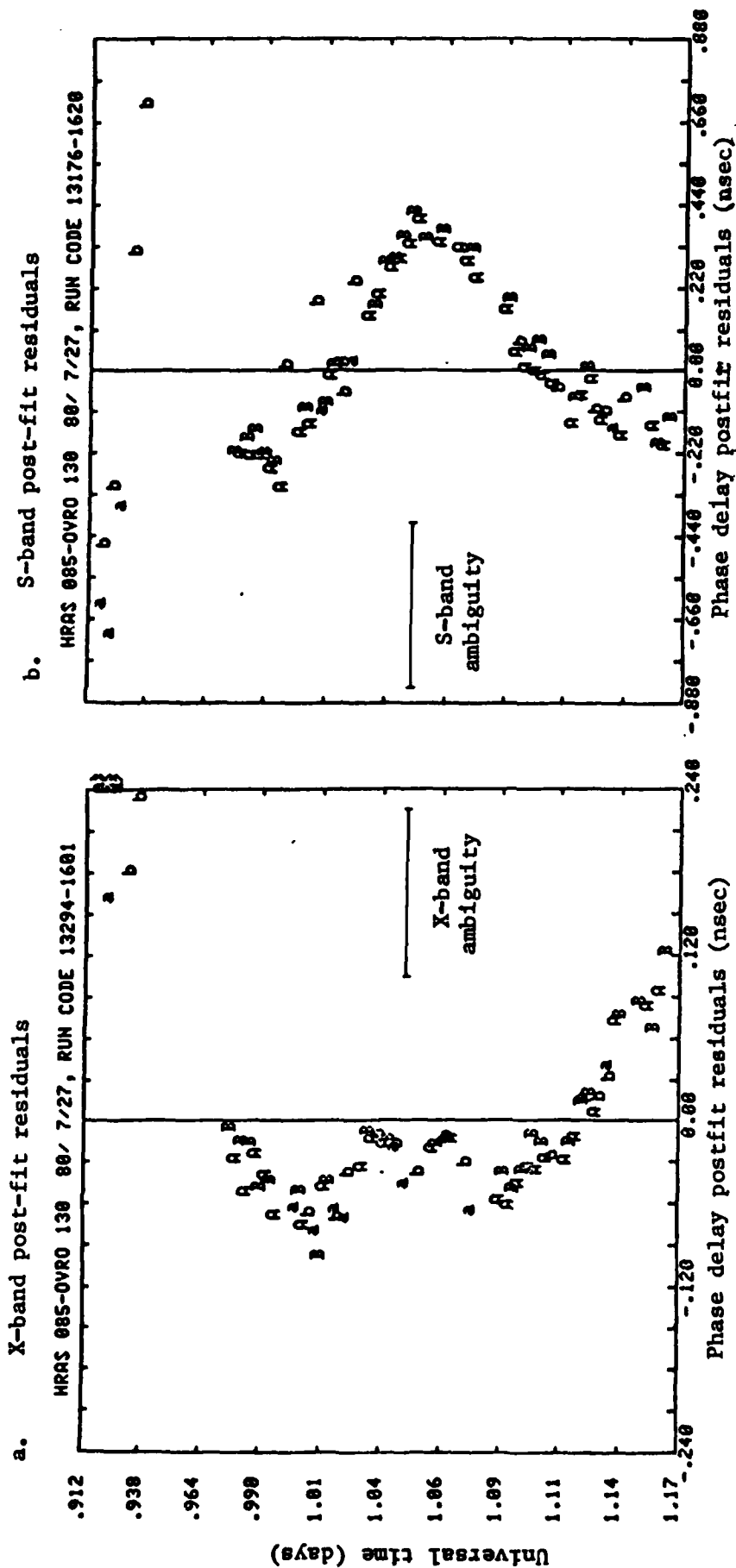
* The letters A and B denote observations of 3C 345 and NRAO 512, respectively. The only parameters estimated were the coefficents of a linear function which was used to represent the relative offset and rate for the clocks at these sites

Figure 4.3.8 Normalized RMS scatter of the implied group delay errors
as a function of the expected RMS scatter

July 1980 X-band

Figure 4.3.8   Continued

b.

July 1980 S-band

curve for $\sigma_t^2 = (1.2\sigma_o)^2 + (0.15)^2$ nsec

Legend
● 3C 345
■ NRAO 512

Expected RMS differences (nsec)

Normalized RMS scatter

Figure 4.3.8   Continued

c.

June 1981 X-band

Legend
● 3C 345
■ NRAO 512
✱ 0234+285
● 0235+164

curve for $\sigma_t^2 = (1.1\sigma_o)^2 + (0.015)^2$ nsec

(y-axis: Normalized RMS scatter, x-axis: Expected RMS differences (nsec))

Figure 4.3.8   Continued

d.

June 1981 S-band

curve for $\sigma_t^2=(1.2\sigma_o)^2+(0.15)^2$ nsec

Legend
● 3C 345
■ NRAO 512
✸ 0234+285
● 0235+164
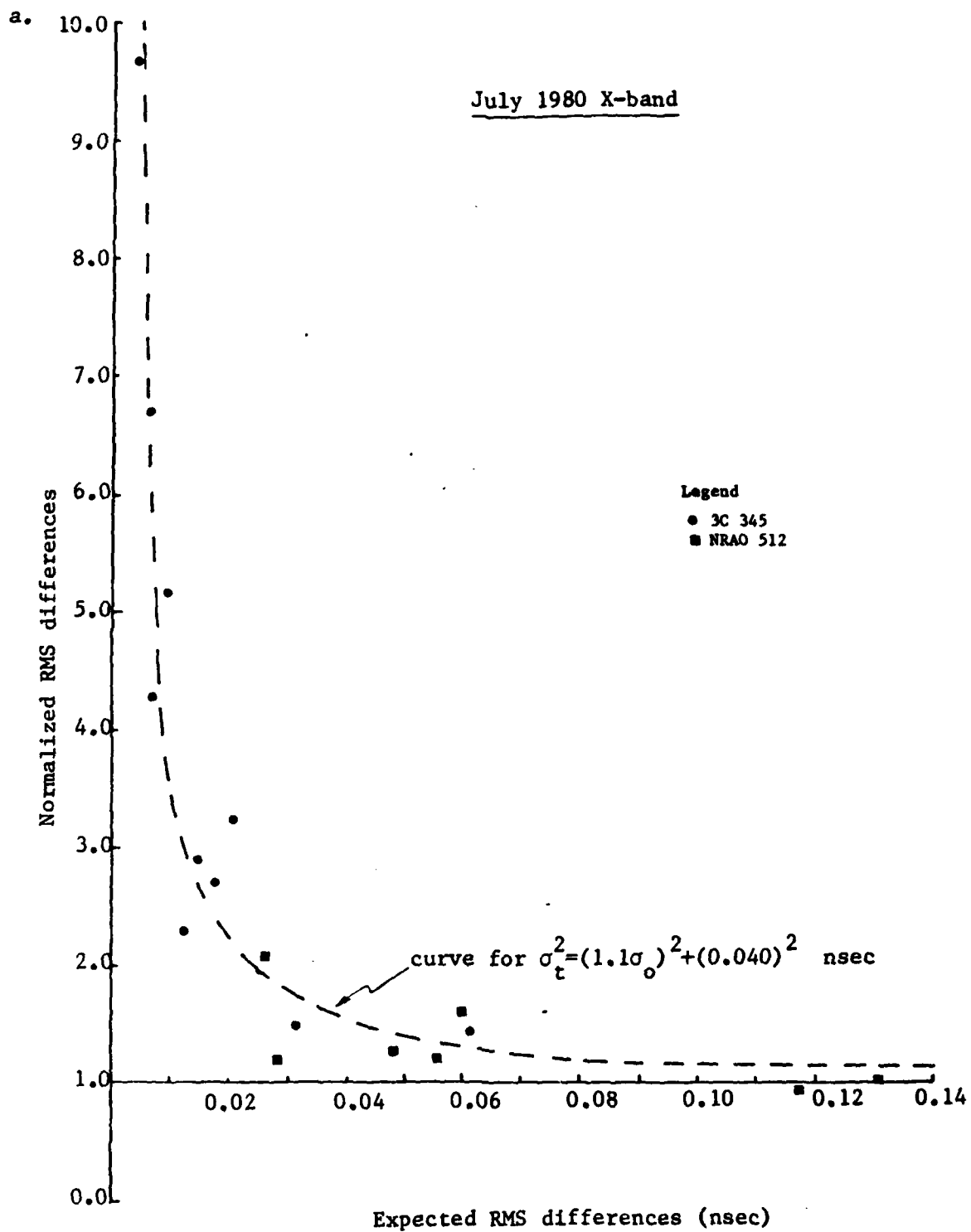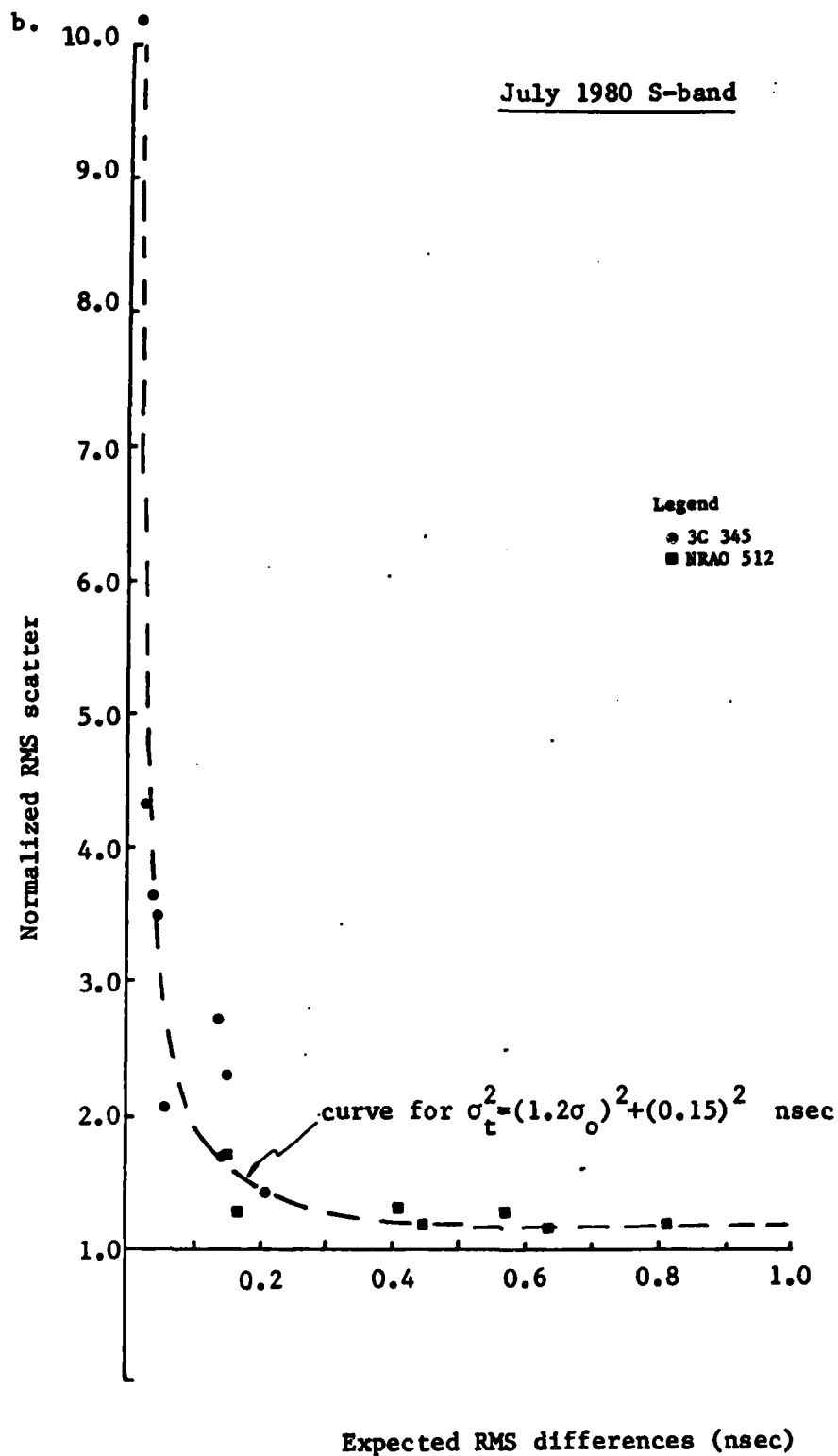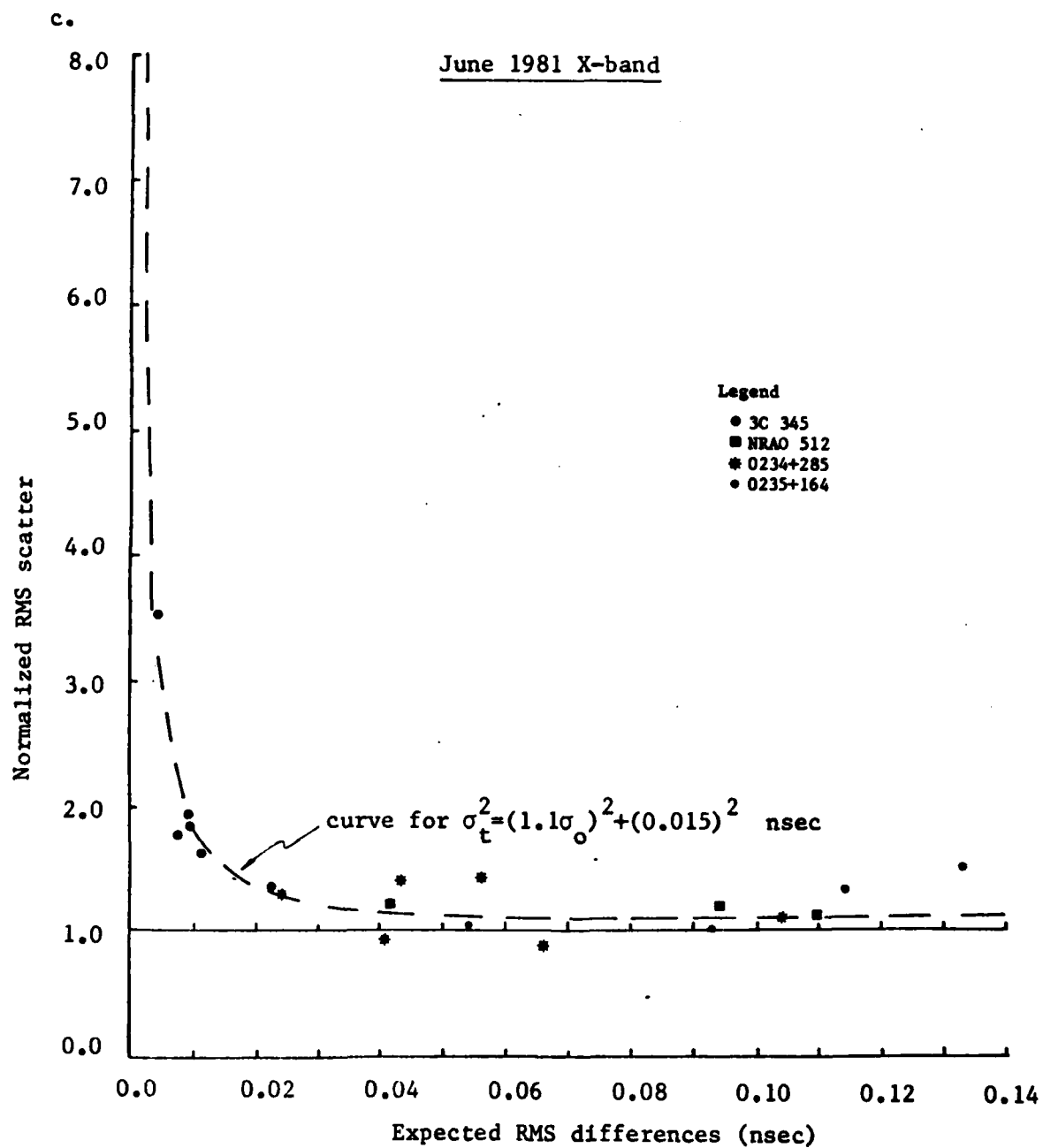
system performance does not seem to have improved between the two experiments.

The performance threshold at S-band implies that we cannot predict phase delay ambiguities from group delays even with infinite signal-to-noise ratio observations at S-band. (See Section 4.2; the upper bound in the S-band uncertainty for phase delay predictions was 0.09 nsec.)

We have yet to discuss the possible origins of the systematic trends. We leave this discussion until we have considered the apparent group delays from a geodetic experiment.

4.4   Results for geodetic experiments: Effelsberg-Onsala

We saw in Section 4.3, that when we evaluated the apparent group-delay errors from the phase-connected experiments, the performance of the S-band system was not good enough to allow the prediction of the phase-delay ambiguities from the group delays. However, in view of the systematic trends seen in the apparent group-delay errors, we have attempted to predict phase-delay ambiguities from the group delays for the Effelsberg-Onsala interferometer, in order to gain a better understanding of these trends.

We chose to analyze the July 26th 1980, geodetic experiment because this experiment had the largest number of observations of any 24hr experiment involving these two antennas. We show in Figure 4.4.1 the apparent group delay errors at X- and S-band. (We have plotted each figure twice,

Figure 4.4.1 Implied group delay errors for the Effelsberg to Onsala baseline for the geodetic
schedule on July 26th, 1980.*

a. X-band implied errors

b. One standard deviation error bars shown

Implied group delay error (nsec)

Effect of X-band ambiguity

Effect of S-band ambiguity
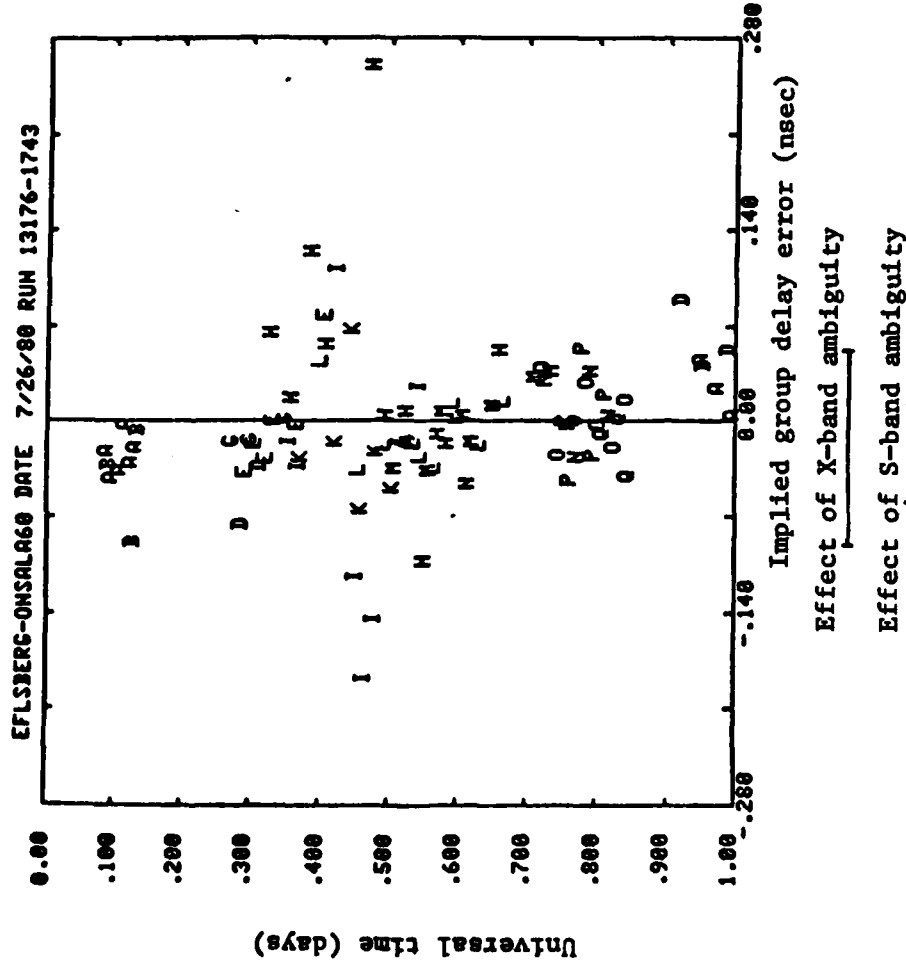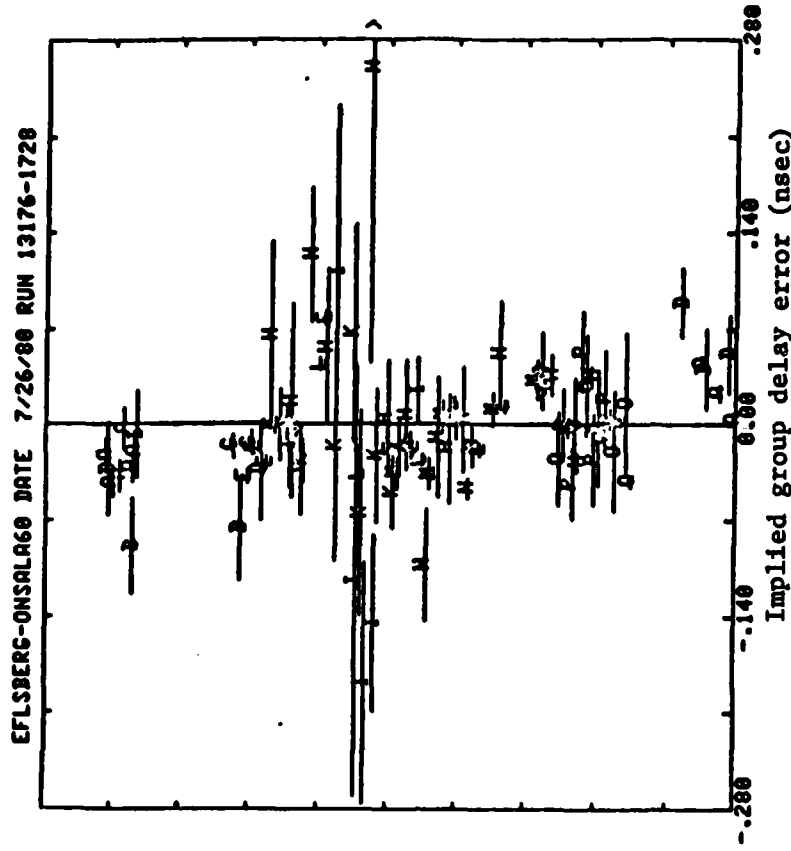
*Each letter denotes observations of a particular source (A=1641+399, B=1633+381, C=1749+096,
D=1642+690, E=2200+420, F=2134+002, G=2251+158, H=0224+671, I=0234+285, J=0106+013, K=0235+164,
L=0355+508, M=0552+398, N=0923+392, O=0953+254, P=1219+285, Q=1226+023)

Figure 4.4.1 Continued

c. S-band implied errors

d. One standard deviation error bars shown

EFLSBERG-ONSALA6Ø DATE 7/26/80 RUN 13176-1744

EFLSBERG-ONSALA6Ø DATE 7/26/80 RUN 13176-1736

Implied group delay errors (nsec)

Effect of X-band ambiguity

Effect of S-band ambiguity

Universal time (days)
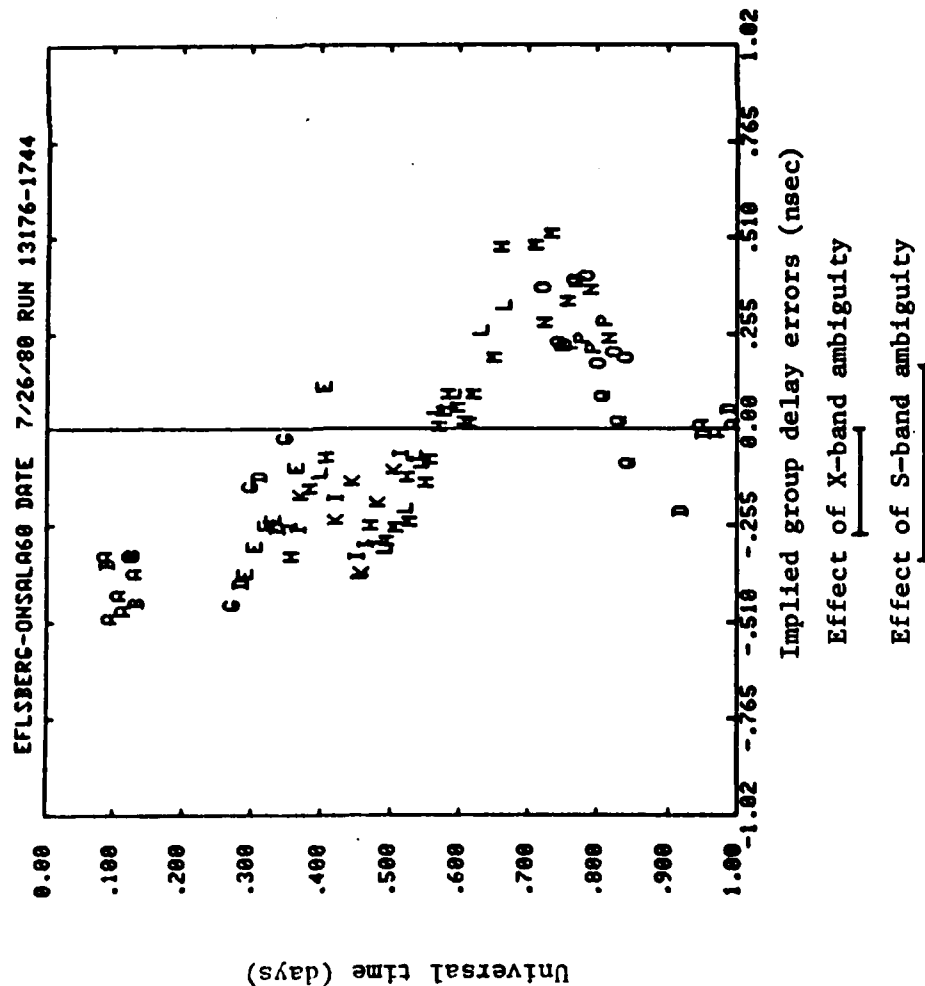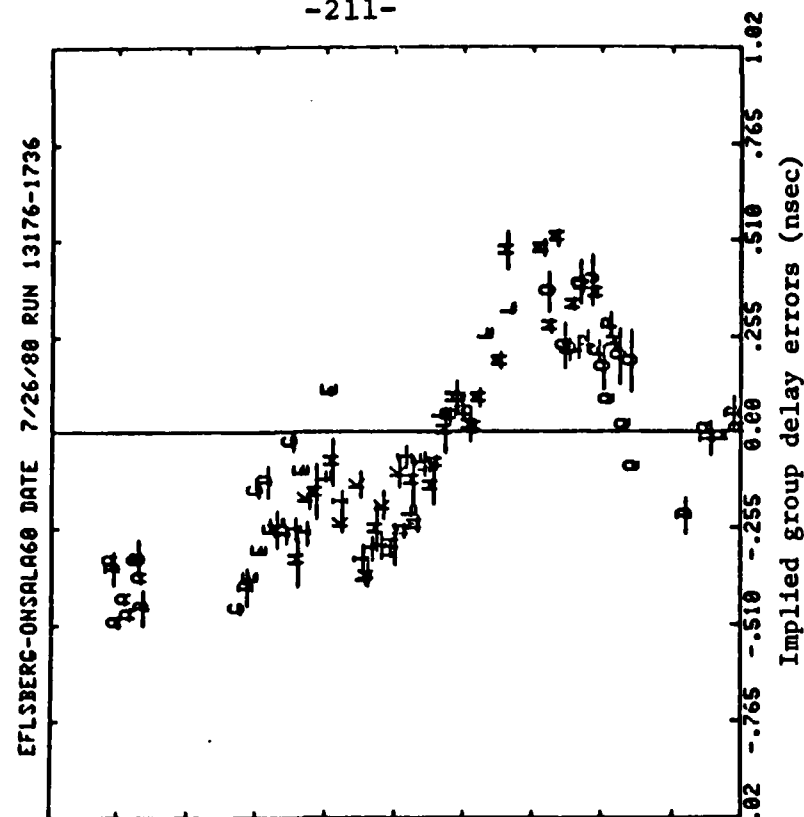
once with one-standard-deviation error bars, the second time
without error bars.) We see in these figures the same types
of systematic trends which were observed in the phase
connected experiments. Again, the S-band results show much
greater trends than the X-band results. The normalized RMS
scatter of the X-band results is 3.4, which indicates that
their scatter is not in accord with their statistics.

We have shown these results because they give us some
insights into the origin of the systematic trends. There are
several important properties of this origin which can now be
deduced from the geodetic results. Firstly, we notice that a
major part of the apparent group delay errors at S-band is
independent of the radio source being observed. We may there-
fore conclude that the differences are most probably not
caused by the brightness distributions of the radio sources
(see Section 3.3). Also, since the radio telescopes look in
many different parts of the sky in rapid succession, the major
portion of the trends is probably not due to antenna
deformation. In Figure 4.4.2, we show the elevation angles at
Effelsberg and Onsala for these observations. From this
figure we see that observations are taken at many different
elevation angles in a short period of time. (In each of these
cases some portion of the error is may be due to these causes
(i.e., source structure and antenna deformations), because the
scatter of errors around smooth curves is greater than would
be expected from the variances of the observations.)

The apparent group delay errors seem to depend mainly on

Figure 4.4.2 Elevation angles at Effelsberg and Onsala for the observations shown in Figure 4.4.1
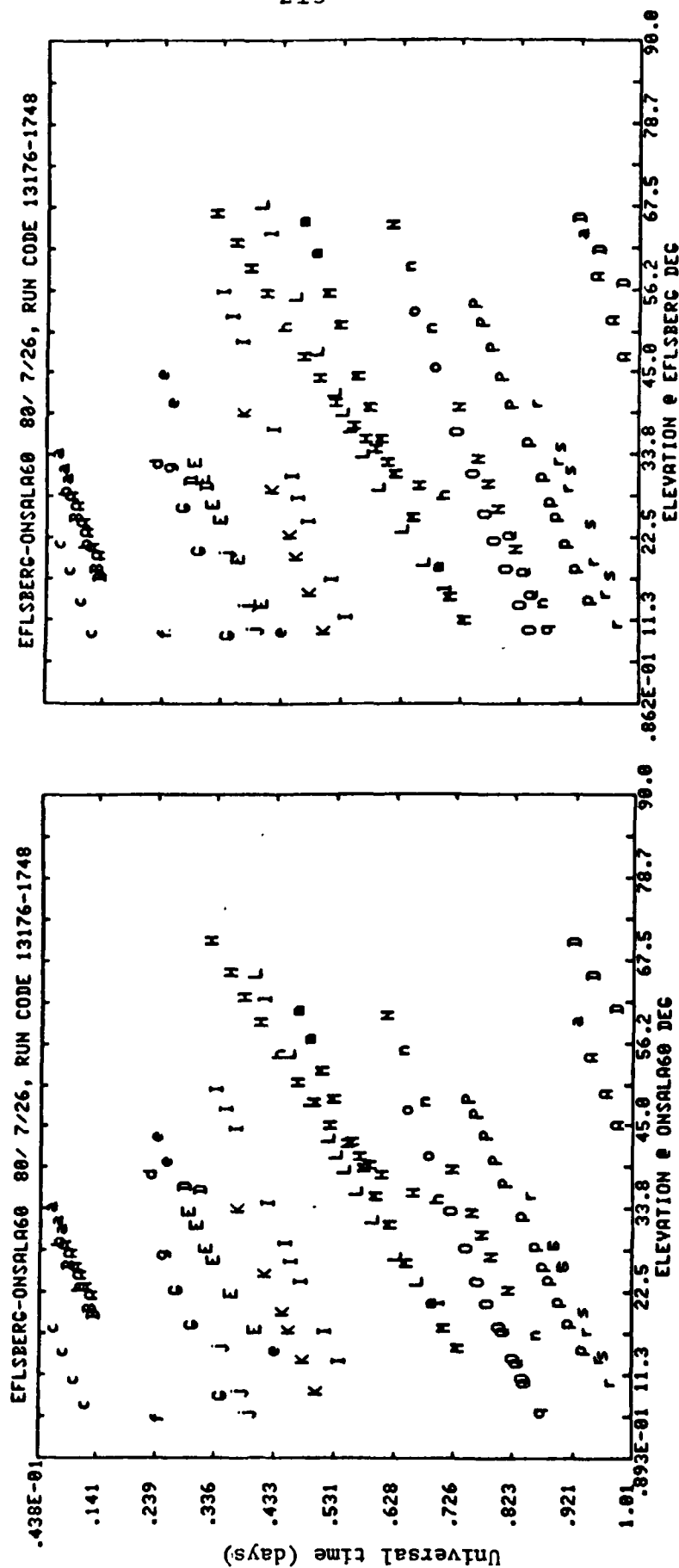
time, and hence it is likely that they arise from the Mark III system itself. What could be the origin of the errors? We saw in Section 3.1 that the Mark III dispersion depends on time and we found lower bounds on the variations of 8 psec/day at X-band and 36 psec/day at S-band. But we stressed in that section that these are lower bounds because we cannot estimate the full delay variations from the residual phases. At this time it seems that variations of the Mark III dispersion with time are the most likely cause of the systematic variations of the apparent group delays.

It is important at this time that we put the systematic variations of the S-band group delays into perspective. In a geodetic data analysis, the S-band group delays are used only to calculate the ionospheric correction. Consequently, the contribution of their errors to the non-dispersive delays are 12.5 times smaller than their errors and, hence, the threshold performance limit of 150 psec at S-band, results in the threshold limit of ≈12 psec for non-dispersive delays. When we compare this limit with the threshold limits of the X-band group delays (40 psec in July 1980 and 14 psec in June 1981), we see that the X-band group delay errors are of equal, or more importance, than the S-band errors.

5. Summary of results

We have now discussed the theory and the operational performance of the Mark III VLBI system. While the performance of the system is not so good as expected (see Chapters 3 and 4), we have not detected any major errors in the measurements made with the system. We now discuss some of the geodetic results which have been obtained with this system and the Mark I VLBI system.

This chapter will consist of three papers -- two of which have already been published and one of which is currently being prepared for publication. For those readers unfamilar with the locations and names of the radio telescopes which have been used in geodetic VLBI experiments, this information is given in Table 5.1. The telescope locations are also shown in Figure 5.1. The IAU (International Astronomical Union) convention and the common names of the radio sources discussed here and elsewhere in the thesis are given in Table 5.2.

The first paper we present, entitled "Geodesy by radio interferometry: Intercontinental distance determinations with subdecimeter precision", was published in 1981 and discusses the baseline length measurements from radio telescopes in North America to a radio telescope in Onsala, Sweden, which were made with the Mark I VLBI system. When this paper was published we thought that these length measurements were of subdecimeter precision. Soon after dual frequency band observations to Onsala were analysized, it became clear that

Table 5.1   Radio telescope descriptions and names

Observatory	Name[‡]	Operated by	Location	Antenna Diameter
North America				
Haystack Observatory	HAYSTACK	Northeast radio Observatory Cor- poration (NEROC)	Westford, Ma.	37 m
Westford Observatory	WESTFORD	NEROC	Westford, Ma.	26 m
Harvard Radio Astron- omy Station	HRAS 085	Harvard University	Ft. Davis, Tx.	26 m
Owens Valley Radio Obser- vatory	OVRO 130	California Insti- tute of Technology	Big Pine, Ca.	40 m
National Radio Astron- omy Obser- vatory	NRAO 140	National Radio Astronomy Obser- vatory (NRAO)	Green Bank, Wv.	43 m
Eupope				
Onsala Space Observatory	ONSALA60 (X-band)	Chalmers Institute of Technology	Onsala, Sweden	20 m
	ONSALA85 (S-band)			26 m
Effelsberg Observatory	EFLSBERG	Max Plank Institut fur Radio- astronomie	Bonn, Federal Republic of Germany	100 m
Chilbolton Observatory	CHLBLTON	Appleton Laboratories	Chilbolton, England	26 m

[‡] These names will be used on the computer generated plots.

Figure 5.1   Locations of radio telescopes
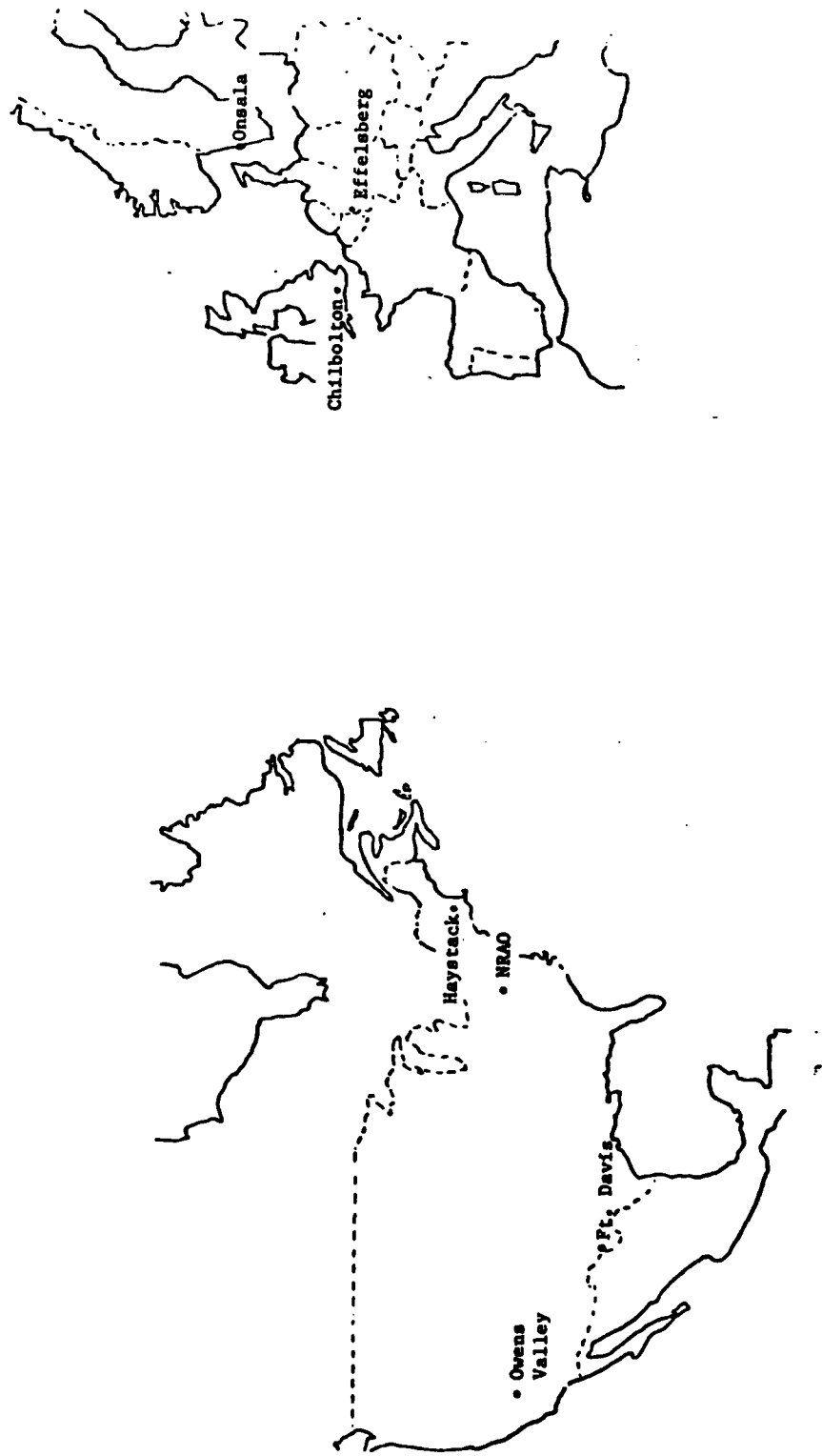
Table 5.2   IAU convention[‡] and common names for VLBI radio sources

Common name	IAU name
4C 67.05	0224+671
NRAO 150	0355+508
OJ 287	0851+202
4C 39.25	0923+392
OK 290	0953+254
3C 273B	1226+023
OQ 208	1404+286
NRAO 512	1638+398
3C 345	1641+399
VR 422201	2200+420
3C 454.3	2251+158

[‡]The IAU (International Astronomical Union) names are defined in the form of rrhh±dd.d, where rrhh are the hours and minutes of the right ascension of the source, and dd.d is the declination of the source in decimal degress.

we had grossly underestimated the effects of the ionosphere delay on the single frequency band (X-band) Mark I observations. Although, at the time we also thought that some of the differences (between the Mark I and the Mark III baseline length estimates) could have been due to the change in the equipment or to the ionospheric delay correction computed from the two antennas at Onsala (see Section 4.2).

We learned a valuable lesson on the effects of un-modeled systematic errors from this experience. This lesson prompted the detailed analysis of the baseline length estimates which are discussed in the second paper in this chapter -- "Geodesy by radio interferometry: Precision and accuracy of inter-continental distance measurements."

The third paper we present discusses the determination of the tidal parameters of the earth using VLBI. The paper is entitled "Determination of the tidal parameters of the earth."

GEODESY BY RADIO INTERFEROMETRY:
INTERCONTINENTAL DISTANCE DETERMINATIONS WITH SUBDECIMETER PRECISION

T. A. Herring,[1] B. E. Corey,[1] C. C. Counselman III,[1] I. I. Shapiro,[1,7]
B. O. Rönnäng,[2] O. E. H. Rydbeck,[2] T. A. Clark,[3] R. J. Coates,[3]
C. Ma,[3] J. W. Ryan,[3] N. R. Vandenberg,[3] H. F. Hinteregger,[4]
C. A. Knight,[4] A. E. E. Rogers,[4] A. R. Whitney,[4]
D. S. Robertson,[5] and B. R. Schupler[6]

Abstract. Analysis of very-long-baseline interferometer (VLBI) observations yielded estimates of the distances between three radio telescopes in the United States and one in Sweden, with formal standard errors of a few centimeters: Westford, Massachusetts-Onsala, Sweden: 5,599,714.66±0.03m; Green Bank, West Virginia-Onsala, Sweden: 6,319,317.75±0.03m; and Owens Valley, California-Onsala, Sweden: 7,914,131.19±0.04m, where the earth-fixed reference points are defined in each case with respect to the axes of the telescopes. The actual standard errors are difficult to estimate reliably but are probably not greater than twice the formal errors.

1. Introduction

Very long baseline interferometry (VLBI) has been used for the past decade to determine distances between radio telescopes and positions of radio sources. Improvement in the precision of baseline length determination in North America over this period has been substantial -- from a precision of about 1 m [Hinteregger et al., 1972] to a precision of under five centimeters [Robertson et al., 1979] for baselines from about 1000 km to just under 4000 km in length. Near the middle of the decade, distances between North America and Europe were determined with a precision of about half a meter [Robertson, 1975; see also Cannon et al., 1979]. In this paper we describe more recent

---

[1]Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

[2]Onsala Space Observatory, S-439 00 Onsala, Sweden.

[3]Goddard Space Flight Center, Greenbelt, Maryland 20771.

[4]Northeast Radio Observatory Corporation, Haystack Observatory, Westford, Massachusetts 01886.

[5]U.S. National Geodetic Survey, Rockville, Maryland 20852.

[6]Computer Science Corporation, Silver Spring, Maryland 20910.

[7]Also at Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

VLBI determinations, with subdecimeter precision, of these intercontinental distances.

2. Data Analysis

The data analyzed consisted of interferometric group delays [Shapiro, 1976] which were obtained from three sessions of observations involving up to four antennas, as described in Table 1. All observations were made with the Mark I VLBI system with a center radio frequency of about 8 GHz. A multichannel bandwidth synthesis technique [Whitney et al., 1976] was used. Each recorded channel had the Mark I bandwidth of 360 kHz, whereas the synthesized bandwidth was ~100 MHz in the first session of observations and ~300 MHz in the last two sessions. A hydrogen maser frequency standard of modern, post-1970, design was used at each telescope for each session.

The Cartesian coordinate system used in the analysis was geocentric and earth-fixed with the Z axis parallel to the mean pole of rotation of 1900-1905, as defined by the International Latitude Service and maintained by the Bureau International de l'Heure (BIH). The X axis was defined to be perpendicular to the Z axis and in the direction of the Greenwich meridian. The Y axis completed the right-handed triad. Operationally, the origin of this system was defined by the coordinates for the intersection of the azimuth and elevation axes of the Haystack radio telescope, obtained from a combination of spacecraft-tracking and VLBI observations made at various sites. The orientation was defined by the BIH values for pole position and UT1 (1968 system), with the addition of fortnightly and monthly terms of small amplitude [Woolard, 1959], and by the models for diurnal polar motion [McClure, 1973] and earth tides (based on ephemeris positions of the moon and sun, on the Love numbers h = 0.61 and ℓ = 0.09, and on the assumption of no dissipation). The small effects of antenna deformations [Carter et al., 1980], ocean loading, plate tectonics, and other errors in the model were ignored in view of the limited precision of the measurements being analyzed.

The orientation of this earth-fixed coordinate system in inertial space (with respect to the mean equinox and equator of 1950.0), was defined by the pole position, by UT1 and by the standard formulas for sidereal time, nutation, and precession, with certain small corrections: Woolard's [1953] theory as modified by Melchior [1971] was used for nutation; and 5,027."878 per tropical century at 1950.0 was taken for the precession constant. The origin of right ascension was defined by the value 12 hr 26 min

TABLE 1. Summary of VLbi Observations

Observation Session	Dates	Duration, hours	Antennas[*]	Number of Sources[†]	Declination Range of Sources, deg	Number of Group-Delay Observations[**]
1	Sept. 21-25, 1977	88[††]	H,N,O	10(6)	-5.5 to 50.8	529 (151)
2	Feb. 24-26, 1978	50	H,O,V	8(4)	0.5 to 50.8	291 (77)
3	May 17-19, 1978	45	H,O,N,V	10(10)	-5.5 to 50.8	830 (258)

[*] H ≡ 37-m diameter telescope of the Haystack Observatory, Westford, Massachusetts; N ≡ 43-m diameter telescope of the National Radio Astronomy Observatory (NRAO), Green Bank, West Virginia; O ≡ 20-m diameter telescope of the Onsala Space Observatory, Onsala, Sweden; and V ≡ 40-m diameter telescope of the Owens Valley Radio Observatory, Big Pine, California.

[†] The sources observed are listed in the caption to Figure 1. The numbers of sources observed at Onsala are given in parentheses.

[**] The numbers of observations involving Onsala are given in parentheses.

[††] Onsala participated for the last 62 hours.

33.246 sec for the radio source 3C 273B (elliptic aberration removed).

Using this framework and a theoretical model for the group delays [Robertson, 1975; Ma, 1978], we estimated by weighted least squares the baseline vectors, the source positions, and the other relevant parameters. Included in the last category are parameters that allow for the estimation of changes in pole position and variations in UT1 when data from two or more sessions of observations are combined [Shapiro et al., 1974, 1976; Robertson et al., 1979]. In addition to these purely geometric parameters there are parameters in the model that represent the effects on the group delays of the propagation medium and of the behavior of the clocks at the various radio telescopes. The representation of the propagation medium includes the zenith tropospheric delay as a parameter. Since the signal delay through the troposphere varies with time and with location, a number of such parameters are utilized with this model. The effects of the ionosphere, always under ∿1 ns in the zenith direction, but not separately modeled, are largely absorbed by these parameters because of the similarity in the signature in the group delays of the ionosphere and the troposphere. The representation of the behavior of the clock at one site relative to that at another consists of a polynomial in time of low order with the coefficients as parameters, the first two denoting the epoch and rate offsets. Since the clocks drift unpredictably with time, a number of such polynomials are usually employed.

The choice of the number of parameters to represent the combined effects of the troposphere and the ionosphere and the behavior of the clocks as well as the choice of the time interval of applicability of each parameter are to a certain extent subjective. We therefore investigated the effects on the results of a variety of such choices. As extremes for the representation of the troposphere and the ionosphere, we considered for each telescope the use of one parameter for the zenith delay for each session of observations and one such parameter for each 12-hour period of observations. For the clock behavior we chose the clock at

Haystack as the reference and considered for each other telescope for each session the use of one parameter each for the relative offsets in epoch, rate, and (in several instances) rate of change of rate. The intervals of applicability of such 'clock polynomials' ranged from 7.5 to 24 hours; only for this latter case was a parameter included for the rate of change of rate. The variations in the estimates of baseline lengths resulting from these different parameterizations were up to 1.8 and 1.2 times the corresponding standard deviation obtained for the adopted parameterization for the data from each of the first two sessions and from the third session, respectively. In view of this stability, we deem it unlikely that either the propagation medium or the behavior of the clocks has introduced errors in baseline length much in excess of the formal standard deviations. The adopted choice of epochs and intervals of applicability for these parameters (see section 4) was based primarily on examination of the postfit residuals, especially those from observations at low elevation angles for the parameters representing the behavior of the propagation medium.

The inverse of the weight for each group delay was obtained from the sum of the variance found from a signal-to-noise analysis [Whitney, 1974] and an ad hoc term included to account for error sources that were not a function of signal strength. The inclusion of this term was motivated by the desire to (1) allow for the effects of systematic errors, and, relatedly, (2) weight the observations of the different sources more evenly, since for some strong sources the signal-to-noise analysis yielded uncertainties that were well below the suspected contributions of other errors. The magnitude of this ad hoc term was assumed constant for the data for a given baseline for a given session of observations. Its value was obtained from the constraint that chi-square per degree of freedom be unity, and its square root ranged from 0.10 to 0.25 ns. Omitting this ad hoc term from the variances changed the estimates of baseline length by up to 1.2 times the standard deviations obtained with this term included.

3. Results

Table 2 gives our estimates of the distances between the radio telescope in Sweden and the three in the United States. The reference point for each telescope is the intersection of the azimuth and elevation axes, except for the NRAO telescope where the reference is the point on the polar axis closest to the (nonintersecting) equatorial axis [Hinteregger et al., 1972].

Typical postfit residuals from the simultaneous analysis of the data from all sessions are shown in Figure 1. The root mean square of these residuals ranged from 0.13 ns for the Haystack-NRAO data from the last session to 0.49 ns for the Onsala-NRAO data from the first session, the ratio being approximately as expected in view of (1) the threefold smaller synthesized bandwidth used in the first session of observations, (2) the generally larger correlated flux densities obtained from observations involving the shorter baselines, and (3) the slightly higher sensitivity of the Haystack-NRAO system compared to the Onsala-NRAO system.

The repeatability of the estimates of the baseline lengths shown in Table 2 indicates that the formal standard errors may not be much smaller than the actual errors. However, one fact detracts from the importance of this indication: The uncertainties of the estimates of baseline lengths from the third session of observations were much smaller than those from the second which, in turn, were smaller than those from the first. The disparity between the uncertainties from the first and second sessions *was due primarily to the contrast in synthesized bandwidths.* The further discrepancy between the uncertainties from the first two sessions and those from the third was due mainly to geometry; the schedule of observations for the first two sessions was controlled by requirements of other experiments involving only the U.S. antennas and therefore did not properly provide for Onsala's participation.

The estimates for the propagation delay through the atmosphere and ionosphere in the zenith direction ranged from 6.7 ns (NRAO) to 8.2 ns (Onsala); the estimates for the clock parameters were also each in approximate accord with expectations, ranging up to 23 μs for epoch offsets and from $\sim 3 \times 10^{-14}$ to $\sim 2 \times 10^{-12}$ for rate offsets.

The estimates for the source positions, the distances between the radio telescopes in the United States, and the pole position and UT1 variations will be given elsewhere and discussed there along with a much larger collection of VLBI observations that involved only the telescopes in the United States.

4. Conclusions

Combination of data from three sessions of VLBI observations involving sites in Sweden and the United States yielded results of subdecimeter precision in the determination of the intercontinental distances. Since sufficiently accurate and well-distributed observations to determine baselines to Sweden with subdecimeter precision were available for only one session, the data are not useful for the establishment of

---

TABLE 2. Baseline Length Estimates

Observation Session(s)*	Haystack – Onsala Length, m	Formal Standard Error, m	Difference,† m	NRAO – Onsala Length, m	Formal Standard Error, m	Difference,† m	Owens Valley – Onsala Length, m	Formal Standard Error, m	Difference,† m
1	5,599,714.948	0.301	0.288	6,319,318.156	0.354	0.404	7,914,131.372	0.169	0.180
2	.4.516	0.120	-0.144	—	—	—	.1.173	0.043	-0.019
3	.4.686	0.028	0.026	7.776	0.036	0.024	—	—	—
1 + 2 + 3**	5,599,714.660	0.027	—	6,319,317.752	0.034	—	7,914,131.192	0.040	—

* See Table 1 for definitions and text for discussion. The number of parameters estimated in each solution, in order of listing, was 47, 41, 56, and 102. The speed of light used to convert light seconds to meters was 299,792.458 m s^{-1}.
† Differences from adopted solution.
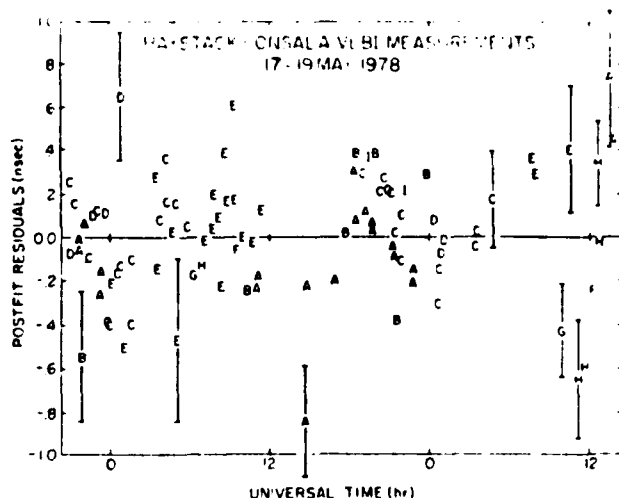** Adopted solution.

Fig. 1. Postfit residuals from the adopted solution (see Table 2) for the Haystack-Onsala group-delay measurements from session 3 (see Table 1). The root mean square of these residuals is 0.25 ns. The vertical bars represent ±1 standard deviation. Each letter in the figure denotes a different radio source: A ≡ NRAO 150; B ≡ 3C 84; C ≡ 4C 39.25; D ≡ 3C 273B; E ≡ 3C 345; F ≡ VRO 42.22.01; G ≡ 2134+00; H ≡ 3C 454.3; and I ≡ OJ 287. One other source, observed on some baselines in session 3, was 3C 279. See Clark et al. [1976] for approximate values of the coordinates for each source.

subdecimeter repeatability. Neither are the earlier VLBI determinations [Robertson, 1975] useful since they were less precise, and, in addition, involved a different telescope at Onsala which has yet to be located accurately with respect to the one we utilized (Table 1). Further, to assess inherent accuracy, comparisons would best be made with results from a more accurate technique. None now exists. The most accurate alternate technique currently available for the determination of these particular intercontinental baselines is based on observations of the Doppler shift of radio signals transmitted from satellites, plus standard ground surveys to tie the reference points for the Doppler measurements to those for the VLBI determinations. Such Doppler data exist (W. E. Carter, private communication, 1979), but the analysis and the ground surveys have not been completed and so no comparisons can yet be made.

Periodic repetitions of these intercontinental VLBI measurements are also planned with more radio telescopes in the interferometer array and with use of our new, more powerful, Mark III VLBI system. This system, among other attributes records data simultaneously from two well separated radio frequency bands, thus allowing ionospheric effects to be virtually eliminated as a source of error. The purpose of these further VLBI measurements will not only be to check on the repeatability of the present baseline determinations but, more importantly, to improve the precision to the level at which the expected [Minster and Jordan, 1978] ~1.7 cm/yr spreading rate between Europe and the United

States could be detected reliably within a decade.

References

Carter, W. E., A. E. E. Rogers, C. C. Counselman, and I. I. Shapiro, Comparison of geodetic and radio interferometric measurements of the Haystack-Westford baseline vector, J. Geophys. Res., 85, 2685-2687, 1980.

Cannon, W. H., R. B. Langley, W. T. Petrachenko, and J. Kouba, Geodesy and astrometry by transatlantic long baseline interferometry, J. Geophys. Res., 84, 229-236, 1979.

Clark, T. A., L. K. Hutton, G. E. Marandino, C. C. Counselman, D. S. Roberston, I. I Shapiro, J. J. Wittels, H. F. Hinteregger, C. A. Knight, A. E. E. Rogers, A. R. Whitney, A. E. Niell, B. O. Rönnäng, and O. E. H. Rydbeck, Radio source positions from very-long-baseline interferometry observations, Astron. J., 81, 599-603, 1976.

Hinteregger, H. F., I. I. Shapiro, D. S. Robertson, C. A. Knight, R. A. Ergas, A. R. Whitney, A. E. E. Rogers, J. M. Moran, T. A. Clark, and B. F. Burke, Precision geodesy via radio interferometry, Science, 178, 396-398, 1972.

Ma, C., Very long baseline interferometry applied to polar motion, relativity and geodesy, Ph.D. thesis, Univ. of Md., College Park, 1978.

McClure, P., Diurnal polar motion, Rep. X-592--73-259, NASA/Goddard Space Flight Cent., Greenbelt, Md., 1973.

Melchior, P., Precession-nutations and tidal parameters, Celestial Mech., 4, 190-212, 1971.

Minster, J. B., and T. H. Jordan, Present-day plate motions, J. Geophys. Res., 83, 5331-5354, 1978.

Robertson, D. S., Geodetic and astrometric measurements with very-long-baseline interferometry, Ph.D. thesis, Mass. Inst. of Technol., Cambridge, 1975.

Robertson, D. S., W. E. Carter, B. E. Corey, W. D. Cotton, C. C. Counselman, I. I. Shapiro, J. J. Wittels, H. F. Hinteregger, C. A. Knight, A. E. E. Rogers, A. R. Whitney, J. W. Ryan, T. A. Clark, R. J. Coates, C. Ma, and J. M. Moran, Recent results of radio interferometric determinations of a transcontinental baseline,

polar motion, and Earth rotation, in Time and the Earth's Rotation, edited by D. D. McCarthy and J. D. Pilkington, pp. 217-224, D. Reidel, Dordrecht, Netherlands, 1979.

Shapiro, I. I., Estimation of astrometric and geodetic parameters from VLBI observations, in Methods of Experimental Physics, edited by M. L. Meeks, pp. 261-276, Academic, New York, 1976.

Shapiro, I. I., D. S. Robertson, C. A. Knight, C. C. Counselman III, A. E. E. Rogers, H. F. Hinteregger, S. Lippincott, A. R. Whitney, T. A. Clark, A. E. Niell, and D. J. Spitzmesser, Transcontinental baselines and the rotation of the earth measured by radio interferometry, Science, 186, 920-922, 1974.

Shapiro, I. I., D. S. Robertson, C. A. Knight, C. C. Counselman III, A. E. E. Rogers, H. F. Hinteregger, S. Lippincott, A. R. Whitney, T. A. Clark, A. E. Niell, and D. J. Spitzmesser, Transcontinental baselines and the rotation of the earth measured by radio interferometry, Science, 191, 451, 1976.

Whitney, A. R., Precision geodesy and astrometry via very long baseline interferometry, Ph.D. thesis, Mass. Inst. of Technol., Cambridge, 1974.

Whitney, A. R., A. E. E. Rogers, H. F. Hinteregger, C. A. Knight, J. I. Levine, S. Lippincott, T. A. Clark, I. I. Shapiro, and D. S. Robertson, A very-long-baseline interferometer system for geodetic applications, Radio Sci., 11, 421-432, 1976.

Woolard, E. W., Theory of the rotation of the earth around its center of mass, Astron. Pap. Am. Ephemeris Nautical Alm., XV (I), 1953.

Woolard, E. W., Inequalities in mean solar time from tidal variations in the rotation of the earth, Astron. J., 64, 140-142, 1959.

GEODESY BY RADIO INTERFEROMETRY: PRECISION AND ACCURACY OF INTERCONTINENTAL DISTANCE MEASUREMENTS

T.A. Herring[*1], I.I. Shapiro[*1], T.A. Clark[2], C. Ma[2], J.W. Ryan[2],
J.C. Pigg[3], B.R. Schupler[3], C.A. Knight[4], D.B. Shaffer[4], N.R.
Vandenberg[4], R.J. Cappallo[5], H.F. Hinteregger[5], A.E.E. Rogers[5],
J.C. Webber[5], A.R. Whitney[5], G. Elgered[6], G. Lundqvist[6], B.E. Corey[7]

[1]Harvard-Smithsonian Center for Astrophysics
60 Garden St., Cambridge, Ma. 02138

[2]Goddard Space Flight Center, Greenbelt, Md. 20771

[3]Computer Science Corporation, Silver Spring, Md. 20910

[4]Interferometrics Incorporated, 8150 Leesburg Hts, Suite 1200,
Vienna, Va. 22180

[5]Haystack Observatory, Westford, Ma. 01886

[6]Onsala Space Observatory, S-439 00 Onsala, Sweden

[7]Department of Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge, Ma. 02139

Abstract

The precision and accuracy of intercontinental baseline length estimates are studied through a number of repeatability tests on 55 Mark III very-long-baseline interferometry (VLBI) observing sessions carried out between November 1979 and January 1982. We also present results from the analysis of a further 61 VLBI observing sessions carried out between January 1982 and June 1983. The estimate of the Westford, Massachusetts to Onsala, Sweden baseline length obtained from the ensemble of observing sessions has a statistical standard deviation of 0.5 cm. The weighted-root-mean-square (WRMS) scatter of the estimates from the 40 individual observing sessions involving these sites (each of approximately 24 hr duration), about the ensemble mean is 2.6 cm. Studies of these solutions indicate that the actual standard deviation of the baseline length estimate is probably more in accord with the repeatability rather than with the statistical formal estimate. The estimate of the rate of change of the length of this baseline is 1.6 ± 0.5 cm/yr, where the standard deviation of the rate is computed from the formal uncertainty of the rate scaled by the ratio of the scatter of the baseline estimates to the expected scatter computed from the standard deviations of the length estimates. The WRMS scatter of the 40 baseline length estimates about this "best-fit" straight line was 2.0 cm. Although the rate of change of the baseline length is over three standard deviations, we do not yet claim to have detected the relative motions of these two sites because our studies indicate that the VLBI estimates of the baseline lengths could be affected by systematic errors at the level of the apparent change in the length.

A unexpected conclusion from these studies is that the precision of the baseline-length estimates from solutions which incorporate low-elevation-angle observations (as low as $3^\circ.5$) is better, in general, than the estimates from solutions which do not include observations at elevation angles below 10°.

Introduction

The Mark III very-long-baseline interferometry (VLBI) system was introduced in April 1979 (Rogers et al., 1983). This system was designed to utilize two widely separated frequency bands, one each near 2.3 GHz and 8.3 GHz, so that the dispersive delay due to the Earth's ionosphere could be calibrated and, hence, eliminated as a major error source in the estimation of the distances between the radio telescopes that form the interferometers. The first intercontinental observing session to partially use dual-frequency-band data was carried out in November 1979, but only six hours of such data were obtained. In July 1980, the first full dual-frequency-band obseιving session, involving North American (N.A.) and European sites, was carried out. Between then and January 1983, over 30,000 dual-frequency observations have been made in observing sessions sponsored by the NASA (National Aeronautics and Space Administration) Crustal Dynamics Project and the National Geodetic Survey (NGS) Polaris program (Polar-motion Analysis by Radio Interferometric Surveying). In this paper we discuss the analysis of VLBI data obtained between November 1979 and June 1983 (36,000 observations). This data set includes 41 observing sessions involving both N.A. and European sites, with 40 of these observing sessions involving Onsala, Sweden.

This paper is the first of a series which will discuss the quality of the data, and the precision and accuracy of the geodetic parameters which can be estimated from these data. Here we address the question of the precision and accuracy of the estimates of the baseline lengths between the radio telescopes participating in these observing sessions.

The "true" accuracy of the baseline-length estimates cannot be determined reliably without a more accurate standard to which the VLBI results may be compared. No such standard exists which will allow accuracy to be checked at the necessary level of one part in 10^9 over intercontinental distances (corresponding to $\simeq 0.5$ cm for a baseline between radio telescopes located in Massachusetts

and Sweden). Without such a standard, we can still address the accuracy question by investigating the variations in the baseline-length estimates with changes in the data set or with changes in parameterization of the models used to analyze the data. These techniques, will, however, only place a lower bound on the estimates of the accuracy of the results.

The main purpose of the Crustal Dynamics geodetic VLBI experiments involving sites in North America and Europe is measure the contemporary rate of motion between these continents, or to place a useful upper bound on this rate. Measurement precision is more important for this purpose than accuracy. But an understanding of accuracy is required to identify sources of error that may change over the time interval used to estimate the rate, in our case several years.

The question of the accuracy of VLBI baseline-length estimates involves many facets of the technique. We limit discussion here to the major error source: the effects of the neutral atmosphere, or more precisely, the methods we have adopted to study these effects. We first discuss the mathematical and statistical models which were used to analyze the observations, and then the effects on the baseline-length estimates of (i) limiting the minimum elevation angle of observations used in the analysis; (ii) using surface weather data to calibrate the neutral atmosphere delay; and (iii) estimating parameters that relate elevation angle to air mass (see below). We also compare our results for the Haystack--to-Onsala baseline length with values previously published.

Mathematical and Statistical Models

The data analyzed consisted of dual-frequency-band interferometric group delays (Shapiro, 1976; Rogers et al., 1983), obtained in 123 observing sessions between November 1979 and June 1983. (Each of these sessions was of approximately 24 hr duration.) The frequency bands used were centered at 8.3 GHz (X-band) and 2.3 GHz (S-band).

Within each frequency band, data were obtained in four to eight channels, each of 2 MHz bandwidth. The total bandwidth spanned by these channels varied between 300 and 360 MHz at X-band, and 50 and 85 MHz at S-band, thus allowing accurate estimates of the group delays (Whitney et al., 1976).

These observations were analyzed within the framework described in Herring et al. (1981). However, the following models were used in place of their counterparts described there: (i) The IAU-1980 nutation series (Wahr, 1981); (ii) the Greenwich sidereal time formula of Aoki et al. (1982); and (iii) the atomic time-coordinate time relationship that includes relativistic corrections (Shapiro, et al., 1983). This new relativistic correction accounts for the change in the light propagation time in the potential fields of the Earth and the Sun. In addition, we added corrections for the ionospheric propagation delay, based on the dual-frequency-band observations, and for the neutral atmospheric propagation delay, based on surface weather data (Marini, 1975).

We assumed that the observation noise was an independent, zero mean, Gaussian random process. The variance of the noise for each observation was computed as the sum of two variances. The first variance was computed from the signal-to-noise ratio (SNR) of the observation and represents the uncertainty in the observation due predominantly to instrumental noise. The second variance was calculated from the condition that the χ^2/degree of freedom of the postfit residuals for each baseline be unity (see below). This variance is used to account for the pseudo-noise contribution from the inadequate models used to represent the data.

Data Analysis

The observations were analyzed using a sequential, weighted-least-squares technique (see, for examples, Kaula, 1966; Morrison, 1969), with the weights being taken as the inverses of the variances of the noise contributions to the observations. Unless otherwise

stated, the following classes of parameters were estimated in the solutions to be discussed below: (i) Coefficients of one to three polynomials in time, each of first or second order, and in some cases, coefficients of diurnal sine and cosine functions. These polynomial and trigonometric functions were used to represent the relative behavior of the hydrogen maser "clocks" at the different sites; (ii) a daily bias correction to the neutral atmosphere propagation delay calculated from surface weather data; (iii) all crust-fixed site coordinates (assumed time independent), except for those of one site which were used to define the origin of the coordinate system (see Herring et al., 1981); (iv) All radio source positions, save for one parameter -- the mean correction to the right ascensions of seven radio sources which served as the right ascension origin. (The seven radio sources chosen were 0851+202, 0923+392, 1226+023, 1641+399, 2134+004, 2200+420, and 2251+158); (v) daily corrections to the pole position and UT1 given by BIH Circular D; (vi) daily corrections to the nutation angles given by the IAU-1980 nutation series (to be discussed separately in Herring et al., 1983).

In addition to these parameters, we sometimes estimated a correction to the relationship between the elevation angle of the measurement and the air mass through which the signal passed (this relationship will be called the "mapping function"). We parameterized the mapping function given by Chao (1972), by introducing a parameter corresponding to a fractional correction to one constant in his empirical formula:

$$t_e = t_z/[\sin(e) + (1+a)r_1/(\tan(e)+r_2)] \ , \qquad (1)$$

where t_e is the atmospheric delay for an observation at an elevation angle, e, when the zenith delay is t_z, and $r_1=0.00143$ and $r_2=0.0445$ are two constants estimated (by Chao) by relating his empirical model to a ray-trace model, and a is the parameter we introduced.

VLBI data analysis also involves a number of subjective aspects:

(i) the parameterization of the relative behaviors of the hydrogen masers at the radio telescopes, (ii) the parameterization of the atmospheric delay behavior at each site, (iii) data editing, and (iv) data reweighting (see discussion below). We have already discussed the the models which are used to parameterize the relative clock behaviors and the atmospheric delays. The polynomial order and the epochs of the boundaries between estimates of the polynomial coefficients are decided on by the analysts after a visual inspection of the postfit residuals from a solution which utilizes relatively simple parameterizations of the clock behavior. Trends in the residuals which are radio-source independent are usually associated with clock variations. Similarly, trends which are elevation-angle dependent are usually associated with atmospheric variations. In our analysis a single atmospheric parameter was estimated each day at each site and served to remove any constant biases in the surface weather model calibration of the atmospheric delays. Data editing, i.e. the elimination of observations which appear to be defective, is also subjective in nature. However, an examination of the observations which in the past have been defective, the reason for the defect can often be established, thus allowing future observations to be checked for this condition. Consequently, as the the use the Mark III has increased, a series of quality checks has been developed which allow defective observations to be detected before the analyst examines the data. These checks have eliminated many of the subjective decisions associated with data editing. We discuss shortly the techniques which we used to ensure that a uniform editing criterion was used throughout our analysis. Data reweighting is the final subjective aspect of the data analysis we need to discuss. The scatter of the postfit residuals from a Mark III least squares solution is nearly always greater than the standard deviations of the observations would indicate. This greater scatter occurs because the observation standard deviations, which are calculated from the signal-to-noise ratio (SNR) of the observations, do not take into account the errors in the mathematical models used

to analyze the data and in the random variations of the frequency standards. The contribution of these additional uncertainties can be accounted for by adding a constant variance to the SNR-calculated variances and by using this new variance to calculate the weights used in the weighted least-squares solution. This operation is call "reweighting." The baseline-dependent variance is computed from postfit residuals and, hence, depends on the solution used to generate these postfit residuals. We discuss below the techniques we used to determine these baseline-dependent variances.

Our analysis was carried out in several stages with the aim of generating a data set which was both uniform in its editing and reweighting criteria. We firstly carried out a solution using all of the observing sessions. We then reanalyzed each session separately with the radio-telescope and radio-source positions constrained to the values obtained from the analysis of all of the data. We rejected all observations which had postfit residuals greater than three (reweighted) standard deviations. Using this edited data set, we calculated baseline-dependent variances which, when added to the SNR variances of the observations, would make the χ^2 per degree of freedom of the postfit residuals for each interferometry approximately unity. Since both the editing and the reweighting depend on the the station and source positions used in the analysis of each session, we repeated the above procedure three times to achieve convergence. The baseline-length estimates changed by up to 3.6 standard deviations (third iteration, formal) during these iterations. (The largest change was 3.6 cm for the Owens Valley to Onsala baseline.) The average change of the baseline-length estimates was 1.5 standard deviations (corresponding to approximately 1-2 cm; see Table 1).

The final data set used to study the accuracy of the baseline length estimates consisted of 22,892 group-delay observations of which 15,928 were used in the sequential weighted least squares solution. The data which were not used either failed quality checks during correlation (5,987 observations), or were rejected by the

editing criterion discussed above (977 observations). The large number of observations rejected by the "three-sigma" criterion is probably due to instrumental errors which were not detected during correlation (especially in the early observing sessions before the development of an extensive set of correlator checks), and to large postfit residuals caused by inadequate models. In addition to this data set, we also analyzed a further 8,688 observations obtained in observing sessions carried out between January 1982 and June 1983, which were not available when the accuracy studies were carried out. The combined data set was used to study the precision of the baseline length estimates.

Results

The data set

The data set of nearly 16,000 group delay observations, described in the previous section, was used to study the effects of the neutral atmosphere on the estimates of baseline length. The specific tests we performed were: (i) limiting the data set to observations whose elevation angles were above a minimum value called the "elevation-angle cutoff"; (ii) comparing solutions with and without a surface weather model prediction of the neutral atmosphere delay; and (iii) comparing the solutions with and without estimation of a correction to the atmospheric mapping function.

The result of limiting the data set to observations whose elevation angles were greater than the elevation cutoff are summarized in Figure 1. (A zero degree elevation-angle cutoff simply means that all available data were used in the solution. The actual minimum elevation angle of any observation in the data set was $3^{\circ}.5$.) The error bars shown in Figure 1 are the standard deviations of the differences in the baseline-length estimates (since the solution with the lower elevation cutoffs contain data which were not used in those with higher elevation-angle cutoffs, there should be differences between the solutions consistent with these standard

deviations). (It is easily shown that the variance of the difference is the difference of the variances of the two solutions, in the sense of the larger variance (solution with the higher elevation-angle cutoff) minus the smaller variance (the solution with the lower elevation-angle cutoff).)

Figure 1 shows that the changes in the baseline-length estimates when the elevation-angle cutoff is changed, are not in accord with the statistics of the difference. Systematic differences of up to 1 cm exist between the 0° and 10° cutoff solutions. The differences between the solutions with the 10° cutoff and those with the other cutoffs do not show such strong systematic trends, but still show scatter larger than the appropriate standard deviations would indicate. In particular the differences between the solution with 10° and 20° cutoffs show large differences (between 3 and 5 cm) for baselines longer than 8,000 km. These large differences, and their associated large uncertainties, could be due to geometric degeneracy. (Many of the low-elevation-angle observations in this data set were included to strengthen the geometry of these very long baselines.)

To further investigate the origin of the systematic differences between the solutions using different elevation-angle cutoffs, we compiled the statistics of the postfit residuals from these solutions. We accumulated the statistics in a matrix whose row and column indices were defined by the air masses (calculated from the elevation angles) at the two sites for each baseline. The air mass bins were 0.5 air masses wide and ranged between 1.0-1.5 and 14.0-14.5 air masses (the upper limit corresponds to an elevation angle of $3^\circ.1$). We present in Figure 2, for the 0° cutoff and the 10° cutoff solutions, the statistics of the column of this matrix for which the air mass bin was 1.0-1.5 at one end of the baseline. We show the weighted mean and the normalized-root-mean square (NRMS) scatter of the residuals in each bin. (The NRMS scatter is the square root of the χ^2/degree of freedom for the residuals in the bin. We plot this quantity, rather than the χ^2/degree of freedom, to decrease

the dynamic range of the plots.) The error bars shown for the weighted means are one standard deviation estimates, based on the reweighted standard deviations of the observations. The 95% confidence intervals shown for the NRMS scatter are also based on the reweighted standard deviations.

It is clear that the NRMS scatters show a dependence on the air mass in both the 10° and 0° cutoff solutions. This elevation angle dependence is not accounted for in the observation standard deviations, i.e., the low elevation angle (high air mass) observations are overweighted weighted in these analyses. The effects of this higher weighting can be seen in the weighted mean results. The positive bias of the weighted means in air mass bins below 10° elevation angle (as seen in the 10° elevation angle cutoff solution) are considerably reduced in the 0° cutoff solution. This reduction in the weighted means implies that this systematic bias has been absorbed into the parameter estimates.

To test the hypothesis that the relationship between air mass and elevation angle could be systematically biased, we carried out a solution using a 0° elevation angle cutoff, but with a site- and observing-session dependent correction to the mapping function estimated. The mapping function correction was estimated at a site only when there were observations with elevations angles less than 10°. We show the difference between this solution and the solution with the 10° cutoff (without a correction to the mapping function estimated) in Figure 3. The error bars shown on the plot are the same as those in Figure 1.a. Figure 3 does not show the systematic trends of Figure 1.a indicating that an error in the mapping function could be partially responsible for the differences between the solutions that use different elevation angle cutoffs. While estimating corrections to the mapping function seems to remove the systematic difference between the solutions with different elevation angle cutoffs, the estimated values of a (see Equation 1) have large scatter with weighted mean values of only a few percent. (The correction to r_1 at Onsala was 4% with a RMS scatter of 11%.)

However, the magnitude of the weighted mean values are consistent with the elevation angle dependence of the postfit residual mean values shown in Figure 2. These studies of the atmospheric mapping function are very preliminary and certainly need to be continued. We could, for example, assume that the correction is site and observing session independent and estimate a single value for the correction from all of the data. (Our software would need some changes to allow this estimation.) An alternative approach would be to carry out specialized observing sessions in which a combination of very low elevation and high elevation angle observations were made in rapid succession. These observing sessions at a number of sites and in different seasons could allow accurate estimation of corrections to the mapping function.

We further investigated the effects of our atmospheric delay models by repeating the solution with the 10° elevation angle cutoff without the surface weather model prediction of the atmospheric delay. (In both solutions we estimated zenith delay corrections once daily at each site.) The differences between the baseline-length estimates from these two solutions are shown in Figure 4. Again we see a bias between the baseline-length estimates which increases with baseline length. (In this case the error bars shown are the standard deviations of the baseline-length estimates.) These results indicate that care should be taken in combining data sets with and without surface weather model calibration of the atmospheric delay.

In Table 1 we give the baseline-length estimates obtained from the solution with the 10° cutoff with surface weather model calibration of the atmospheric delay We also give the WRMS and NRMS scatters of the baseline-length estimates, about these values, obtained from a solution in which the source positions were constrained to the values obtained from the ensemble of data and the station positions were estimated separately for each observing session. We also tabulate the WRMS and NRMS scatters of baseline-length estimates from same type of analysis with the 0° cutoff. These scatters highlight why we believe low elevation angle observations

should be studied. In nearly all cases the WRMS scatter of the baseline-length estimates from the solution with the 0° cutoff are less than those from the solution with the 10° cutoff, indicating that the solution with the 0° cutoff is more precise than that with the 10° cutoff. In both cases the normalized RMS scatters (which should be unity if the scatter of the results were in accord with their statistics), generally range between 1.5 and 2.0. These values indicate that the uncertainties of the baselines-length estimates are "optimistic" by a factor of between 1.5 and 2.0. The final entry in Table 1 is the contribution of the new relativistic correction to the baseline length. (We include these values so that our results may be compared with previously published values of these baseline lengths.) This contribution has an average value of -14 parts in 10^9 (Shapiro et al., 1984).

We show the estimates of the Westford-to-Onsala baseline length for each observing session in Figure 6. The slope of a linear least squares regression through these results (from 3.0 years of data) is 1.6 ± 0.5 cm/yr (the uncertainty is the statistical estimate of standard deviation, scaled by the NRMS scatter of the baseline-length estimates).

Estimates of some of these baseline lengths have been published in the past; we now compare our estimates with those previously published values. In 1981, the first estimates of the Haystack-to-Onsala baseline were published (Herring et al., 1981). At the time we felt that this estimate of the baseline length was uncertain at only the subdecimeter level. However, that value (after compensation for the relativistic correction; see Table 1) is 21 cm longer than the value given here; more than twofold worse than we had expected. We now believe that the estimate obtained from the single-frequency-band Mark I data was positively biased by an amount greater than expected because of the neglect of the ionospheric delay.

More recently published values (Robertson and Carter, 1982) for the Haystack-to-Onsala baseline length, estimated from a subset of the data used here, have also been positively biased by 10 cm.

In these solutions some source positions were held fixed at the values obtained from the analysis of Mark I data (Robertson and Carter, 1982). The value given by Rogers et al., 1983, agree within 2 cm with the value given here. As we have seen, differences of 2 cm can be accounted for because of slightly different solution techniques. (These results were also obtained from a subset of the data analyzed in this paper.)

Conclusions

We have presented results which show intercontinental distance measurement using the Mark III VLBI system have a repeatability of ~2.6 cm for 40 observing sessions. We have also attempted to study the accuracy of these results by carrying out a number of repeatability tests. These tests have shown that the estimates of the baseline lengths can change by several centimeters depending on the data set and the atmospheric calibrations used in the analysis. (For the Westford-to-Onsala baseline, the changes in the baseline lengths of up to 2 cm occurred.)

The estimated rate of change of the Westford to Onsala baseline is 1.6 ± 0.5 cm/yr. The standard deviation given is the statistical estimate of the rate from a weighted least squares regression through the baseline length estimates, scaled by the NRMS scatter of the these values about their mean. Although the rate of change of the baseline length seems to be significant, we do not yet claim to have detected motion between these two sites because the total change in the baseline length over the three year period is only 4.8 cm. There is a possibility that this apparent change in the length could be due to changes in the systematic errors which affect our results. However, these errors are bounded at about the 5 cm level and a continued lengthening of the baseline will be difficult to dismiss as due to systematic error.

Our studies of the dependencies of the baseline length estimates on elevation angle cutoff and on the statistics of the postfit

residuals have indicated that the relationship between elevation angle and air mass may be systematically biased. This subject needs further study because our results also show that the repeatability of the baseline lengths is better when low elevation angle observations are included in the analysis. However, this improvement in the repeatability does not necessarily imply an improvement in accuracy.

A problem which we need to solve is the assessment of "realistic" standard deviations of the baseline-length estimates. In nearly all cases the WRMS scatter of the baseline length estimates about their values obtained from the ensemble of data are 1.5 to 2.0 times the amount expected from the standard deviations. There are two reasons for this underestimation of the standard deviations. Firstly, unmodeled or inadequately modeled effects such as water vapor delay, frequency standard variations, radio source structure, and ocean loading will contribute to the variability of the baseline length estimates. These variations are not totally accounted for in the least squares estimates of the uncertainties (to some extent they are accounted for through the reweighting of the observation uncertainties). In a future paper we will discuss methods which can be used to assess these contributions to the uncertainties to the baseline length estimates. Secondly, there is a correlation between the pseudo-noise contributions (due to the unmodeled site dependent errors) to the observations involving a given site at each epoch. (There are also possibly temporal correlations between the model errors.) We currently do not account for these correlations in the least squres solutions which will cause the parameter estimate uncertainties to be overly optimistic. We are currently developing software which will calculate the expected correlations between observations based on the variances added to the SNR variances.

The VLBI data used in the analysis for this paper, and additional data obtained in more recent observing sessions, can be used for many purposes. In future papers we will discuss the performance

of the Mark III VLBI system, the positions of the pole with respect to the crust of the earth and in inertial space (or strictly the quasar reference frame) which were estimated in these solutions, techniques to eliminate the polynomial modeling of the relative behavior of the hydrogen masers (by use of Kalman filters), and the precision and accuracy of the calibration of water-vapor delay with water vapor radiometry.

Acknowledgments

* Formerly at the Massachusetts Institute of Technology.

References

Aoki, S., B. Guinot, G.H. Kaplan, H. Kinoshita, D.D. McCarthy, and P.K. Seidelmann, The new definition of universal time, *Astron. Astrophys.*, *105*, 359-361, 1982.

Chao, C.C., A model for tropospheric calibration from daily surface and radiosonde balloon measurements, *Tech. Mem. 391-350*, Jet Propulsion Laboratory, Pasadena, Calif., 1972.

Herring, T.A., B.E. Corey, C.C. Counselman III, I.I. Shapiro, B.O. Ronnang, O.E.H. Rydbeck, T.A. Clark, R.J. Coates, C.Ma, J.W. Ryan, N.R. Vandenberg, H.F. Hinteregger, C.A. Knight, A.E.E. Rogers, A.R. Whitney, D.S. Robertson, and B.S. Schupler, Geodesy by Radio Interferometry: Intercontinental distance determinations with subdecimeter precision, *J. Geophys. Res.*, *86*, 1647, 1981.

Kaula, W.M., *Theory of satellite geodesy*, Blaisdell Publishing Company, Waltham, Mass., pp. 125, 1966.

Marini, J.W., Correction of radio range tracking data for atmospheric refraction at elevations above 10 degrees, *Internal NASA Memorandum*, Feb., 1974.

Morrison, N., *Introduction to sequential smoothing and prediction*, Mc.Graw-Hill Book Company, New York, pp. 645, 1969.

Robertson, D.S., Geodetic and astrometric measurements with very-long-baseline interferometry, Ph.D. Thesis, Mass. Inst. of Technol., Cambridge, 1975.

Robertson, D.S. and W.E. Carter, in *IAU Colloquium No. 63: High Precision Earth rotation and Earth-Moon Dynamics*, O.

Calame, Ed. Reidel, Dordrecht, 1982.

Rogers, A.E.E., R.J. Cappallo, H.F. Hinteregger, J.R. Levine, E.F. Nesman, J.C. Webber, A.R. Whitney, T.A. Clark, C. Ma, J. Ryan, B.E. Corey, C.C. Counselman, T.A. Herring, I.I. Shapiro, C.A. Knight, D.B. Shaffer, N.R. Vandenberg, R. Lacasse, R. Mauzy, B. Rayhrer, B. Schupler, and J.C. Pigg, Very-long-baseline radio interferometry: the Mark III system for geodesy, astrometry, and aperture synthesis, _Science,_ _219,_ 51-54, 1983.

Shapiro, I.I., D.S. Robertson, C.A. Knight, C.C. Counselman III, A.E.E. Rogers, H.F. Hinteregger, S. Lippincott, A.R. Whitney, T.A. Clark, A.E. Niell, and D.J. Spitzmesser, Transcontinental baselines and the rotation of the earth measured by radio interferometry, _Science,_ _191,_ 451-454, 1976.

Shapiro, I.I. _et al._, in preparation, 1983.

Wahr, J. M., The forced nutations of an elliptical, rotating, elastic and oceanless earth, _Geophys. J. R. Astron. Soc.,_ _64,_ 705-727, 1981.

Whitney, A.R., A.E.E. Rogers, H.F. Hinteregger, C.A. Knight, J.I. Levine, S. Lippincott, T.A. Clark, I.I. Shapiro, and D.S. Robertson, A very-long-baseline interferometer system for geodetic applications, _Radio Sci.,_ _11,_ 421-432, 1976.

Table 1 Summary of baseline-length estimates

Baseline	# of exps.	length[+] (m)	sigma[+] (m)	WRMS[+] (m)	NRMS[+]	WRMS[*] (m)	NRMS[*]	Relativity correction
HA-CH	7	5,072,314.364 ± 0.006	0.029	2.16	0.028	2.28		−0.069
HA-HR	30	3,135,640.956 ± 0.004	0.017	1.25	0.017	1.28		−0.045
HA-OV	22	3,928,881.556 ± 0.003	0.015	1.50	0.015	1.54		−0.054
HA-ON	25	5,599,714.373 ± 0.005	0.026	1.60	0.023	1.82		−0.077
HA-WE	6	1239.395 ± 0.002	0.004	1.14	0.004	1.14		−0.000
HA-EF	6	5,591,903.441 ± 0.006	0.024	1.53	0.026	1.74		−0.077
HA-NR	3	845,129.839 ± 0.002	0.010	2.31	0.009	2.31		−0.012
CH-HR	7	7,663,737.269 ± 0.010	0.060	1.88	0.063	1.98		−0.105
CH-OV	7	7,846,991.133 ± 0.009	0.047	1.81	0.050	1.92		−0.106
CH-ON	7	1,109,864.304 ± 0.003	0.004	0.46	0.008	1.00		−0.016
CH-WE	0	5,073,265.270 ± 0.007						−0.069
CH-EF	0	589,796.542 ± 0.004						−0.008
CH-NR	0	5,834,122.784 ± 0.008						−0.078
HR-OV	26	1,508,195.356 ± 0.002	0.012	1.41	0.012	1.41		−0.023
HR-ON	37	7,940,732.104 ± 0.009	0.063	1.49	0.054	1.38		−0.111
HR-WE	120	3,134,927.971 ± 0.003	0.021	1.47	0.021	1.51		−0.045
HR-EF	5	8,084,184.751 ± 0.011	0.064	1.54	0.043	1.05		−0.113
HR-NR	4	2,354,633.977 ± 0.003	0.010	1.52	0.009	1.25		−0.033
OV-ON	26	7,914,130.853 ± 0.008	0.039	1.41	0.035	1.44		−0.110
OV-WE	28	3,928,579.286 ± 0.004	0.017	1.63	0.015	1.51		−0.055
OV-EF	6	8,203,742.365 ± 0.010	0.049	1.99	0.048	2.00		−0.114
OV-NR	4	3,324,244.142 ± 0.004	0.009	1.32	0.009	1.35		−0.046
ON-WE	40	5,600,741.326 ± 0.006	0.026	1.43	0.023	1.58		−0.078
ON-EF	5	832,210.499 ± 0.003	0.008	1.28	0.007	1.18		−0.013
ON-NR	2	6,319,317.417 ± 0.008	0.027	1.19	0.018	0.88		−0.087
WE-EF	0	5,592,850.979 ± 0.007						−0.077
WE-NR	4	844,148.075 ± 0.002	0.003	0.70	0.003	0.71		−0.012
EF-NR	1	6,334,648.392 ± 0.009	0.067	2.74	0.069	2.82		−0.087

Table 1 Continued.

$^+$ 10° elevation angle cutoff solution

*0° elevation angle cutoff solution

The codes for the stations are: HA Haystack, Massachusetts; WE Westford, Massachusetts; ON Onsala, Sweden; CH Chilbolton, England; EF Effelsberg, Federal Republic of Germany; NR National Radio Astronomy Observatory, Virginia; HR Harvard Radio Astronomy Telescope, Texas; OV Owens Valley Radio Observatory, California. The repeatability of the estimates of baselines involving Westford have been calculated including observing sessions which used the Haystack observatory, by inferring the lengths to Westford from the lengths to Haystack. The weighted RMS scatter of the postfit residuals from the 10° and the 0° cutoff solutions were 0.090 nsec (for 15,615 observations) and 0.091 nsec (for 15928 observations), respectively. The speed of light used to convert measured delays to distances was 2.99792458×10^8 m/sec.

Figure Captions

Figure 1. Differences in the baseline length estimates between solutions with different elevation angle cutoffs. The error bars represent the standard deviation of the difference between the estimates due to the additional data included in the lower elevation cutoff solutions. a) 0° elevation angle cutoff solution minus the 10° elevation angle cutoff solution; b) 10° elevation angle cutoff solution minus the 15° elevation angle cutoff solution; c) 10° elevation angle cutoff solution minus the 20° elevation angle cutoff solution

Figure 2. The accumulated statistics of the postfit residuals in air mass bins between 1 air mass and 10 air masses for the 0° cutoff and the 10° cutoff solutions. The normalized RMS (NRMS) scatter is the square root of the chi-squared per degree of freedom for the postfit residuals. The 95% confidence intervals for each air mass bin give the region in which the NRMS scatter should lie (with 95% probability) if the scatter of the postfit residuals were in accord with their statistics.

Figure 3. Differences of the baseline-length estimates between the 10° elevation angle cutoff solution and the 0° elevation angle cutoff solution with a correction to the atmospheric mapping function estimated.

Figure 4. Differences of the baseline-length estimates between the 10° elevation angle cutoff solutions with and without a surface weather model calibration of the neutral atmospheric delay. In each solution we estimated corrections to the zenith delay at each site.

Figure 5. Estimates of the Westford-to-Onsala baseline length from the 10° elevation angle cutoff solution. The station positions

were estimated separately for each observing session with the source positions constrained to the values obtained from the global solution based on all of the data. The uncertainty of the slope has been scaled by the NRMS scatter of the individual length estimates.

Figure 1.

a.
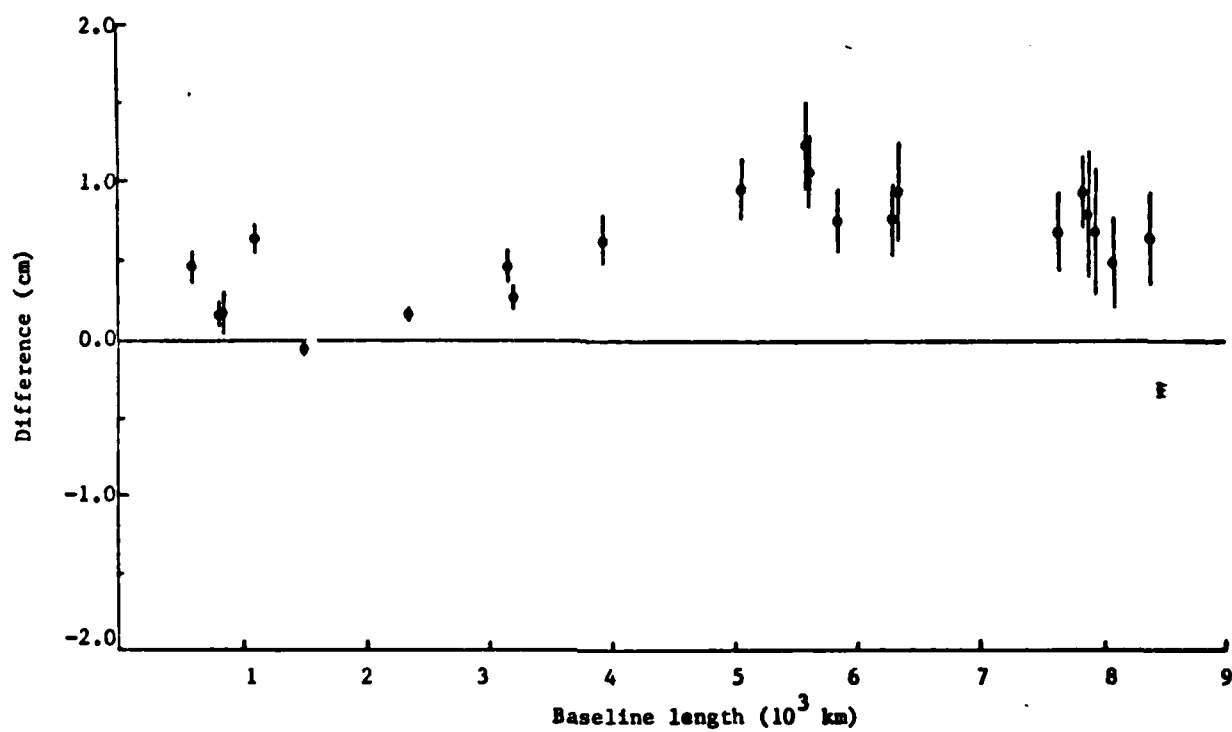
b.
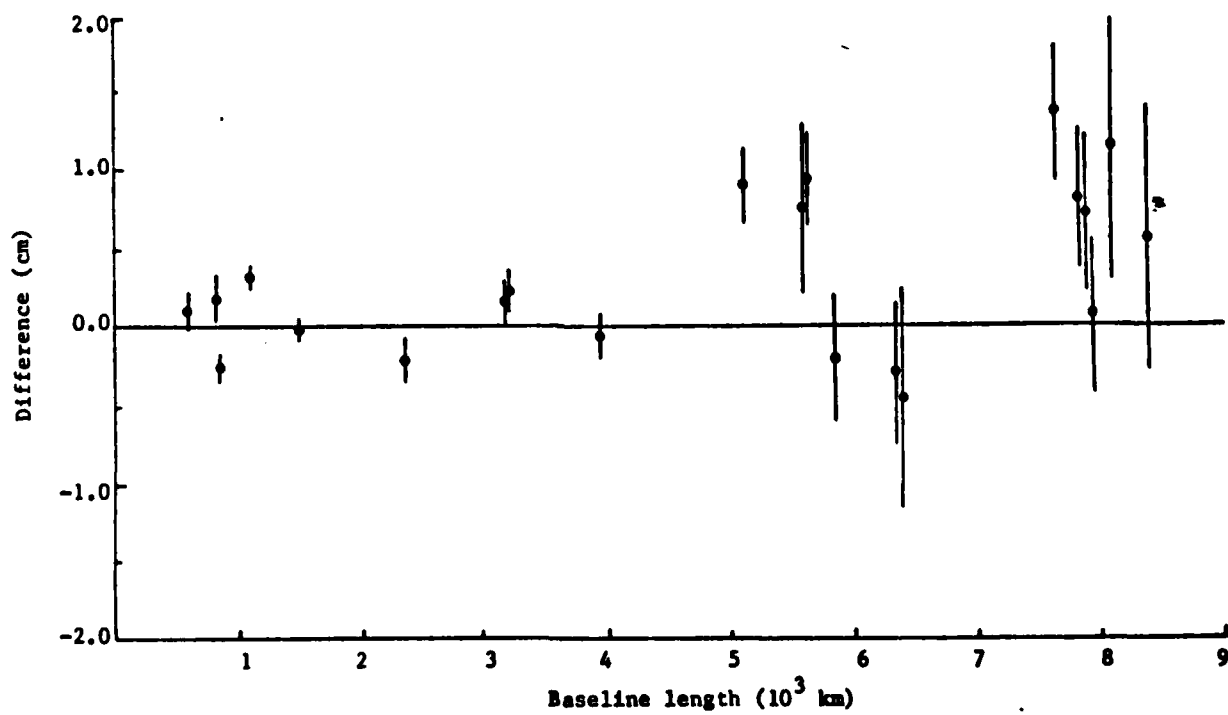
Figure 1 Continued

Figure 2

Figure 3

Figure 4
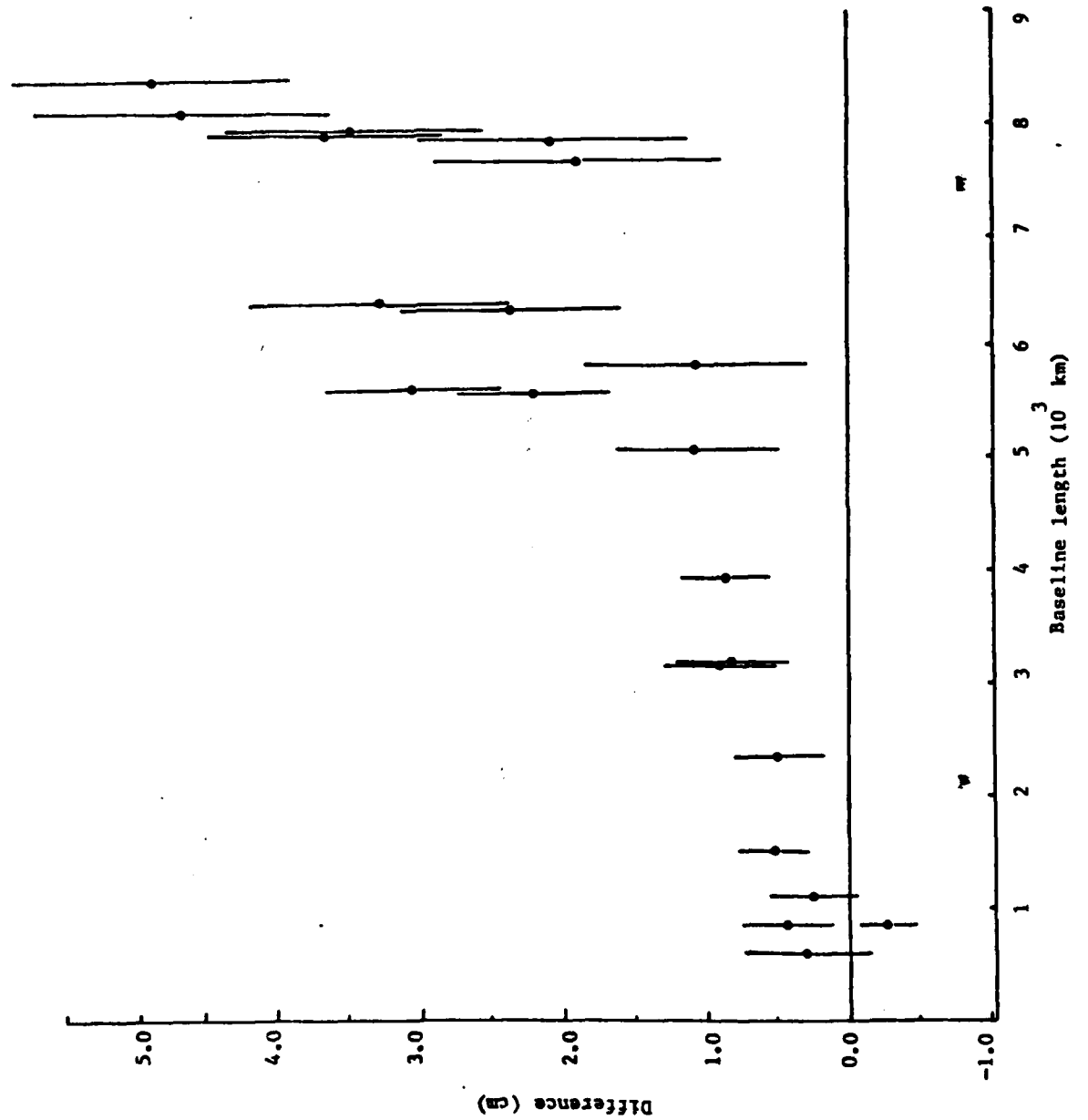
Figure 5
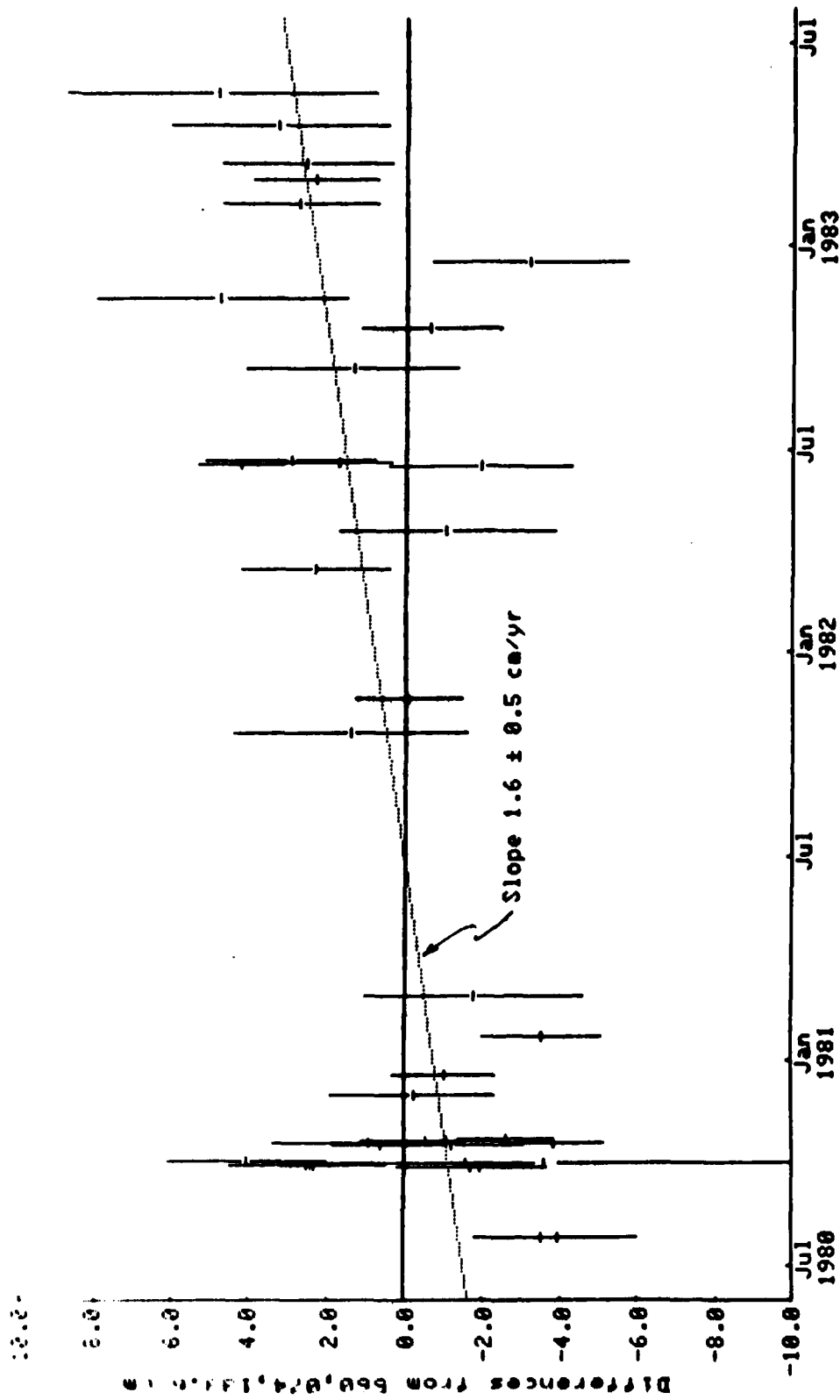
Determination of Tidal Parameters from VLBI Observations

T.A. HERRING, B.E. COREY, C.C. COUNSELMAN, and I.I. SHAPIRO,
Cambridge, MA, USA

A.E.E. ROGERS and A.R. WHITNEY, Westford, MA, USA

T.A. CLARK, C.A. KNIGHT, C. MA, J.W. RYAN, B.R. SCHUPLER, and N.R.
VANDENBERG, Greenbelt, MD, USA

G. ELGERED, G. LUNDQVIST, and B.O. RÖNNÄNG, Onsala, Sweden

J. CAMPBELL, Bonn, FRG
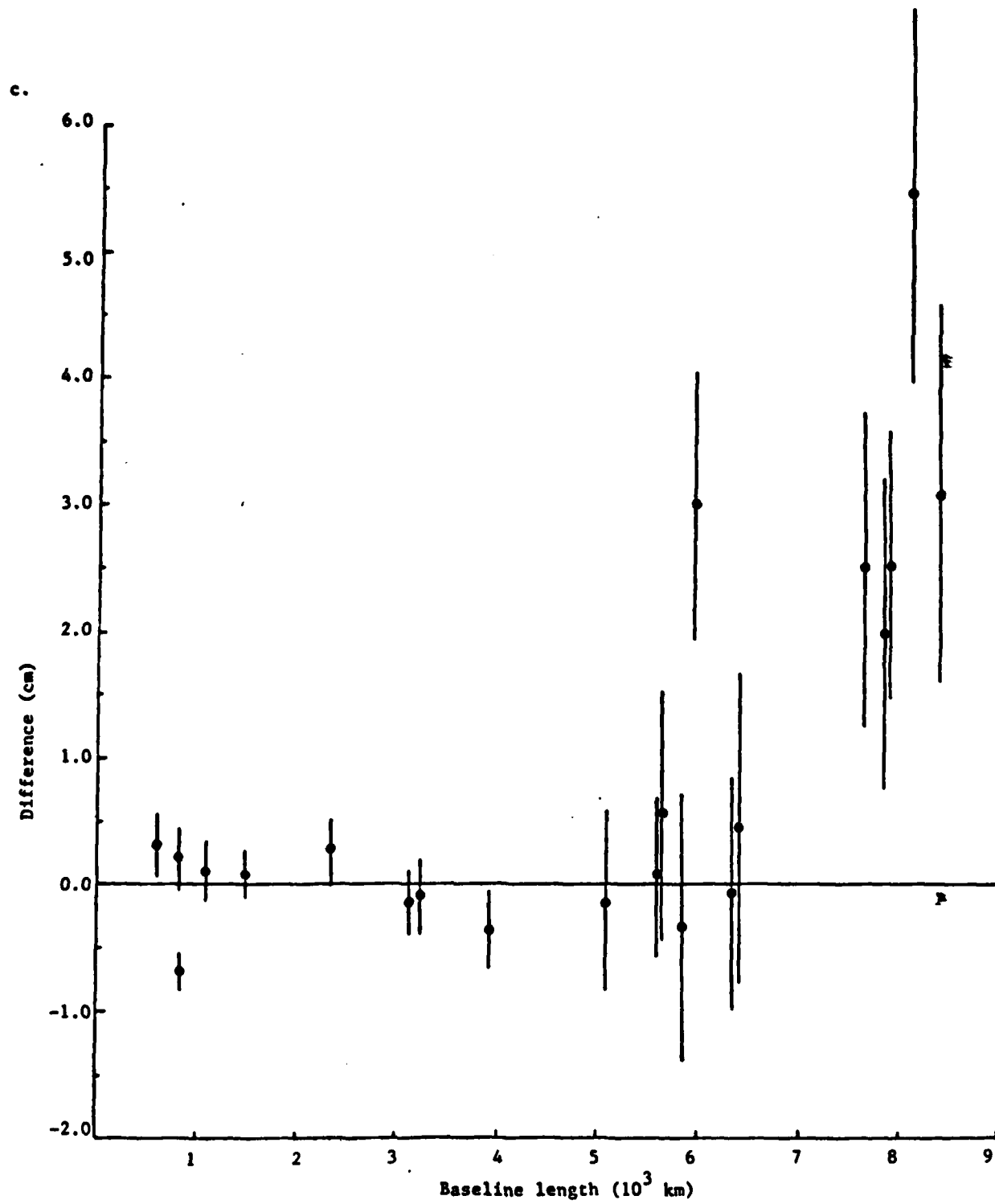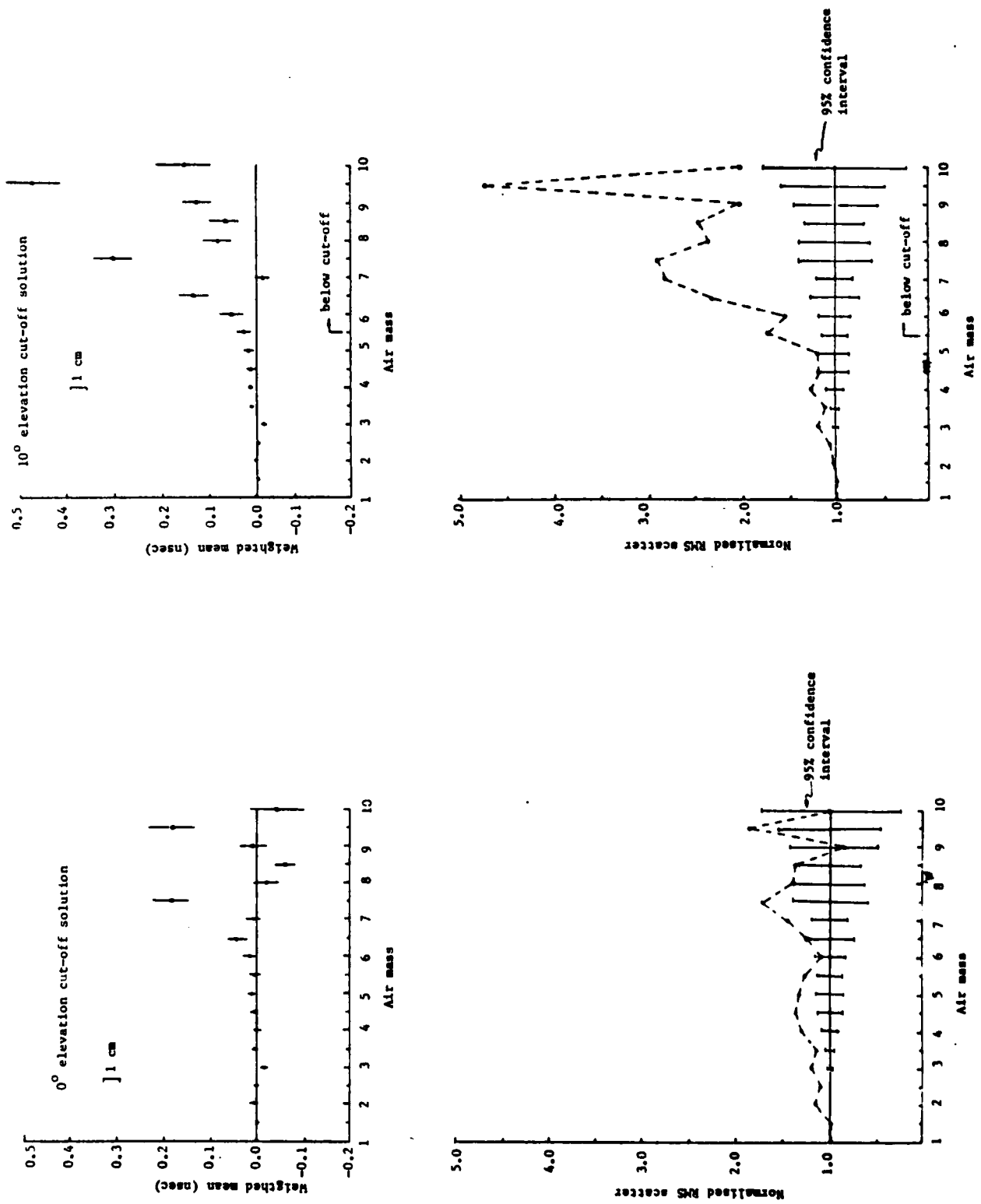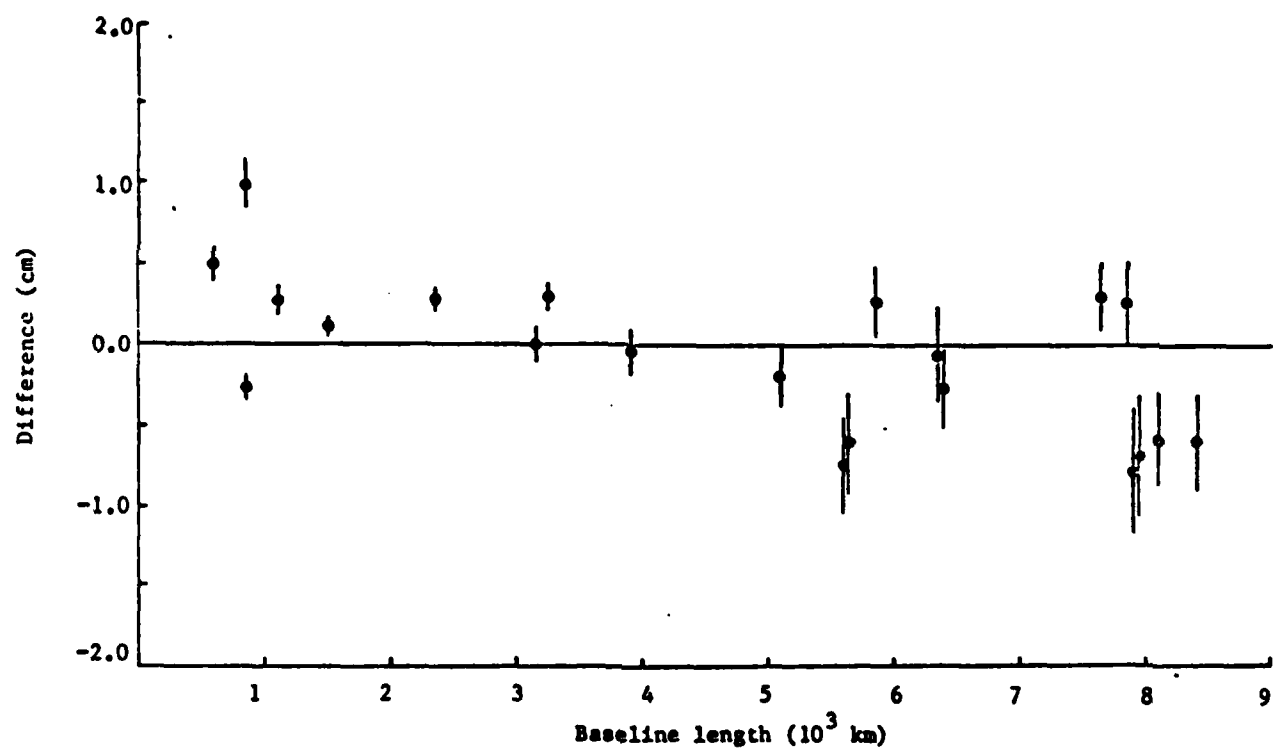
P. RICHARDS, Chilbolton, England

with 3 figures

Abstract: VLBI observations made in September and October 1980
with up to five telescopes distributed over northern Europe
and the United States yielded values for solid earth tide
parameters:

$$h = 0.62 \pm 0.01 \text{ (radial Love number)}$$
$$\ell = 0.11 \pm 0.03 \text{ (horizontal Love number)}$$
$$\phi = 1° \pm 1° \text{ (tidal lag angle)}$$

Keywords: Love numbers, tidal lag, radio interferometry

1. Introduction

Very Long Baseline Interferometry (VLBI) has been used since 1969
to determine vector separations between radio telescopes and posi-
tions of radio sources. Initially, the precision of baseline length
determinations was about 1 m (HINTEREGGER et al., 1972). Since
then the precision of baseline length determinations has improved
considerably and is now under 5 cm for baselines from 1000 km to
4000 km (ROBERTSON et al., 1979) and subdecimeter for intercontin-
tal lengths of about 6,000 km (HERRING et al., 1981). Observations
made with the new Mark III VLBI system should allow baseline
lengths to be determined with a precision of approximately 1 cm.
As the precision of observations increases, we must ensure that the
models are also improved for the interpretations of the observa-
tions.

Solid earth tides have long been incorporated in the model used
to analyze VLBI observations, and preliminary results have already

been obtained for the radial Love number, h (ROBERTSON, 1975).
Simultaneous analyses of six (6) years of VLBI data, obtained with
the Mark I system (RYAN et al., 1981) yielded h = 0.620±0.005. The
standard deviation of the radial Love number is based solely on the
scatter of the postfit residuals; the true standard error is likely
to be about a factor of three larger, partly owing to the scatter
of results from disjoint data sets. In this paper we present our
determinations of radial and horizontal Love numbers and the tidal
lag angle, from analyses of two weeks of VLBI data obtained during
September and October 1980 with the Mark III system.

2. Data analyses

The analyzed data consisted of interferometric group delays
(SHAPIRO, 1976) which were obtained during two one-week sessions,
separated by two weeks, in September and October, 1980. Up to five
antennas were involved in each session. Radio telescopes in West-
ford, Massachusetts, Ft. Davis, Texas, Big Pine, California, and
Onsala, Sweden participated in all sessions. In the first two
days of the first week, the radio telescope at Bonn, FRG, partici-
pated, and the radio telescope at Chilbolton, England, participated
for all of the second week. All observations were made with the
dual-band Mark III system with center radio frequencies of about
8.3 GHz, and 2.3 GHz for the two-frequency bands. A multichannel
bandwidth synthesis technique (WHITNEY et al., 1976) was used, with
each recorded channel having a bandwidth of 4 MHz (upper & lower side-
bands); synthesized bandwidths were 300 MHz at 8.3 GHz; 75 MHz at 2.3 GHz.

The theoretical models and the coordinate system definitions
used for the analyses are summarized in HERRING et al. (1981). The
dual-band observations were used to calibrate the ionospheric
delays and surface weather data were used to estimate the tropo-
spheric delays. Appropriate parameters for each day were included
in the theoretical model to represent relative clock behavior and
possible residual errors in the tropospheric path delay.

Because this paper deals with the determination of the Love
numbers for the solid Earth, only the tidal models will be described
in detail. For convenience in calculation, the tidal disturbing
potential is modified and written as

$$V = GM \left(1/|\vec{r}| - 1/|\vec{a}| - \vec{r} \cdot \vec{a}/|\vec{r}|^3 \right), \tag{1}$$

where $\vec{r}$ is the geocentric position vector of the disturbing body, rotated about the Earth's rotation axis by the tidal lag angle ϕ, $\vec{a}$ is the geocentric position vector of the radio telescope and GM is the product of the gravitational constant and the mass of the disturbing body.

Using Eq. 1, the disturbing potential and its derivatives with respect to latitude and longitude, $\partial V/\partial \theta$ and $\partial V/\partial \lambda$, respectively, are calculated for each observation at each telescope from ephemeris positions for the sun and moon. These values are then converted to displacements using

$$u_r = h \ (V) g^{-1}$$
$$u_\theta = \ell \ (\partial V/\partial \theta) g^{-1} \tag{2}$$
$$u_\lambda = \ell \ (\partial V/\partial \lambda) \ (g\cos \theta)^{-1},$$

where h and ℓ are the radial and horizontal Love numbers, respectively, g is the normal acceleration of gravity at the site, and u_r, u_θ, and u_λ are, respectively, the radial, north, and east displacements of the telescopes.

The displacements are represented very accurately using Eq. 2, since the disturbing body positions are available from other computations in the VLBI data analyses. The disadvantage in using Eq. 2 is the required use of Love numbers and lag angle that are independent of the harmonic content of the forcing function. The parameters h, ℓ, and ϕ and their covariances are estimated simultaneously with the other relevant parameters form the VLBI data using standard weighted least squares.

3. Results

Figures 1, 2 and 3 present our estimates of the tidal lag angle and the radial and horizontal Love numbers. Two features are apparent: 1) the scatter of each of the results about its weighted mean is up to 10 times the formal standard deviations of a single estimate, and 2) the distribution of the individual results about their weighted mean seems to be systematic in each case. These systematic trends might arise at least in part from neglected perturbations on the earth tides due to ocean-loading effects.

The variation in the formal standard deviations of the estimates of the tidal parameters is approximately in accordance with that of the number of usable observations obtained each day. These

numbers ranged between 236 for October 1, 1980 with four stations
participating to 777 for October 22, 1980 with five stations par-
ticipating. The small numbers of observations obtained on October
1 and 2 were due to a telescope pointing problem, subsequently
fixed , at the antenna in Westford, Massachusetts. Even accounting
for the variations in formal standard deviations, the scatter in
the results from the first week are greater than that from the
second week. There are several possible explanations for this
difference: 1) During the first two days of the first week, the
antenna in Bonn, FRG participated. If the Love numbers are a
function of the station locations, then the average of the estimates
from these two days could be expected to differ from that for the
remainder of the week; and 2) For two days of the first week, all
observations west of the meridian were lost at the Westford antenna.
This bias in the distribution of the observations on the sky could
increase the effects of systematic errors on the results.
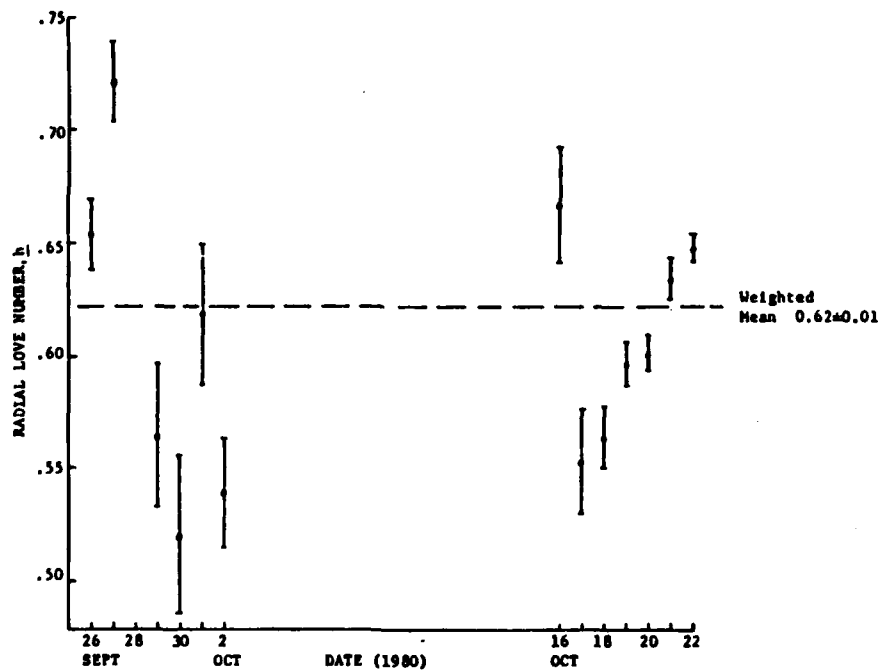
Figure 1. Independent estimates of the radial Love number, $\underline{h}$.
The error bars represent plus and minus one standard
deviation, determined by the addition of a constant to the
variance for each measurement such that x^2 per degree of
freedom was unity from each weighted least squares solution.
The quoted uncertainty for the weighted mean was calculated by
multiplying its theoretical standard deviation by the ratio of
the scatter of the individual results about their weighted
mean to the expected scatter due to random error.

Figure 1 shows the weighted mean of the results from this analysis:

$$h = 0.62 \pm 0.01$$
$$\ell = 0.11 \pm 0.03 \tag{3}$$
$$\phi = 1^{\circ} \quad \pm 1^{\circ}$$

where the quoted uncertainties are based on the scatter of the individual results about the weighted mean.

The determination of h is consistent with the values obtained previously from other VLBI data (ROBERTSON, 1975 and RYAN et al., 1981).

Love numbers and lag angles for each telescope location were also estimated. The largest fractional difference between such an estimate for h and the corresponding global value occurred for Big Pine, California, which has a value of $h = 0.73 \pm 0.03$. In view of the systematic, time-dependent behavior and the scatter of our various results, we do not consider this difference to be significant.

The values of the Love numbers obtained in this study also agree well with values determined previously using other techniques. Theoretical values of the Love numbers typically range from 0.60 to 0.62 for h, 0.08 to 0.09 for ℓ (MELCHIOR, 1978), with the variation resulting from the choice of earth model and core rigidity. Experimental determination of the Love numbers using gravimetric, tilt and extensometric observations generally give combinations of the Love numbers (h, ℓ and k) which agree with the theoretical estimates although the results can be significantly perturbed by ocean tides and local geology. The tidal lag angle is also obtained from these techniques. Estimates of its value range between $\pm 5^{\circ}$, depending on tidal component, with quoted uncertainties that range from $0^{\circ}.1$ to 3° with the larger lag angles associated with the larger uncertainties (MELCHIOR, 1978).

Analyses of orbits of earth satellites also yield estimates of the lag angle. Recently, CAZENAVE & DAILLET (1981) obtained lag angles of $0^{\circ}.32 \pm 0^{\circ}.28$ for the M_2 earth tide. They also obtained $1^{\circ}. 41 \pm 0^{\circ}.28$ and $0^{\circ}.76 \pm 0^{\circ}.29$, depending on the choice of M_2 and O_1 ocean tide models.

- 210 -
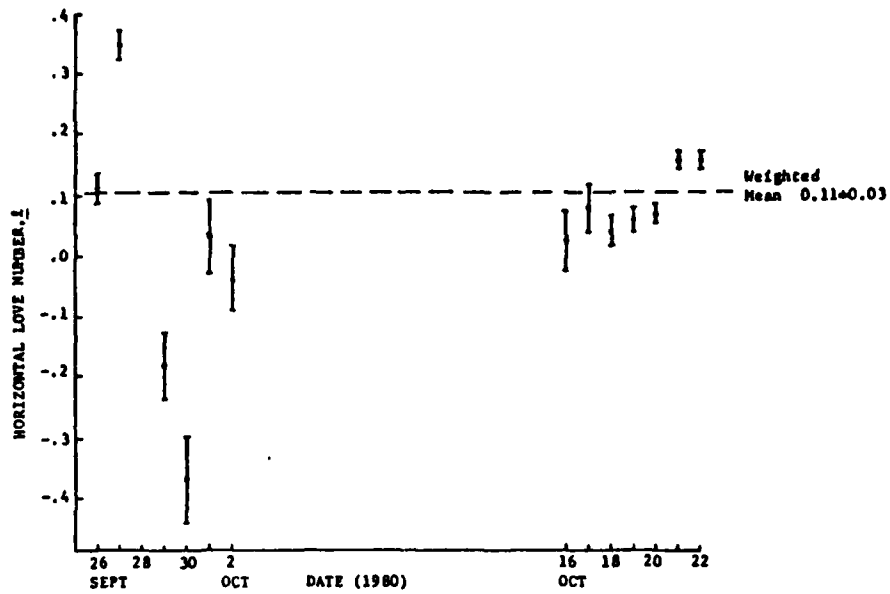
Figure 2. Independent estimates of the horizontal Love number, ℓ. (See Figure 1 for explanation of error bars.)
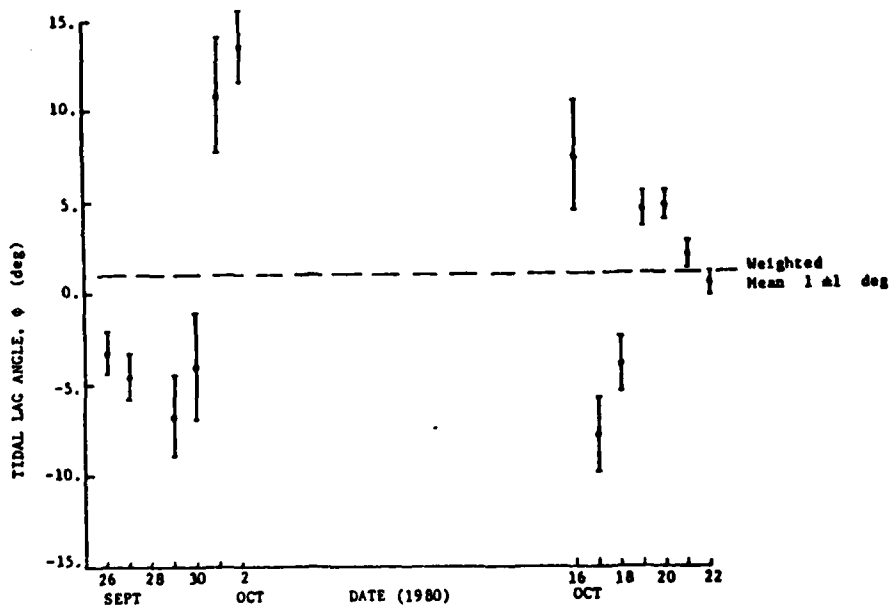
Figure 3. Independent estimates of the solid-earth tidal lag angle, φ. (See Figure 1 for explanation of error bars.)

ACKNOWLEDGEMENTS

REFERENCES

Cazenave, A., Daillet, S., 1981. Lunar tidal acceleration from
    earth satellite orbit analysis, J. of Geophys. Res., 86,(B3):
    1659-1663.
Herring, T.A., Corey, B.E., Counselman, C.C., Shapiro, I.I.,
    Rönnäng, B.O. Rydbeck, O.E.H., Clark, T.A., Coates, R.J., Ma, C.,
    Ryan, J.W., Vandenberg, N.R., Hinteregger, H.F., Knight, C.A.,
    Rogers, A.E.E., Whitney, A.R., Rogers, A.E.E., Whitney, A.R.,
    Robertson, D.S. & Schupler, B.R., 1981. Geodesy by radio inter-
    ferometry: Intercontinental distance determinations with sub-
    decimeter precision, J. Geophys. Res., 86: 1647-1651.
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. & Burke, B.F., 1972. Precision geodesy via radio interfer-
    ometry, Science, 178: 396-398.
Melchior, P., 1978. The Tides of the Planet Earth, pp 609, Perga-
    mon Press, New York.
Robertson, D.S., 1975. Geodetic and astrometric measurements with
    very long baseline interferometry. Ph.d. thesis, MIT
Robertson, D.S., Carter, W.E., Corey, B.E., Cotton, W.D., Counsel-
    man, C.C., Shapiro, I.I., Wittels, J.J., Hinteregger, H.F.,
    Knight, C.A., Rogers, A.E.E., Whitney, A.R., Ryan, J.W., Clark,
    T.A., Coates, R.J., Ma, C. & Moran, J.M., 1979. Recent results
    of radio interferometric determinations of a transcontinental
    baseline, polar motion, and Earth rotation, in McCarthy &
    Pilkington, eds., Time and the Earth's Rotation, pp 217-224,
    Dordrecht.
Ryan, J.W., et al., 1981. in preparation.
Whitney, A.R., Rogers, A.E.E., Hinteregger, H.F., Knight, C.A.,
    Levine, J.I., Lippincott, S., Clark, T.A., Shaprio, I.I. &
    Robertson, D.S., 1976. A very long baseline interferometer
    system for geodetic applications, Radio Science, 11: 421-432.

6. CONCLUSIONS AND DISCUSSION

The Mark III VLBI system is potentially capable of measuring intercontinental distances with one centimeter precision. In this thesis, we have tried to determine the proximity of the Mark III system to this goal.

The analysis of the algorithms used to determine the group delay, the phase-delay rate, and the visibility phase, from the signals recorded at the radio telescopes, did not reveal any large errors in these algorithms. However, when a series of tests was performed on the VLBI measurements, we found that the apparent noise in these measurements did not have the statistical properties which were calculated from the signal-to-noise-ratio (SNR) of the observations. There were clearly instrumental deficiencies.

Analysis of the phase versus frequency response of the Mark III showed that the system has large dispersions. The origin of the dispersion has not yet been determined, but it is hoped that the properties of the dispersion given in this thesis will aid in isolating the origin. The main properties of the dispersion are (i) it (the dispersion) varies rapidly with frequency (variations of up to 40° have been observed over a 4 MHz change in frequency, see Figure 3.1.6.); (ii) it can be very large; (iii) it varies in amplitude smoothly on time scales of months, and occasionally shows discontinuities, see Figure 3.1.10. One possible cause for the dispersion which seems to match much of the available data is a

corruption of the phase calibration pulses. In particular, the possibility that extra pulses from the pulse generator, which should be removed (or gated) from the phase calibration signals, are being injected into the receiver seems to require further investigation. There are also other possible causes. Among these, "ripple" in the video convertors, i.e., non-linear variations of phase with frequency across the video converter bandwidth, could explain the discontinuity which was seen in the Westford-Ft. Davis interferometer residual phases.

Other quality checks indicated that not all of the apparent errors in the group delays can be explained by phase calibration system errors or video converter ripple. The group delay closure errors did not match the SNR statistics. These closure errors seem to be explained by the leakage of left-circularly polarized radiation into the output of the nominally right-circularly polarized feed horns.

Probably the most powerful test we applied to the Mark III group-delay measurements was the investigation of the prediction of phase-delay ambiguities from the group-delay measurements. These studies revealed that the group-delay measurements seem to drift relative to the phase delays. The application of the phase-delay ambiguity prediction techniques to observations from a geodetic experiment yielded results which indicated that the systematic variations of the group delays relative to the phase delays, was mostly <u>independent</u> of the radio source being observed and the orientation of the radio telescope. These results indicate that the group delay

errors probably arise from the Mark III equipment, rather than
from external sources, such as antenna deformation and source
structure. These latter two effects do affect the apparent
group delay errors, but they to not appear to be the major
source of these errors.

A compilation of the statistics of the implied group
delay errors i.e., the differences between the observed values
of the group delays and the estimates of these values calcu-
lated from the phase delays, indicate that the standard devia-
tions of the group delays, computed from the assumption of
Gaussianly distributed random noise, are deficient in two
respects. Firstly, there appears to be a proportional error
in the computed standard deviations of between 1.1 and 1.2,
which is probably due to correlator non-reproducibility and
the dispersion in the system. Secondly, there appears to be a
threshold performance limit which cannot be penetrated, even
with very high SNR observations. At X-band ($\approx$8.34 GHz), this
limit was 0.040 nsec ($\approx$1.2 cm) and 0.015 nsec ($\approx$0.5 cm), for
the two experiments analyzed (see Figure 4.3.8). The S-band
($\approx$2.3 GHz) system has a performance limit of 0.15 nsec ($\approx$4.5
cm) for both of the experiments analyzed. The difference in
the performance of the X- and S-band systems can be partially,
but not totally (see Table 3.1.1), explained by the difference
in the spanned bandwidths in these two frequency bands.
(Typically 300 to 350 MHz is spanned at X-band while only
75-80 MHz is spanned at S-band.)

We also presented results which show intercontinental

distance measurements with a repeatablity of 2.0 cm, from 40 experiments spanning 3 years, after a "best-fit" linear trend was removed from the data. Studies of the solutions which generated these results show that the accuracy of the length estimate, from the ensemble of experiments, is at the several centimeter level rather than at the sub-centimeter level which is implied by the repeatability and the number of experiments in the ensemble.

Studies of the the dependencies of the baseline length estimates on the minimum elevation angle of observations included in the data set, have indicated that the relationship we used between the elevation angle of the observation and the airmass through which the signal propagated, may be systematically biased.

We need to determine "realistic" standard deviations of the baseline length estimates. In nearly all cases the weighted-root-mean-square (WRMS) scatter of the baseline length estimates about the global solution values were 1.5 to 2.0 times the amount expected from their standard deviations. There are two reasons for this under estimation of the standard deviations of the baseline length estimates. Firstly, unmodelled or inadequately modelled effects such as water vapor delay, radio source structure, and ocean loading will contribute to the variability of the baseline length estimates. These variations are not totally accounted for in the least squares estimates of the standard deviations (to some extent they are accounted for through the reweighting of

the observations). Secondly, there are probably correlations between the pseudo-noise contributions to the observations from site dependent (unmodelled) systematic errors. We currently do not account for these correlations in the least-squares solutions, which will cause the uncertainities of the parameter estimates to be too small.

VLBI data acquisition and analysis has progressed considerably in the past 14 years. The precision of baseline length determinations has decreased from the meter level to a few centimeters. The factors that have limited the precision of the baseline length estimates has also changed considerably over this period of time. Initially, the equipment was the major limitation (Whitney, 1974). Before the introduction of the Mark III dual-frequency band system, the ionospheric delay was probably the major limiting error source (Robertson, 1975). Currently, it is not clear which deficencies in the data analysis or the equipment are the major error sources. To considerably improve upon the current precision will probably require the solution of a number of distinct problems. The delay due to atmospheric water vapor is likely to be the most serious error source at the moment. Recent improvements in the water vapor radiometer data acquisition techniques may allow this delay to be accurately determined. However, there are other sources of errors which are probably important, and also need to be studied. These error sources include frequency standard instabilities, instrumental errors, radio source brightness distribution, oceaning loading, the dry

atmospheric delay and antenna deformations.

In 14 years, the precision of VLBI baseline length deter-
minations has improved by almost two orders of magnitude.
With further study and development, we should be able to
improve the precision by another factor of three in the next
14 years.

Appendix A. Relationship between the delay resolution function and least squares.

The delay resolution function (Rogers, 1970) is one of the most important functions in geodetic VLBI. Maximization of this function provides a means of estimating the group delay, phase delay and phase delay rate. (These will be refer- red to as the VLBI observables.) As we approach being able to measure intercontinental distances with 1 cm precision, it becomes critical that we fully understand the interaction between these estimation algorithms and the VLBI instrument- ation and environment. In this appendix, the relationship between the usual method of maximization of the delay resolu- tion function and a least-squares approach to the estimation problem will be described. We introduce the least-squares approach because it allows quantitative determination of the effects of phase errors, dispersions, and data loss on the VLBI observables.

Before developing the equivalence of the maximization of the delay resolution function and the least squares estimator, we review the concepts of group delay, phase delay and phase delay rate. We then review the development of the delay resolution function which is shown to be a maximum likelihood estimator, highlighting the assumptions made during the deri- vation. Finally, before showing the equivalence, we consider some of the practical details of computing the maximum likeli- hood or least squares estimates of VLBI observables.

The maximum likelihood estimator for the VLBI observables is given by Rogers (1970) and developed in detail by Whitney (1974). The review given here will follow closely the development of Whitney with some modifications.

Two radio telescopes are assumed to be observing an emitter of electromagnetic energy at radio frequencies. The output of noiseless receivers, in noiseless surroundings, at the two telescopes will be given by $\sqrt{Ta_1}\ s_1(t)$ and $\sqrt{Ta_2}\ s_2(t)$ where Ta_1 and Ta_2 are the antenna temperatures at each site and the dimensionless functions $s_1(t)$ and $s_2(t)$ are assumed to be white noise with zero mean and unit variance. The antenna temperatures are simply a convenient method of representing the power being received by each antenna from the radio source. In terms of the source strength S_o (given in Janskys, where 1 Jansky $= 10^{-26}$ w/m^2-Hz), the antenna temperature for observations in one polarization from a broadband, unpolarized source will be (Krauss, 1966)

$$Ta = A_e\ S_o\ /\ 2k \tag{A.1}$$

where A_e is the effective area of the antenna and k is Boltzmann's constant (1.38×10^{-23} Joule/K).

For the purposes of this section it is more convenient to use the spectra of $s_1(t)$ and $s_2(t)$ which will be denoted by $\bar{S}_1(\omega)$ and $\bar{S}_2(\omega)$. The vector symbol denotes complex quantities. We assume that real and imaginary components of the spectra are independent and Gaussianly distributed. While these may seem somewhat arbitrary assumptions, the central limit theorem of probability ensures that the sum of many

small independent sources of energy will indeed be close to Gaussianly distributed, independent of the original probability distributions obeyed by the individual energy sources (see Section 2.1). We further assume that the spectral components at each sampled frequency ω are independent of each other. There are two aspects to this assumption. Since we are only able to observe the signal using a finite bandwidth B, and for a finite duration T, the Nyquist sampling theorem states that, provided we sample the signal at the Nyquist rate (1/2B), there are only BT independent complex frequency components of the spectra which can be determined from the data. The processing algorithms used in the correlator ensure that the correct number of cross spectral components are computed (although as we will see later, these components are not computed at the minimum frequency spacing determined by the duration of the data). The recording equipment is set to sample the data at the correct rate. The second aspect of this assumption is that the original spectra must have independent components at the frequencies which we are able to sample. If this assumption is to be valid then the phases of the spectral components of the signals should vary randomly between frequencies. (We are considering the spectrum of the signal not the cross spectrum of the data recorded at each site. The cross spectrum phase at each frequency will be related to the difference in arrival time of the signals at each site.) Since the signal spectrum is the sum of the spectra of the emission from many electrons (at least 10^{47}

electrons; see Section 2.1), the phase of this spectrum should vary randomly.

The signals from which $\bar{S}_1(\omega)$ and $\bar{S}_2(\omega)$ are determined originate from a common source and hence should be related if these signals are spatially coherent on scales of several thousand kilometers. Thus, we assume that for each frequency we can define a wavefront such that the phase of any spectral component of the signal remains constant as we move along this wavefront. Since the properties of the propagation medium between any point on the wavefront and the source may change as we move along the wavefront, the wavefront is not necessarily plane even for sources an infinite distance from the radio telescopes.

The observed spectra are $\sqrt{Ta_1}\,\bar{S}_1(\omega)$ and $\sqrt{Ta_2}\,\bar{S}_2(\omega)$ and the relationship between these spectra can be expressed as $\bar{S}_2(\omega) = \bar{S}_1(\omega)\,e^{-i\phi_\omega}$ where ϕ_ω is a phase change which accounts for the difference in arrival times of the signal at the two sites (we will neglect, for the moment, that ϕ_ω is a function of time because of the rotation of the Earth). This phase change in the general case is given by

$$\phi_\omega = \int_0^{\bar{r}_2} \bar{k}_\omega(\bar{r}) \cdot d\bar{r} \;-\; \int_0^{\bar{r}_1} \bar{k}_\omega(\bar{r}) \cdot d\bar{r} \qquad (A.2)$$

where the integration is along the ray path, $\bar{k}_\omega(\bar{r})$ is the wave vector (see, e.g., Jackson, 1975, Chapter 7) which may be a function of position and frequency, and $\bar{r}_1$ and $\bar{r}_2$ are the position vectors of the receivers with respect to the source.

If we assume $\bar{k}_\omega(\bar{r})$ is independent of position, then Equation (A.2) reduces to $\phi_\omega = \bar{k}_\omega \cdot \Delta\bar{r}$ where $\Delta\bar{r} = \bar{r}_2 - \bar{r}_1$.

We now consider a simple case where the radio waves have been emitted from a point source, an infinite distance from the radio telescopes, and have propagated through a homogeneous nondispersive medium. For this case the wave vector $\bar{k}_\omega$ will be given by

$$\bar{k}_\omega = -\frac{\omega}{v_p}\,\hat{e}_s$$

where v_p is the phase velocity of the propagation (independent of frequency in this case, by definition) and $\hat{e}_s$ is a unit vector in the direction from the receiver to the point source. The phase delay, τ_p, will be $-\hat{e}_s \cdot \Delta\bar{r}/v_p$ and hence the phase difference, ϕ_ω, between corresponding spectral components becomes $\phi_\omega = \omega\tau_p$. We could use this formula directly to find the phase delay by comparing the spectra of the signals received at two sites and determining the difference in phase between the corresponding spectral components. An alternative (and equivalent) approach would be to multiply the spectra together to form the cross spectrum. The phase of the cross spectrum (neglecting the noise contribution) at frequency ω would be ϕ_ω. However, we are only able to measure ϕ_ω modulo 2π radians. We have no direct means of determining the total number of cycles through which the phase has turned as the signal propagated from the source to each of the sites. If we were to use <u>a priori</u> information about the magnitude of τ_p to resolve the number of cycles or ambiguities, the initial

estimate of τ_p would need to be sufficiently accurate that we could predict ϕ_ω to within much less than π radians. For observations made at 8 GHz, this condition requires the error in the $\underline{a\ priori}$ value of τ_p to be much less than 0.06 nsec, equivalent to ≈ 2 cm. In general, this condition cannot be met. Chapter 4 deals with some attempts to eliminate this ambiguity.

To overcome the ambiguity problems with the phase delays, the bandwidth synthesis technique was developed (Rogers, 1970). The foundation of this technique is that while the total number of cycles at a given frequency may not be known, the number of cycles between two relatively close frequencies may be calculated, with uncertainty far less than half a cycle, from $\underline{a\ priori}$ information. The delay is then found from $\Delta\phi/\Delta\omega$. As $\Delta\omega$ tends to 0, $\Delta\phi/\Delta\omega$ approaches $\partial\phi/\partial\omega$ which is the group delay.

Using the group delay, we could model the observed phase with $\phi_\omega = \omega\tau_g$, but this expression is inadequate for two reasons. If the radio waves have propagated through different dispersive regions (e.g. the ionosphere) before arriving at each site, then the phase delay and group delay will not be equal (see Section 4.1) and the projection of the group delay to zero frequency will not pass through the origin (Figure A.1.a), a condition imposed by the above form of ϕ_ω. Even without dispersive effects, the uncertainty in the number of cycles and the value of the phase at each frequency will lead to a similar problem (Figure A.1.b). To avoid these problems

Figure A.1 Graphical representations of group delay and visibility phase

a. Dispersive medium

The slope $\Delta\phi/\Delta\omega$ is the estimate of the group delay

The intercept at ω_o is the visibility phase, ϕ_t

The phase delay is $(\phi_t+2\pi N)/\omega_o$

$\phi_t+2\pi N$

dispersion curve

frequency ω

ϕ_o

Projection of the group delay to zero frequency

b. Ambiguity uncertainity

integer cycle (2π in phase)

$\phi_t+2\pi N$
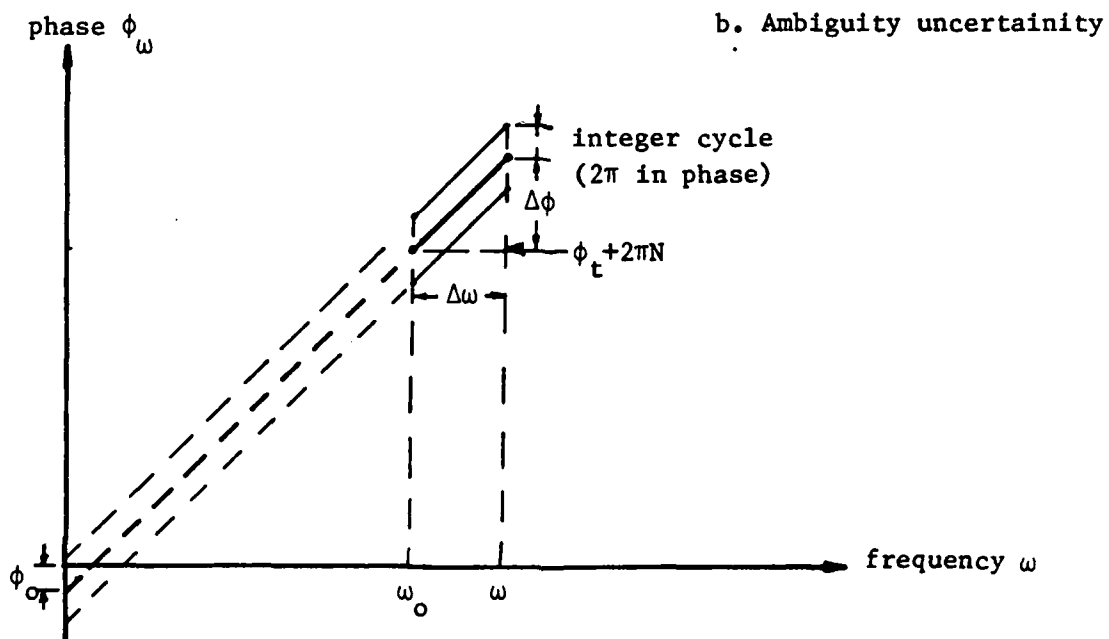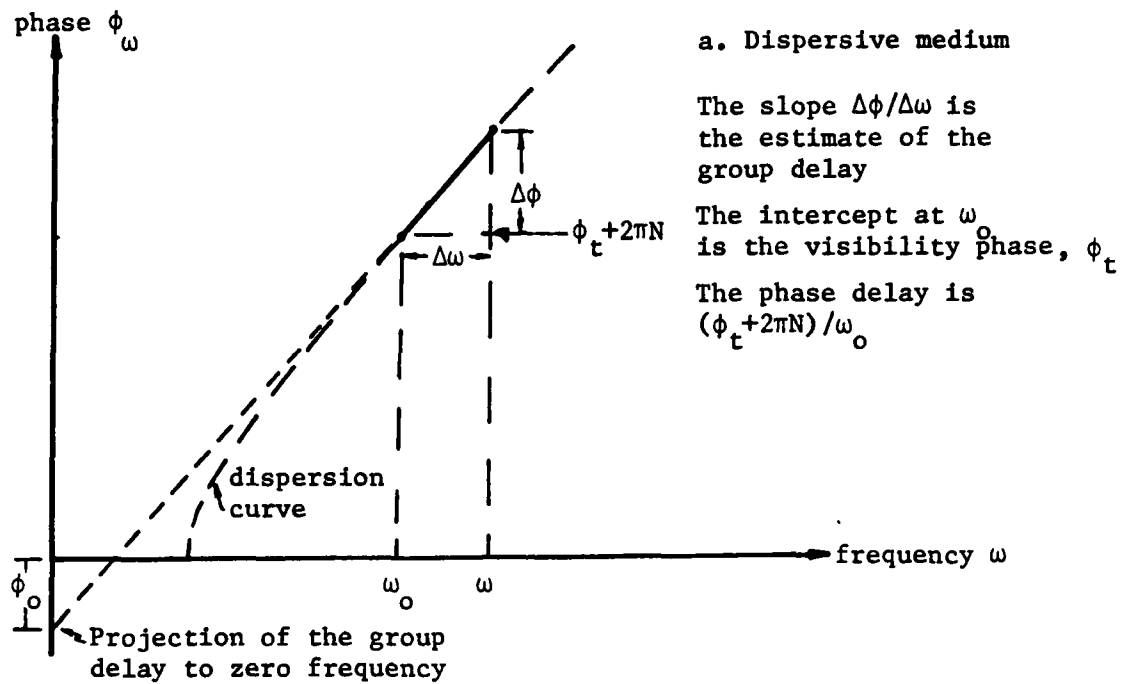
frequency ω

we model the observed phase using

$$\phi_\omega = \omega\tau_g + \phi_o$$

where ϕ_o is a phase offset referred to zero frequency.

To more readily see the meaning of ϕ_o in the above expression, we can write an equivalent expression for the observed phase: $\phi_\omega = (\omega-\omega_o)\tau_g + \phi_t$, where ϕ_t is the phase referred to frequency ω_o. This phase is called the visibilty phase. ϕ_t is now related to the phase delay by $\omega_o\tau_p = \phi_t + 2\pi N$, where N is the number of integer cycles of phase needed to achieve equality.

The discussion so far has assumed that the differences in phase between the corresponding spectral components of the signal recorded at two sites are independent of time. In reality, the two sites are located on a rotating earth and the phase differences will change with time. To account for this effect we introduce the phase delay rate $\dot{\tau}_p$. The phase can now be modeled as

$$\phi_\omega = (\omega-\omega_o)\tau_g + \phi_t + \omega\dot{\tau}_p(t-t_c) \qquad (A.3)$$

where t_c is the epoch of the observation (i.e. the time to which τ_g and ϕ_t are referred) and $\underline{t}$ is the epoch of the phase determination. In this case we can use the phase delay rate directly, without resorting to a group delay rate, because we are only interested in the change of phase with time and do not need to know the total number of ambiguities at any time.

In summary, we started with the spectra of signals at

each site which were related by

$$\bar{S}_2(\omega) = \bar{S}_1(\omega) \, e^{-i\phi}\omega$$

and now we have a relationship between ϕ_ω and the VLBI obser-
vables (Equation A.3) for the case of the radio waves origin-
ating from a point source and propagating through a homogene-
ous non-dispersive region. In terms of the VLBI observables
the relationship between the spectra becomes

$$\bar{S}_2(\omega) = \bar{S}_1(\omega) \exp[-i(\omega-\omega_0)\tau_g - i\phi_t - i\omega\dot{t}_p\Delta t] \qquad (A.4)$$

where $\Delta t = t - t_c$.

We now investigate how to determine τ_g, ϕ_t and $\dot{t}_p$ from
the signals recorded at each site, $\sqrt{Ta_1}s_1(t)$ and $\sqrt{Ta_2}s_2(t)$.

The data recorded at each site consist not only of
$\sqrt{Ta_1}s_1(t)$ and $\sqrt{Ta_2}s_2(t)$ but also of additive noise due mostly
to the receivers. Returning to the frequency domain, we can
write the spectra of the data recorded at each of two sites as

$$\bar{X}_1(\omega) = \sqrt{Ta_1}\bar{S}_1(\omega) + \sqrt{Ts_1}\bar{N}_1(\omega)$$

$$\bar{X}_2(\omega) = \sqrt{Ta_2}\bar{S}_2(\omega) + \sqrt{Ts_2}\bar{N}_2(\omega)$$

where $\sqrt{Ts_1}N_1(\omega)$ and $\sqrt{Ts_2}N_2(\omega)$ are the noise spectra at each
site, and Ts_1 and Ts_2 are the "off-source" system tempera-
tures. As with the antenna temperature, the system tempera-
ture is a convenient method of measuring the noise power. We
assume the components of the noise spectra have Gaussianly
distributed, independent real and imaginary components. The
components at each frequency are also assumed to be indepen-
dent. These assumptions can be justified by the same argu-

ments used for the spectra from the source.

Before progressing further it is convenient to change notation to reduce the number of symbols used in subsequent equations. We define

$$\bar{S}(\omega) \equiv \sqrt{Ta_1} \; \bar{S}_1(\omega)$$

$$K\bar{S}'(\omega) \equiv \sqrt{Ta_2} \; \bar{S}_2(\omega)$$

where $K = \sqrt{Ta_2/Ta_1}$. The scaling K ensures that the components of $\bar{S}(\omega)$ and $\bar{S}'(\omega)$ have the same variance, which is equal to $Ta_1/2$. The noise spectra are similarly redefined

$$\bar{N}_1' \equiv \sqrt{Ts_1} \; \bar{N}_1(\omega)$$

$$\bar{N}_2' \equiv \sqrt{Ts_2} \; \bar{N}_2(\omega) \quad .$$

Also for convenience we will define

$$\alpha \equiv (\omega - \omega_0)\tau_g + \phi_t + \omega t_p \Delta t \; .$$

Using the new symbols the expression for the spectra of the recorded data at each site will be

$$\bar{X}_1(\omega) = \bar{S}(\omega) + \bar{N}_1'(\omega)$$

$$\bar{X}_2(\omega) = K\bar{S}'(\omega) + \bar{N}_2'(\omega) \tag{A.5}$$

and the relationship between spectra will be $\bar{S}'(\omega) = \bar{S}(\omega)e^{-i\alpha}$. All of the spectra given in Equation (A.5) vary with time. However, this time dependence does not mean that we cannot determine estimates of the above spectra. Any finite duration or data can always be expanded in Fourier series. In a later part of this appendix and in Appendix B we will discuss the determination of these spectra. (We will see later, that the spectra of the signals recorded at each site never need be computed, only the cross spectrum of the signals

is computed.)

We now wish to develop an estimation technique to deter-
mine τ_g, ϕ_t and t_p given the spectra $\bar{X}_1(\omega)$ and $\bar{X}_2(\omega)$ which we
can determine at a number of different frequencies and times.
For an observation of recorded bandwidth B and duration T, a
maximum of BT independent estimates of the components of $\bar{X}_1(\omega)$
and $\bar{X}_2(\omega)$ can be calculated, assuming that the signals have
been sampled at the Nyquist rate 1/2B. If we observe for 100
seconds with a total recorded bandwidth of 16 MHz, then the
number of independent estimates of the components of the
spectra is 1.6×10^9 (a very large number).

There are many estimation techniques we could use to find
estimates of τ_g, ϕ_t and $\hat{\tau}_p$. Of these we will develop the
Maximum Likelihood (ML) estimate. This technique requires a
priori knowledge of the joint probablity density function
(PDF) of the observations (in our case $\bar{X}_1(\omega)$ and $\bar{X}_2(\omega)$) which
are functions of the VLBI observables. The ML estimates of
τ_g, ϕ_t and t_p, which will be denoted by $\hat{\tau}_g$, $\hat{\phi}_t$ and $\hat{t}_p$, are
estimated by maximizing the conditional PDF $p(\bar{X}_1,\bar{X}_2|\tilde{\tau}_g,\tilde{\phi}_t,\tilde{t}_p)$
where $\tilde{\tau}_g$, $\tilde{\phi}_t$ and $\tilde{t}_p$ are the values to be varied until the
conditional PDF is maximized. The values of $\tilde{\tau}_g$, $\tilde{\phi}_t$ and $\tilde{t}_p$
which maximize the conditional PDF are, as stated, the maximum
likelihood estimates $\hat{\tau}_g$, $\hat{\phi}_t$ and $\hat{t}_p$.

At this stage the conditional PDF we need is not avail-
able. It will have to be derived. We commence with the PDF's
which we do know (or strictly, have assumed). From the
recorded data we have BT spectral components. For each of

these components the signal and noise values will be denoted
by $\bar{S}_j$, $\bar{N}'_{1j}$ and $\bar{N}'_{2j}$. (The index j implicitly implies that these
components are functions of both frequency and time; see
Section 2.2 and Appendix B for discussion.) Since we have
assumed Gaussian distributions for the real and imaginary
parts of the signal and noise components, the individual PDF's
will be

$$p(\bar{S}_j \ ) = \frac{1}{2\pi(\sigma_s)_j^2} \ \exp(\ -|\bar{S}_j|^2/\ 2(\sigma_s)_j^2)$$

$$p(\bar{N}_{1j}) = \frac{1}{2\pi(\sigma_{n1})_j^2} \ \exp(\ -|\bar{N}_{1j}|^2/\ 2(\sigma_{n1})_j^2)$$

$$p(\bar{N}_{2j}) = \frac{1}{2\pi(\sigma_{n2})_j^2} \ \exp(\ -|\bar{N}_{2j}|^2/\ 2(\sigma_{n2})_j^2)$$

where $(\sigma_s)_j^2$, $(\sigma_{n1})_j^2$ and $(\sigma_{n2})_j^2$ are the variances of each of
the real and the imaginary components of the respective
signals and noises.

Since all the signals and noises have independent compon-
ents we can form their joint PDF, $p(\bar{S},\bar{N}'_1,\bar{N}'_2)$, which is given
by the product of their individual PDF's.

$$p(\bar{S},\bar{N}'_1,\bar{N}'_2) = \prod_{j=1}^{BT} \frac{1}{(2\pi)^3(\sigma_s)_j^2(\sigma_{n1})_j^2(\sigma_{n2})_j^2}$$

$$\exp[\ -|\bar{S}_j|^2/2(\sigma_s)_j^2- \ |\bar{N}_{1j}|^2/2(\sigma_{n1})_j^2 \ - \ |\bar{N}_{2j}|^2/2(\sigma_{n2})_j^2]$$

Remembering that the observed spectra are $\bar{X}_1$ and $\bar{X}_2$, we can
substitute for $\bar{N}'_1$ and $\bar{N}'_2$ in the above PDF using $\bar{N}'_1 = \bar{X}_1 - \bar{S}$
and $\bar{N}'_2 = \bar{X}_2 - K\bar{S}\exp(-i\tilde{\alpha}_j)$ where $\tilde{\alpha}_j$ is defined to be $(\omega-\omega_o)\tilde{\tau}_g +$

$\tilde{\phi}_t + \omega \tilde{\tau}_p \Delta t_j$. Hence we obtain

$$p(\bar{X}_1, \bar{X}_2, \bar{S} \mid \tilde{\tau}_g, \tilde{\phi}_t, \tilde{\tau}_p) = \prod_{j=1}^{BT} \frac{1}{(2\pi)^3 (\sigma_s)_j^2 (\sigma_{n1})_j^2 (\sigma_{n2})_j^2} \cdot$$

$$\exp\left[-\frac{|\bar{S}_j|^2}{2(\sigma_s)_j^2} - \frac{|\bar{X}_{1j} - \bar{S}_j|^2}{2(\sigma_{n1})_j^2} - \frac{|\bar{X}_{2j} - K\bar{S}_j e^{-i\tilde{\alpha}}|^2}{2(\sigma_{n2})_j^2} \right] \quad \text{(A.6)}$$

The above expression is not directly usable because it still depends on the spectrum $\bar{S}$ of the signal which can not be directly observed (because of the presense of the noise). The desired conditional PDF can be generated from Equation (A.6) by integrating over all possible values of $\bar{S}$ to form a marginal PDF (see e.g. Drake, 1967, pp. 69-72, for a discussion on generating marginal PDF's):

$$p(\bar{X}_1, \bar{X}_2 \mid \tilde{\tau}_g, \tilde{\phi}_t, \tilde{\tau}_p) = \int_{-\infty}^{\infty} p(\bar{X}_1, \bar{X}_2, \bar{S} \mid \tilde{\tau}_g, \tilde{\phi}_t, \tilde{\tau}_p) \, d\bar{S}$$

The details of carrying out this integration are given by Whitney (1974, pp. 64-65). The result is

$$p(\bar{X}_1, \bar{X}_2 \mid \tilde{\tau}_g, \tilde{\phi}_t, \tilde{\tau}_p) = \prod_{j=1}^{BT} \frac{\sigma_j^2}{(2\pi)^2 (\sigma_s)_j^2 (\sigma_{n1})_j^2 (\sigma_{n2})_j^2} \cdot$$

$$\exp\left[-\frac{|\bar{X}_{1j}|^2}{2(\sigma_{n1})_j^2} - \frac{|\bar{X}_{2j}|^2}{2(\sigma_{n2})_j^2} + \frac{|\bar{Y}_j|^2}{2\sigma_j^2} \right] \quad \text{(A.7)}$$

where

$$\sigma_j^2 \equiv \frac{(\sigma_s)_j^2 (\sigma_{n1})_j^2 (\sigma_{n2})_j^2}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2 + (\sigma_s)_j^2 (\sigma_{n2})_j^2 + K(\sigma_s)_j^2 (\sigma_{n1})_j^2} \quad ,$$

$$\bar{Y} \equiv \frac{\bar{X}_{1j}^* \sigma_j^2}{(\sigma_{n1})_j^2} + \frac{K\bar{X}_{2j}^* e^{-i\tilde{\alpha}_j} \sigma_j^2}{(\sigma_{n2})_j^2} \quad ,$$

and $\bar{X}_{ij}^*$ is the complex conjugate of $\bar{X}_{ij}$, $i=1,2$.

Although Equation (A.7) looks very complicated (and not very close to a usable form), its interpretation is quite simple. Remembering that $\bar{X}_{1j}$ and $\bar{X}_{2j}$ are the observed quantities, we can choose a set of trial values for $\tilde{\tau}_g$, $\tilde{\phi}_t$ and $\tilde{\tau}_p$ which would determine the complex vector $\bar{Y}_j$ (assuming all the sigma's are known). We would then form the product over the BT values to determine $p(\bar{X}_1,\bar{X}_2|\tilde{\tau}_g,\tilde{\phi}_t,\tilde{\tau}_p)$. When this operation is performed for many sets of trial values of $\tilde{\tau}_g$, $\tilde{\phi}_t$ and $\tilde{\tau}_p$, the set giving the maximum value for $p(\bar{X}_1,\bar{X}_2|\tilde{\tau}_g,\tilde{\phi}_t,\tilde{\tau}_p)$ will be the maximum likelihood estimates $\hat{\tau}_g$, $\hat{\phi}_t$ and $\hat{\tau}_p$.

We can save a great deal of time in computing the product (A.7) by noting that $|X_{1j}|^2/2(\sigma_{n1})_j^2$ and $|X_{2j}|^2/2(\sigma_{n2})_j^2$ are not affected by our choice of $\tilde{\tau}_g$, $\tilde{\phi}_t$ and $\tilde{\tau}_p$ and hence will not affect the maximization process. Also, by expanding $|\bar{Y}_j|^2$ we obtain

$$|\bar{Y}_j|^2 = \frac{\bar{X}_{1j}^*\bar{X}_{1j}\sigma_j^4}{(\sigma_{n1})_j^2} + \frac{K\bar{X}_{2j}^*\bar{X}_{2j}\sigma_j^4}{(\sigma_{n2})_j^2} + \frac{2\sigma_j^4 K \ \mathrm{Re}\{\bar{X}_{1j}\bar{X}_{2j}^* e^{-i\alpha}j\}}{(\sigma_{n1})_j^2(\sigma_{n2})_j^2}$$

in which again the first two terms are not affected by the choice of $\tilde{\tau}_g$, $\tilde{\phi}_t$ and $\tilde{\tau}_p$. To find the maximum likelihood estimate it is thus sufficient to maximize

$$\prod_{j=1}^{BT} \frac{\sigma_j^2}{(\sigma_s)_j^2(\sigma_{n1})_j^2(\sigma_{n2})_j^2} \exp\left[\frac{\sigma_j^2 K}{(\sigma_{n1})_j^2(\sigma_{n2})_j^2} \ \mathrm{Re}\{\bar{X}_{1j}\bar{X}_{2j}^* e^{-i\alpha}j\}\right]$$

This expression may be further simplified by taking the natural log of the product, which yields a summation to be

maximized

$$\sum_{j=1}^{BT} \ln\left(\frac{\sigma_j^2}{(\sigma_s)_j^2 (\sigma_{n1})_j^2 (\sigma_{n2})_j^2}\right) + \sum_{j=1}^{BT} \frac{\sigma_j^2 K}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2} \, \text{Re}\{\bar{X}_{1j}\bar{X}_{2j}^* e^{-i\tilde{\alpha}_j}\}$$

The first summation will not affect the maximization and we are left with maximizing

$$\sum_{j=1}^{BT} \frac{\sigma_j^2 K}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2} \, \text{Re}\{\bar{X}_{1j}\bar{X}_{2j}^* e^{-i\tilde{\alpha}_j}\} \tag{A.8}$$

to find the maximum likelihood estimates of τ_g, ϕ_t and $\dot{\tau}_p$.

At this stage, it is normally assumed (Whitney, 1974) that the sigmas given in Equation (A.8) are independent of j and hence can be taken outside the summation reducing the expression to be maximized to

$$\sum_{j=1}^{BT} \text{Re}\{\bar{X}_{1j}\bar{X}_{2j}^* e^{-i\tilde{\alpha}_j}\} \tag{A.9}$$

Equation (A.9) is very close to the formula used by the Mark III correlator and processing software (FRNGE), to estimate the VLBI observables τ_g, ϕ_t and $\dot{\tau}_p$. If we examine Equation (A.9) more closely, we can derive the actual formula used by the Mark III processing programs. Remembering that $\tilde{\alpha}_j = (\omega_j-\omega_o)\tilde{\tau}_g + \tilde{\phi}_t + \omega_j\dot{\tilde{\tau}}_p\Delta t_j$, we see that Equation (A.9) may be rewritten as

$$\text{Re}\{e^{-i\tilde{\phi}_t} \sum_{j=1}^{BT} \bar{X}_{1j}\bar{X}_{2j}^* \exp[-i(\omega_j-\omega_o)\tilde{\tau}_g - i\omega_j\dot{\tilde{\tau}}_p\Delta t_j]\}$$

where $\tilde{\phi}_t$ has been taken outside the summation because it is not a function of the cross spectral frequencies. Equation

(A.9) can be maximized by maximizing the magitude of

$$\sum_{j=1}^{BT} \bar{X}_{1j}\bar{X}_{2j}^* \exp[\ -i(\omega_j-\omega_o)\tilde{\tau}_g\ -\ i\omega_j\tilde{t}_p\Delta t_j]$$
<div align="right">(A.10)</div>

and rotating the value of Equation (A.10), evaluated at $\hat{\tau}_g$ and $\hat{t}_p$, by $-\hat{\phi}_t$ such that it is then purely real (rotated back onto the real axis). Equation (A.10) when divided by BT is called the delay resolution function. Use of the delay resolution function instead of Equation (A.9) leads to a considerable saving in computation since it is only necessary to search over two parameters, $\tilde{\tau}_g$ and $\tilde{t}_p$, instead of the three parameters which would be needed if Equation (A.9) were used.

The maximization of Equation (A.8) represents the maximum likelihood estimator of the VLBI observables when the noise statistics are assumed to be Gaussian. Equation (A.10) which is used by the Mark III processing software makes a further assumption that the standard deviations of the Gaussian statistics are independent of frequency and time. The equivalence of the ML estimator and a least squares estimator will be shown for Equation (A.8). The relationship between the Mark III algorithms and the least squares estimator will then be shown.

The input for the delay resolution function is the cross spectral components $\bar{X}_{1j}X_{2j}^*$. The derivation of the delay resolution function assumed that these cross spectral components were available near the radio frequency (RF) for the observation. As discussed in Section 2.1, the signals recorded at each site have been heterodyned from RF to video fre-

quencies so that the signals can be sampled. The required input for the delay resolution function can be easily generated from the video signals because the heterodyning only translated the spectrum in frequency. (There were local oscillator phases added to the phases of the spectra from each site, but these local oscillator phases are calibrated by the phase calibration system; see Section 2.3.) In the discussion that follows we will compute the cross spectrum from data recorded at video frequencies. The frequencies of the cross spectrum will have the local oscillator frequencies added to them to produce the correct cross spectral frequency.

Before proceeding further, it is wise to look at some of the practical aspects of performing the summation in either Equation (A.8) or (A.10). In most Mark III geodetic experiments the recorded bandwidth is typically 16 MHz (consisting of eight, 2 MHz channels) and the observation of a single source is usually of 100 seconds duration. For these values, 3.2×10^9 equi-spaced data samples are recorded at each site participating in the observation. From 3.2×10^9 real data samples, 1.6×10^9 complex components of the cross spectrum can be calculated. The frequencies at which these components are evaluated are however not arbitrary. The frequency spacing will be the inverse of the duration of the observation. The highest frequency component which can be estimated will be the inverse of twice the sampling interval. In addition, since the observations are recorded at a number of widely separated frequencies the cross spectrum can only be computed at fre-

quencies within the bandwidths which have been sampled.
Therefore within each 2 MHz band which has been recorded we
can compute 2×10^8 complex components of the cross spectrum.
We restrict the following discussion to a single such band.

A cross spectrum computed directly from the recorded data
may not be very meaningful because of the rotation of the
earth. For each frequency the phase of the cross spectrum
will be a function of the difference in arrival times of the
signals from the radio source. As the earth rotates this
phase will change. If the difference in arrival times of the
signals were changing at a constant rate then the phases of
each frequency component would be changing at a constant rate
which would be equivalent to a change in the frequency of each
component in the cross spectrum. Even this situation does not
apply to geodetic VLBI because the acceleration of the differ-
ence in arrival times is not negligible. For a trans-conti-
nental baseline this acceleration term could change the fre-
quencies in the cross spectrum by up to 1 Hz/sec ($\omega_0 = 8$ GHz).
The resolution in frequency for a 100 sec observation would be
0.01 Hz and the change in frequency introduced by the accel-
eration would be a severe problem. In order to solve this
problem we compute the cross spectrum using shorter intervals
of time. Hence we subdivide the 100 sec observation into seg-
ments. From each segment of data we can obtain estimates of
the components of the cross spectrum and then add the cross
spectral estimates together accounting for both the Doppler
shift and the change in Doppler shift for each segment of

data. Since the duration has been shortened the frequency spacing of the cross spectral components which can be computed will be coarser than if we had used the full 100 sec. This is not a severe problem because the power spectra of the signals and noise are almost constant over the 2 MHz recorded bandwidth (see Section 2.1) and the expectation of the estimates of the components of the power spectra will not change rapidly between adjacent frequencies (the estimates themselves could change rapidly, however).

For a typical geodetic observation what should be the duration of the segments? A cursory consideration may imply that the segment length could be almost one second. If a one second segment of data were used to compute estimates of the cross spectral components the change in frequency of a cross spectral component would be 1 Hz during the interval of time used to compute the cross spectrum, and the frequency resolution would also be 1 Hz, i.e. there would be a "smearing" of the cross spectrum comparable to the frequency resolution. Unfortunately this duration for the segment is far too long. To see why this is so, we need to consider the methods used to compute the cross spectral components.

Two different methods could be used, either the Fourier transform of the cross correlation function of the data from two sites, or the product of the Fourier transforms of the two individual data sets. If we try cross correlating the data from two sites, a problem develops immediately because the delay between the signals is changing with time, and hence, as

we cross correlate the data, the peak in the cross correlation
will move and "smear" the peak out. The peak in the cross
correlation function will translate at the group delay rate
which can be as high as 2 μsec/sec. If we try to cross cor-
relate data for too long an interval, the data will decor-
relate because of this changing delay. The computation of the
Fourier transform of the data from each site does not, at
first, seem to suffer from this problem because the changing
delay will only shift the frequency of the spectrum. With
this approach the problem arises when the product of the
spectra from two sites is computed. Although the spectra are
computed at the same frequency, the frequencies of the spectra
are Doppler shifted with respect to each other. This shift in
frequency causes problems because the phase of the spectral
components from each station vary randomly with frequency.
(Even if we assume no noise, the random variation in phase
occurs because the signals from the radio source are them-
selves random quantities. It is only the phase of the cross
spectrum which varies smoothly with frequency.) This random
variation of phase with frequency requires that the proper
frequencies of each spectrum must be carefully matched when
the product of the spectra is formed to ensure that the
correct cross spectral phase is computed.

To choose an appropriate duration for a segment of data,
we need to consider how the actual cross spectrum changes
during the interval which is used to compute the estimates of
its components. The shorter the duration used the fewer the

number of frequency components we will be able to determine,
but the better the match between the expectation of the
computed cross spectrum for that epoch and the actual cross
spectrum. Since the phase of the signal part of the cross
spectrum changes smoothly with frequency and since its ampli-
tude is constant we do not require very fine frequency resolu-
tion.

We will use 4 μsec (16 data samples) as the duration of a
segment. This choice is based on the rapid decay of the cross
correlation function (see Appendix B). The use of the short
data segments is not the technique used in the Mark III cor-
relator, but this concept of short segments will provide a
convenient method for developing and studying the actual
methods used by the correlator to estimate the cross spectral
components. In Appendix B we will investigate, in detail, the
method used by the correlator and the approximations made in
developing the method. The 4 μsec segment allows eight
complex components of the cross spectrum to be computed.
These components are spaced at 0.25 MHz intervals over the 2
MHz bandwidth. The cross spectral components computed from
all of the data segments then form the input for the delay
resolution function. (For the sign convention adopted in
deriving the delay resolution function, $\bar{X}_1 \bar{X}_2^*$ would be the
complex conjugate of the cross spectrum computed from the
Fourier transform of the cross correlation function defined by
$\int x_1(t) x_2(t+\tau) dt$, where $x_1(t)$ and $x_2(t)$ are the data from each
site. $\bar{X}_1 \bar{X}_2^*$ can be generated from the Fourier transform of
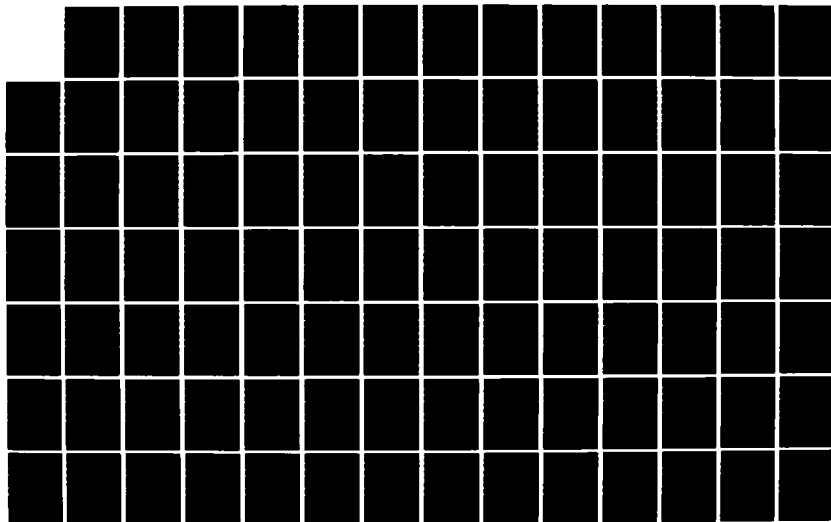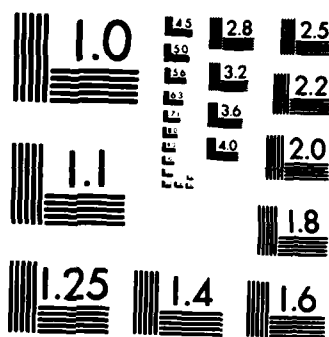
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

$\int x_1(t) x_2(t-\tau) dt)$.

For a given frequency we can consider the behavior of the cross-spectral phase with time. During the 4 μsec between the midpoints of successive intervals used for cross spectral determinations, the phase of the signal part will change by a maximum of $23°$ due to the rotation of the earth (corresponding to a phase delay rate of 2×10^{-6} sec/sec and an observing frequency of 8 GHz). In the current formulation, for each set of search parameters, we sum over all BT components of the cross spectrum. By improving the estimate of the phase delay rate, how much do we expect to improve the estimate of the phase change over 4 μsec? The a priori phase delay rate is normally known to better than 10^{-11} sec/sec and hence the improvement in the estimate of the phase change over 4 μsec would be less than $0°0001$. This small improvement in the value of the phase change would imply that we need only compute the change once (from a priori information) rather than for each set of search values. Actually the improvement is so small we could sum together several thousand estimates of the cross spectrum using just a priori information without introducing a significant amount of "phase smearing." This approach leads to major saving in computation time. We are, in effect, removing the Doppler shift between the spectra of the data from each site. The interval of time over which the cross spectral components from segments of data are summed together using only a priori information is called an accumulation period.

This presummation approach is used in VLBI data process-
ing but it is not implemented as I have described (i.e. by
summing cross spectral components). There are two methods of
computing a cross spectrum, either by finding the spectrum of
each data stream and multiplying the spectra together or by
cross correlating the data streams and determining the
spectrum of the cross correlation function. The Mark III
correlator uses the second approach. The presummation is
implemented by summing together the complex cross correlation
functions and then finding the spectrum of the summed cross
correlation function (see Appendix B for details). The
results obtained from these two techniques are **not** exactly
equal. Their differences will be discussed in Appendix B.

The Mark III correlator computes the complex cross cor-
relation coefficents for a set of eight lags. The cross
correlation coefficients from each accumulation period are
Fourier transformed and the maximum likelihood estimates of
the VLBI observables are determined by a program called FRNGE.

The presummation may be included in Equation (A.8) by
writing the equation as

$$\max_{\tilde{\alpha}} \quad \sum_{q=1}^{QJ} \text{Re}\{\bar{z}_q^a e^{-i\tilde{\alpha}_q}\} \qquad \qquad (A.11)$$

$$\text{where } \bar{z}_q^a \equiv \sum_{j=1}^{n(q)} \frac{\sigma_j^2 K}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2} \bar{X}_{1j} \bar{X}_{2j}^* e^{-i(\alpha_{ap})_j} \quad ,$$

$n(q)$ is the number of estimates of the cross spectral compo-
nents at each frequency in the q th accumulation period of
data, Q is the number of accumulation periods in the observa-

tion, J is the number of frequencies at which the cross
spectrum has been computed and $(\alpha_{ap})_j$ is the <u>a priori</u> value of
$(\omega_j - \omega_o)\tau_g + \phi_t + \omega_j \dot{t}_p \Delta t_j$. The multiplication of the cross
spectrum by $\exp(-i(\alpha_{ap})_j)$ is called counter rotation. The
$\tilde{\alpha}_q$'s now represent corrections to the <u>a priori</u> values of τ_g,
ϕ_t and $\dot{t}_p$. The summation over j is restricted to cross
spectral components at one frequency. (In Equation (A.11) we
are using one summation index ($\underline{q}$) to represent summations over
both frequency and time. The single summation index has been
chosen to simplify the equations.)

We have now developed all the formulas necessary to show
equivalence of the ML and least squares estimators. The ML
estimator is represented by Equation (A.11). We may rewrite
this equation as

$$\sum_{q=1}^{Q_J} \mathrm{Re}\{ r_q e^{i(\psi_q - \tilde{\alpha}_q)} \}$$

where r_q and ψ_q are the magnitude and argument of $\bar{z}_q^a$. From
each accumulation period we will obtain one estimate of this
magnitude and argument at each frequency in the cross spectrum
which has been evaluated. Taking the real component of the
above expression we obtain

$$\sum_{q=1}^{Q_J} r_q \cos(\psi_q - \tilde{\alpha}_q) \tag{A.12}$$

for the expression to be maximized. When Equation (A.12) is
near its maximum value $\tilde{\alpha}_q$ will approximately equal ψ_q and we

may expand (A.12) in series which yields

$$\sum_{q=1}^{Q_J} r_q (1 - (\psi_q - \bar{\alpha}_q)^2/2 + \text{higher order terms})$$

Clearly the above expression may be maximized by minimizing

$$\sum_{q=1}^{Q_J} r_q (\psi_q - \alpha_q)^2$$

But this minimization is simply the weighted least squares condition with the r_q's being interpreted as weights.

We would expect the r_q's to be inversely proportional to the variances of the phases of the cross spectrum ψ_q. To investigate this expectation we return to the form of $\bar{Z}_q^a$. From Equation (A.11) we have

$$|\bar{Z}_q^a| = \left| \sum_{j=1}^{n(q)} \frac{\sigma_j^2 K}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2} \bar{X}_{1j} \bar{X}_{2j}^* e^{-i(\alpha_{ap})_j} \right| \qquad \text{(A.13)}$$

We will investigate this equation in two parts, firstly the variance term and secondly the cross spectral term.

The variance term is $\sigma_j^2 K/((\sigma_{n1})_j^2 (\sigma_{n2})_j^2)$ where

$$\sigma_j^2 = \frac{(\sigma_s)_j^2 (\sigma_{n1})_j^2 (\sigma_{n2})_j^2}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2 + (\sigma_s)_j^2 (\sigma_{n2})_j^2 + K(\sigma_s)_j^2 (\sigma_{n1})_j^2}$$

and $(\sigma_s)_j^2$, $(\sigma_{n1})_j^2$ and $(\sigma_{n2})_j^2$ are the variances of the real or imaginary components of the spectra of the signal at the first site and the noise at each site. The constant K accounts for the difference in the amplitude of the signal at the output of the receiver between the first and the second site. These variances are given by $(\sigma_s)_j^2 = Ta_1/2$, $(\sigma_{n1})_j^2 = Ts_1/2$ and

$(\sigma_{n2})_j^2 = Ts_2/2$. K is equal to $\sqrt{Ta_2/Ta_1}$.

For nearly all VLBI observations, of nearly all extragalactic radio sources, the antenna temperatures are much less than the system temperatures. Most radio telescope, source combinations have Ta's < 1 K, whereas currently available wideband receivers have system temperatures between 100 K and 200 K. Even with modern cooled wideband receivers it is doubtful that system temperatures will drop much below 20 K. Assuming that Ta<<Ts the expression for σ_j^2 reduces to

$$\sigma_j^2 = (\sigma_s)_j^2 = Ta_1/2$$

Incorporating the other variance terms we obtain

$$\frac{\sigma_j^2 K}{(\sigma_{n1})_j^2 (\sigma_{n2})_j^2} = \frac{2\sqrt{Ta_1 Ta_2}}{Ts_1 Ts_2} \qquad (A.14)$$

The above term is considered independent of both j and q in the algorithms used by FRNGE.

To evaluate the remaining part of r_q it is convenient to make one further assumption. We assume that variance terms given above are constant over the duration of an accumulation period, i.e. for a period typically less than 2 seconds. We may then write

$$\left| \bar{Z}_q^a \right| = \left| \frac{\sigma_q^2 K}{(\sigma_{n1})_q^2 (\sigma_{n2})_q^2} \sum_{j=1}^{n(q)} \bar{X}_{1j} \bar{X}_{2j}^* e^{-i(\alpha_{ap})_j} \right| \qquad (A.15)$$

We will define the coherently averaged cross spectrum, $\bar{C}_q^a$, as

$$\bar{C}_q^a \equiv \sum_{j=1}^{n(q)} X_{1j}^* X_{2j} e^{i(\alpha_{ap})_j} \qquad (A.16)$$

We should note that C_q^a is the conjugate of the summation in Equation (A.15); see discussion after Equation (A.10).

From Equation (A.5) we have

$$\bar{X}_{1j}\bar{X}_{2j}^* \, e^{-i(\alpha_{ap})j} = K\bar{S}\bar{S}^* e^{i(\alpha_j - (\alpha_{ap})_j)} + \bar{S}\bar{N}_2^* e^{-i(\alpha_{ap})j} +$$

$$K\bar{S}^*\bar{N}_1 e^{i(\alpha_j - (\alpha_{ap})_j)} + \bar{N}_1\bar{N}_2^* e^{-i(\alpha_{ap})j}$$

$$(A.17)$$

The magnitudes of these terms are, respectively, $|K\bar{S}\bar{S}^*| = \sqrt{Ta_1 Ta_2}$, $|\bar{S}\bar{N}_2^*| = \sqrt{Ta_1 Ts_2}$, $|K\bar{S}^*\bar{N}_1| = \sqrt{Ta_2 Ts_1}$ and $|\bar{N}_1\bar{N}_2^*| = \sqrt{Ts_1 Ts_2}$. If $\dot{\tau}_p \simeq \dot{\tau}_{ap}$, then as we sum the n(q) values in the accumulation period, the signal term $K\bar{S}\bar{S}^* \exp[i(\alpha_j - (\alpha_{ap})_j)]$ will have constant phase and the net signal magnitude will be $n(q)\sqrt{Ta_1 Ta_2}$. The noise terms on the other hand will have random phases for each value in the summation and the contribution of the noise components will increase only as the $\sqrt{n(q)}$. The relationship between the terms in the coherently averaged cross spectra is represented graphically in Figure A.2. The variance of the amplitude of the noise terms will be Rayleigh distributed (Whitney, 1974, Section 3.2). When the real and imaginary components of the noise terms have equal variance (σ^2), the variance of their amplitude will be $(2-\pi/2)\sigma^2$.

From this figure we can also estimate the uncertainty in the phase of the cross spectrum. Since Ta<<Ts, the dominant noise term will be due to the product of the system noise terms for the various sites. Further, assuming that n(q) is sufficiently large that $\sqrt{Ta_1 Ta_2} \gg \sqrt{Ts_1 Ts_2}/n(q)$, we may

Figure A.2 Representation of the components of the cross spectrum

$$\bar{C}_q^a = \sum_{j=1}^{n(q)} \bar{X}_{1j}\bar{X}_{2j}^* e^{-i\alpha_{ap_j}}$$

= the coherently averaged
cross spectrum
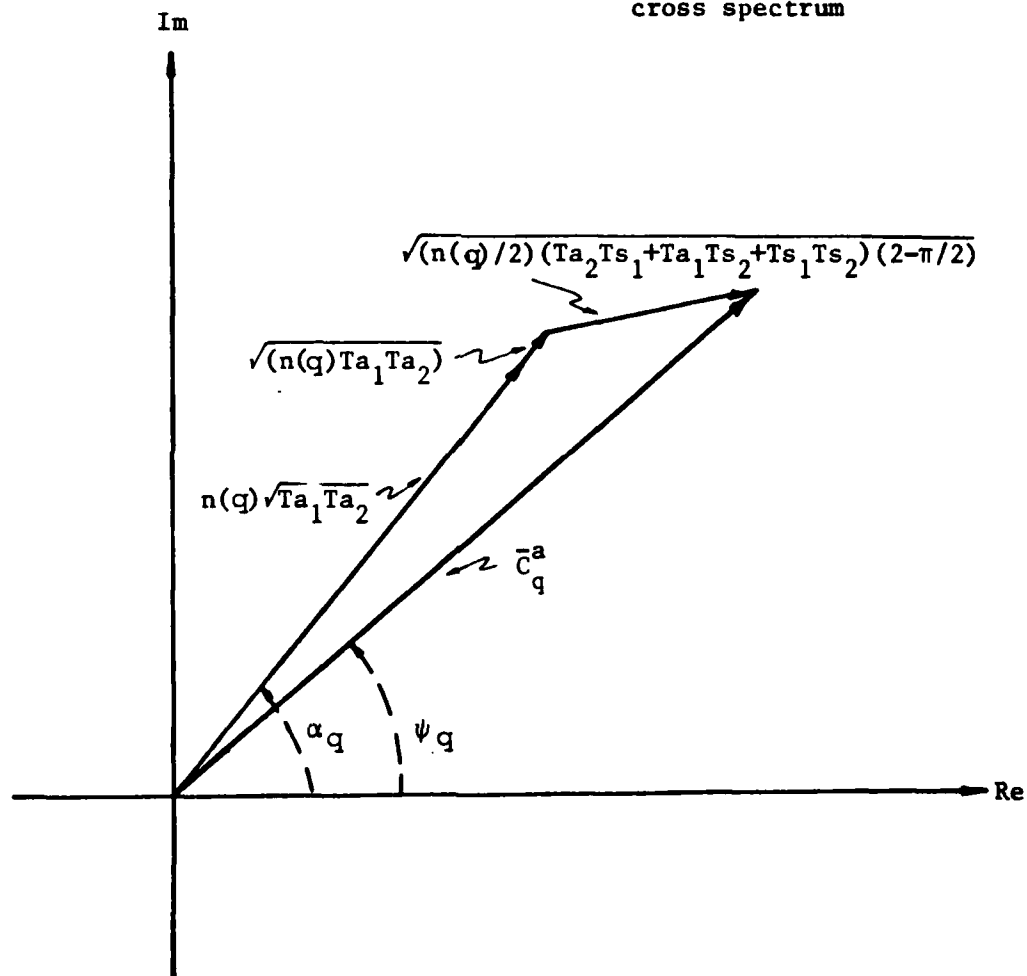
project the noise term in a direction normal to the signal term; it will have a variance of $n(q)(Ta_2Ts_1+Ta_1Ts_2+Ts_1Ts_2)/2$. The variance of the phase will be

$$\sigma_\phi^2 \simeq \frac{Ts_1 Ts_2}{2\, n(q)\, Ta_1 Ta_2} \tag{A.18}$$

(Note: n(q) can nearly always be made sufficently large to satisfy this last assumption because we can always repeat the correlation process with an updated phase delay rate if the initial phase delay rate is not sufficiently accurate to allow n(q) to be selected large enough.)

Also from Figure (A.2) we see that

$$\left| \sum_{j=1}^{n(q)} \bar{X}_{1j} X_{2j}^* \; e^{-i(\alpha_{ap})j} \right| \simeq n(q)\; \sqrt{Ta_1 Ta_2}$$

Combining the above result with Equation (A.14) we have that

$$r_q \simeq \frac{2Ta_1 Ta_2}{Ts_1 Ts_2}\; n(q) \tag{A.19}$$

which by comparing with Equation (A.18) is seen to be inversely proportional to the variance of the phase of the cross spectrum, as expected. This completes the proof of equivalence of the maximum likelihood estimator and weighted least squares.

The above result was not unexpected since it can be proved generally for linear estimation problems with Gaussian statistics that the maximum likelihood and least squares estimators are equivalent. It should be noted that the equivalence is valid only when the linearization in Equation (A.12)

is valid. (In Section 3.1 we saw that the phase differences between the model and the observed cross spectral phases are sufficently small for the linearization of Equation (A.12) to be valid.)

The FRNGE algorithms and least squares are equivalent when the system and antenna temperatures are independent of frequency and time. Now that we know where the FRNGE algorithms are deficient we can use the least squares approach to correct the results from FRNGE if there is evidence that the system and antenna temperatures are not independent of frequency and time.

The least squares approach also provides an excellent method of visualizing how the VLBI observables are estimated. The quantity to be minimized is the weighted sum of the squares of $(\psi_q - \tilde{\alpha}_q)$ where the ψ_q's are the observed phases of the cross spectrum and $\tilde{\alpha}_q$ is simply a mathematical model of the dependence of these phases on the group delay, visibility phase and phase delay rate (Equation (A.3)). The VLBI measurements are made by least squares fitting of straight lines to the observed phases as a function of frequency and time. The slope of the phase versus frequency at a given epoch is the group delay referred to that epcoh; the slope of the phase versus time multiplied by frequency is the phase delay rate, and the intercept of these lines at a given epoch and frequency is the visibility phase. The visibility phase divided by the frequency to which it is referred (ω_o) is the (ambiguous) phase delay.

Finally, before leaving this appendix we ask why FRNGE uses a maximization approach rather than the least squares estimator. Figure (A.3) shows the sum of the squares of the residuals $(\phi_q - \alpha_q)$ when τ_g is varied and $\tilde{\phi}_t$ and $\tilde{t}_p$ are estimated to produce the mininum sum of the squares of the residual phases, i.e. we carried out a series of least square fits to "hypothetical" cross spectral phases but with the group delay constrained to a series of values between -20 and 20 nsec. (See Appendix E for a discussion on "constrained" least squares.) To produce this figure we assumed that the succession of four frequencies in the synthesized band were separated by 100, 250, and 300 MHz. (A review of the equations for the model used to fit the group delay to the cross spectral phases, Equation (A.3), will show that the sum of the squares of the phase residuals does not depend on the absolute frequency of each channel when ϕ_t is estimated as τ_g is varied.) It is clear that there are a number of local minima, to any one of which the least squares may converge (remember, we are dealing with a nonlinear problem and hence we would need to iterate the least squares solution). So even with the least squares approach it would still be necessary to search over many trial values to ensure that the global minimum was found. However, once the region of the correct minimum is located the least squares estimator is an excellent way to estimate the values of the VLBI observables.

We should note in Figure A.3 that after the group delay has been changed by 20 nsec, the sum of the squares of the

Figure A.3 Sum of the squares of the differences between the observed phases and the calculated phases as the trial group delay is varied about the actual value
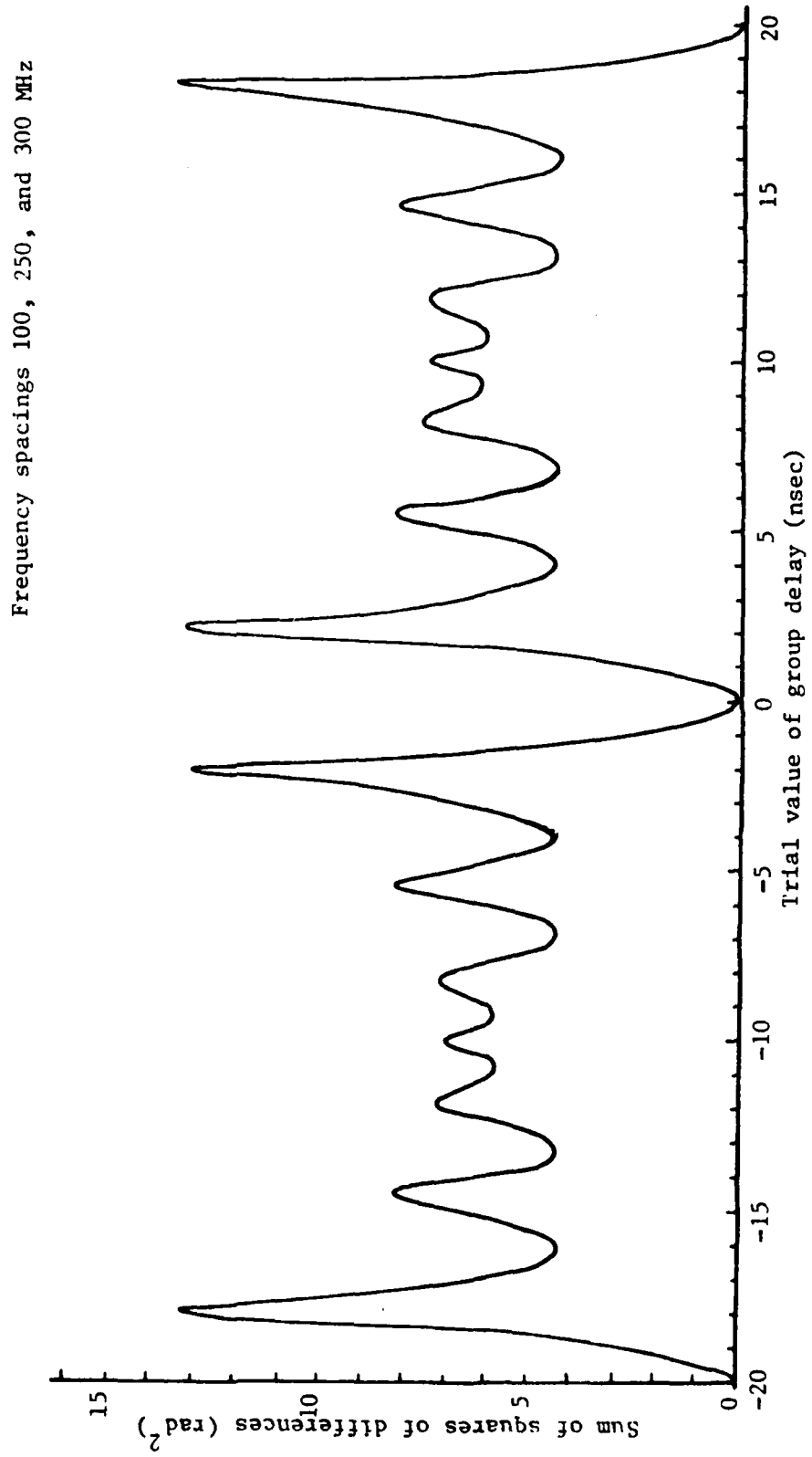
residual cross spectral phases is zero. This perfect fit of the model (Equation (A.3)) to the observations ocurrs because, for the the sequence of frequencies used to produce Figure A.3, a change of 20 nsec in the group delay will change the model phases at each frequency by exact multiples of 2π radians. The spacing between the group delay values which fulfill this relationship is called the group delay ambiguity. Its value is given by the inverse of the largest common denominator of the frequency spacings. It may appear that the group delay ambiguities could be resolved because the fits of the group delay to the cross spectral phases, within each 2 MHz bandwidth channel, will be affected by changes in the group delay by multiples of 20 nsec. This is not the case, however, because in the implementation of the maximization of the delay resolution function, two distinct group delays are estimated: the multiband group delay and the singleband group delay. The multiband group delay is estimated from the change in cross spectral phase between the (widely spaced) channels. The singleband group delay is estimated from the change in phase within the 2 MHz channels. We discuss these two delay measures in detail in Section 2.2 and Appendix B.3 (Equation B.3.3), (see also Whitney, 1974, pp 156-158).

Appendix B.   Evaluation of the delay resolution function

In Appendix A we developed the maximum likelihood esti-
mator for VLBI observables. We assumed that the cross
spectrum of the signals received at each site was available in
the radio frequency (RF) band, i.e. at frequencies around 8
GHz for X-band and 2 GHz for S-band. However, in Chapter 2 we
saw that the signals were recorded at each site after the RF
signals had been mixed with a local oscillator (LO) signal
such that the spectra of the recorded signals were shifted in
frequency by the LO frequency. In this appendix we
investigate the relationship between the spectra of the
recorded ("video") signals, and the spectra of the signals
before they were heterodyned (Section B.1). We will then
investigate the cross correlation function of the video
signals and compare this cross correlation function with the
cross correlation function which would be obtained if the
signals at RF were recorded (Section B.2). Once we have
investigated the cross correlation functions, we will derive
the algorithms used by the program "FRNGE" to estimate the
cross spectral components of the recorded data and to obtain
the maximum likelihood estimates of the group delay, phase
delay rate and visibility phase (Section B.3). Finally we
will investigate the effects of the approximations made in the
correlator algorithms on the group delay, phase delay and
phase delay rate estimates (Section B.4).

Firstly, to avoid any confusion concerning sign conven-

tions and definitions of cross correlation functions and cross spectra, we describe these quantities in Table B.1. An important point to note is that the delay resolution function requires the complex conjugate of the cross spectrum (see discussion concerning Equation (A.10)), which may be obtained by correlating with the sign of the lag (τ) opposite to that shown in Table B.1 (we will discuss this topic further in Section B.2). We have continued to use the complex conjugate for consistency with earlier works (Rogers, 1971; Whitney, 1974; Whitney et al., 1976). (Part of the confusion about conjugates can also be attributed to the cross spectrum $\bar{C}_{21}(\omega)$ being the complex conjugate of the cross spectrum defined in Table B.1, i.e. $\bar{C}_{12}^{*}(\omega) = \bar{C}_{21}(\omega)$, for real signals.)

B.1 Relationship between the video and RF cross spectra

In Chapter 2 we saw that the spectrum of the upper sideband (USB) video signals were related to the spectrum of the RF signals by (see Equation (2.1.1)),

$$\bar{S}_1^u(\omega_v) = \bar{S}_1(\omega_v + \omega_1)\, e^{i\phi_1^u}\, , \qquad \text{for } \omega_v > 0$$

and

$$\bar{S}_1^u(\omega_v) = \bar{S}_1(\omega_v - \omega_1)\, e^{-i\phi_1^u}\, , \qquad \text{for } \omega_v < 0,$$

where $\bar{S}_1^u(\omega_v)$ is the spectrum of the video signals or simply video spectrum, $\bar{S}_1(\omega_v \pm \omega_1)$ is the spectrum of the RF signal, ω_1 is the sum of the frequencies of the first and second local oscillators, and ϕ_1^u is the sum of the phases of the local oscillators for the upper sideband. For the lower sideband

Table B.1 Definitions of the sign conventions for delay, cross correlation function and cross sprectrum.

Delay $\tau_g \equiv t_2 - t_1$,
where t_1 and t_2 are the travel times
from the source to each station
(In this table τ_g will be assumed
constant and equal to the phase delay τ_p.)
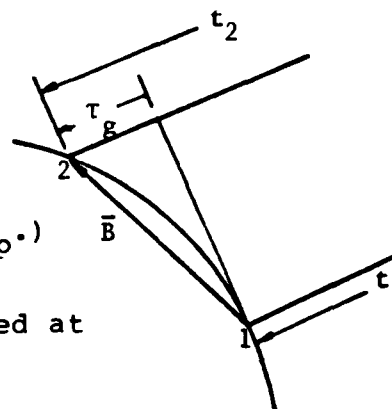
The relationship between signals recorded at each site is (neglecting noise)

$$s_1(t) = s_2(t+\tau_g)$$
or
$$s_2(t) = s_1(t-\tau_g) \ .$$

The spectra of the signals will be

$$\bar{S}_1(\omega) = \int_{-\infty}^{\infty} s_1(t) e^{-i\omega t} dt$$

and

$$\bar{S}_2(\omega) = \int_{-\infty}^{\infty} s_2(t) e^{-i\omega t} dt$$

From the definition of τ_g, these spectra will be related by

$$\bar{S}_2(\omega) = \bar{S}_1(\omega) \ e^{-i\omega\tau_g}$$

The cross correlation function $R_{12}(\tau)$ is

$$R_{12}(\tau) \equiv \int_{-\infty}^{\infty} s_1(t-\tau) s_2(t) dt = \int_{-\infty}^{\infty} s_1(t) s_2(t+\tau) dt$$

The cross spectrum, $\bar{C}_{12}(\omega)$, of the signals from each site is

$$\bar{C}_{12}(\omega) = \int_{-\infty}^{\infty} R_{12}(\tau) \ e^{-i\omega\tau} \ d\tau$$

From the definition of the cross spectrum we have

$$\bar{C}_{12}(\omega) = \bar{S}_1^*(\omega)\bar{S}(\omega) = \bar{S}_1^*(\omega)\bar{S}_1(\omega) e^{-i\omega\tau_g} = \bar{C}_{21}^*(\omega)$$

The delay resolution function requires, as input, the complex conjugate of $\bar{C}_{12}(\omega)$, as defined in this table (see text).

(LSB) video signals, the relationship between the video spectrum and the RF spectrum is (see Equation (2.1.2)),

$$\bar{S}_1^{\ell}(\omega_v) = \bar{S}_1(\omega_v - \omega_1)\, e^{-i\phi_1^{\ell}} \ , \qquad \text{for } \omega_v > 0$$

and

$$\bar{S}_1^{\ell}(\omega_v) = \bar{S}_1(\omega_v + \omega_1)\, e^{i\phi_1^{\ell}} \ , \qquad \text{for } \omega_v < 0,$$

where $\bar{S}_1^{\ell}(\omega_v)$ is the LSB video spectrum and ϕ_1^{ℓ} the sum of the phases of the local oscillators for the lower sideband. The same local oscillator is used to heterodyne both upper and lower sidebands from IF to video, but since the USB and LSB signals propagate through different circuits in the single sideband mixer (see Section 2.1), they could have different phase shifts added to them. Hence we use the two different phases ϕ_1^u and ϕ_1^{ℓ}. The specifications for the video converters indicate that the difference in phase of the upper and lower sidebands should be less than 5° (Mark III System Documentation, 1980, p. VC-8).

At the second site of the interferometer similar relationships exist between the video spectra and the RF spectrum:

$$\bar{S}_2^u(\omega_v) = \bar{S}_2(\omega_v + \omega_2)\, e^{i\phi_2^u} \ , \qquad \text{for } \omega_v > 0, \qquad \text{(B.1.1.a)}$$

$$\bar{S}_2^u(\omega_v) = \bar{S}_2(\omega_v - \omega_2)\, e^{-i\phi_2^u} \ , \qquad \text{for } \omega_v < 0, \qquad \text{(B.1.1.b)}$$

and

$$\bar{S}_2^{\ell}(\omega_v) = \bar{S}_2(\omega_v - \omega_2)\, e^{-i\phi_2^{\ell}} \ , \qquad \text{for } \omega_v > 0, \qquad \text{(B.1.1.c)}$$

$$\bar{S}_2^{\ell}(\omega_v) = \bar{S}_2(\omega_v + \omega_2)\, e^{i\phi_2^{\ell}} \ , \qquad \text{for } \omega_v < 0. \qquad \text{(B.1.1.d)}$$

We now develop the relationship between the video spectra of the data from each site. In Appendix A we derived the

relationship between the spectra for signals at RF. From Equation (A.4) we have

$$\bar{S}_2(\omega) = \bar{S}_1(\omega) \exp[-i(\omega-\omega_0)\tau_g - i\phi_t - i\omega \hat{t}_p \Delta t]$$

where ω_0 is the frequency to which the visibility phase, ϕ_t, is referred, and Δt is the time difference between the epoch of the observation (i.e., the time to which τ_g and ϕ_t are referred) and the time of the determination of the cross spectral components. (Note that the components of the cross spectrum at a specific time can be determined from a 4 μsec segment of data.) If we take ω equal to $\omega_v + \omega_1$ in Equation (A.4), we obtain

$$\bar{S}_2(\omega_v+\omega_1) = \bar{S}_1(\omega_v+\omega_1)\exp\{-i[(\omega_v+\omega_1-\omega_0)\tau_g+ \phi_t +(\omega_v+\omega_1)\hat{t}_p\Delta t]\} \ .$$

But

$$\bar{S}_1(\omega_v+\omega_1) = S_1^u(\omega_v) \ e^{-i\phi_1^u} \ , \qquad\qquad \text{for USB and } \omega_v > 0$$

and

$$\bar{S}_1(\omega_v+\omega_1) = S_1^\ell(\omega_v) \ e^{-i\phi_1^\ell} \ , \qquad\qquad \text{for LSB and } \omega_v < 0.$$

Hence

$$\bar{S}_2(\omega_v+\omega_1) = \bar{S}_1^u(\omega_v)\exp\{-i[(\omega_v+\omega_1-\omega_0)\tau_g+ \phi_t+ (\omega_v +\omega_1)\hat{t}_p\Delta t$$
$$+ \phi_1^u]\} \ , \qquad\qquad \text{for USB, } \omega_v > 0$$

and

$$\bar{S}_2(\omega_v+\omega_1) = \bar{S}_1^\ell(\omega_v)\exp\{-i[(\omega_v+\omega_1-\omega_0)\tau_g+ \phi_t+ (\omega_v +\omega_1)\hat{t}_p\Delta t$$
$$+ \phi_1^\ell]\} \ , \qquad\qquad \text{for LSB, } \omega_v < 0.$$

On the left hand side of the equations we still have the RF spectra of the signals from the second site. To relate these RF spectra to the video spectra we again need to translate the spectra by the local oscillator frequencies. Hence

the spectra of the signals at site 2 at video frequencies become

$$\bar{S}_2(\omega_v + \omega_1) = \bar{S}_2^u(\omega_v + \omega_1 - \omega_2)\, e^{-i\phi_2^u} \quad , \quad \text{for USB, } \omega_v > 0$$

and

$$\bar{S}_2(\omega_v + \omega_1) = \bar{S}_2^\ell(\omega_v + \omega_1 - \omega_2)\, e^{-i\phi_2^\ell} \quad , \quad \text{for LSB, } \omega_v < 0.$$

Therefore the relationship between the video spectra of the recorded data at each site becomes

$$\bar{S}_2^u(\omega_v + \omega_1 - \omega_2) = \bar{S}_1^u(\omega_v)\, \exp\{-i[(\omega_v + \omega_1 - \omega_0)\tau_g + \phi_t$$
$$+ (\omega_v + \omega_1)\dot{\tau}_p \Delta t + (\phi_1^u - \phi_2^u)]\} \quad , \quad \text{for USB, } \omega_v > 0$$

and

$$\bar{S}_2^\ell(\omega_v + \omega_1 - \omega_2) = \bar{S}_1^\ell(\omega_v)\, \exp\{-i[(\omega_v + \omega_1 - \omega_0)\tau_g + \phi_t$$
$$+ (\omega_v + \omega_1)\dot{\tau}_p \Delta t + (\phi_1^\ell - \phi_2^\ell)]\} \quad , \quad \text{for LSB, } \omega_v < 0$$

In the equations given above the video spectra from each site are referred to different frequencies (ω_v for site 1 and $\omega_v + \omega_1 - \omega_2$ for site 2). However, ω_1 and ω_2 are usually quite closely matched since they are derived from hydrogen maser frequency standards. (Even if these frequencies are not closely matched, it is unlikely that they will be greater than the frequency offset due to the phase delay rate; see comments below.) For the data we will be analysing the maximum frequency offset between the hydrogen masers will be 0.1 Hz at X-band (for the Effelsberg hydrogen maser during the July, 1980 experiment). More typical values for frequency differences are 0.01 Hz (again at X-band). These frequency differences are very small compared to the frequency spacings to be

considered in the cross spectrum (0.25 MHz) and the frequency offset due to the rotation of the Earth (this latter offset is implicitly contained in the phase delay rate term, and may be as high as 16.8 kHz at X-band). We can change the frequency offset of the video signals at the second site by introducing a time varying phase term similar to the phase delay rate term.

To see how this term may be introduced, we consider the calculation of the video cross spectrum. The video spectrum at the second site is computed from a finite duration of data (ΔT) by

$$\bar{S}_2^u(\omega_v + \omega_1 - \omega_2) = \int_{\Delta T} s_2(t)\ e^{-i\omega_v t}\ e^{-i(\omega_1 - \omega_2)t}\ dt$$

If $(\omega_1 - \omega_2)\Delta T \ll 1$ (ΔT must be selected to fullfil this condition), then the $(\omega_1 - \omega_2)t$ term may be treated as a constant during the integration which would yield

$$\bar{S}_2^u(\omega_v + \omega_1 - \omega_2) \simeq \bar{S}_2^u(\omega_v)\ e^{-i(\omega_1 - \omega_2)t}.$$

Substituting this equation into the relationship between the video spectra we obtain

$$\bar{S}_2^u(\omega_v) \simeq \bar{S}_1^u(\omega_v)\ \exp\{-i[(\omega_v + \omega_1 - \omega_o)\tau_g + \phi_t + (\omega_v + \omega_1)\dot{\tau}_p\Delta t$$
$$+ (\phi_1^u - \phi_2^u) - (\omega_1 - \omega_2)t]\},\ \text{for USB},\ \omega_v > 0$$

$$(B.1.2.a)$$

and

$$\bar{S}_2^\ell(\omega_v) \simeq \bar{S}_1^\ell(\omega_v)\ \exp\{-i[(\omega_v + \omega_1 - \omega_o)\tau_g + \phi_t + (\omega_v + \omega_1)\dot{\tau}_p\Delta t$$
$$+ (\phi_1^\ell - \phi_2^\ell) - (\omega_1 - \omega_2)t]\},\ \text{for LSB},\ \omega_v < 0$$

$$(B.1.2.b)$$

From Equations (B.1.4) we see that the phase changes due to
the phase delay rate and the oscillator frequency difference
terms have opposite sign for the upper and lower sidebands
(remember $\omega_1 \gg \omega_v$ and ω_1 is always positive). We will see
later in this appendix that this sign reversal causes some of
the errors introduced by these time varying phase terms, to
cancel when upper and lower side band data are combined.

The video cross spectrum will be, for upper sideband

$$\bar{S}_1^u(\omega_v)\bar{S}_2^{u*}(\omega_v) = |\bar{S}_1^u(\omega_v)| \exp\{+i[(\omega_v \pm \omega_1 \mp \omega_o)\tau_g \pm \phi_t$$

$$+ (\omega_v \pm \omega_1)\dot{t}_p\Delta t \pm \Delta\phi^u(t)]\} \qquad (B.1.5.a)$$

and for lower side band

$$\bar{S}_1^\ell(\omega_v)\bar{S}_2^{\ell*}(\omega_v) = |\bar{S}_1^\ell(\omega_v)| \exp\{+i[(\omega_v \mp \omega_1 \pm \omega_o)\tau_g \mp \phi_t$$

$$+ (\omega_v \mp \omega_1)\dot{t}_p\Delta t \mp \Delta\phi^\ell(t)]\} \quad . \qquad (B.1.5.b)$$

If we compare these expressions, with those for the cross
spectrum of the signals at RF, we see that the video and RF
cross spectra are related by

$$\bar{S}_1^u(\omega_v)\bar{S}_2^{u*}(\omega_v) = \bar{S}_1(\omega_v \pm \omega_1)\bar{S}_2^*(\omega_v \pm \omega_2) \exp[\pm i\Delta\phi^u(t)], \quad \text{for USB}$$

$$(B.1.6.a)$$

and

$$\bar{S}_1^\ell(\omega_v)\bar{S}_2^{\ell*}(\omega_v) = \bar{S}_1(\omega_v \mp \omega_1)\bar{S}_2^*(\omega_v \mp \omega_2) \exp[\mp i\Delta\phi^\ell(t)], \quad \text{for LSB.}$$

$$(B.1.6.b)$$

Hence, whenever the delay resolution function requires
the cross spectrum at RF, we can substitute the video cross
spectrum. However, when the video cross spectrum is used
there is an additional time varying term introduced because of

the difference in oscillator frequencies. We should note that the delay resolution function only requires the positive frequency components from the cross spectrum (because only the independent cross spectral components were used to derive the delay resolution function; see Appendix A), and hence the cross spectrum will only need to be computed at positive video frequencies for USB and negative video frequencies for LSB.

Before developing the correlator algorithms it will be instructive to compute the cross correlation function of signals (without noise) recorded at video frequencies. We may find the cross correlation function by taking the Fourier transform of the cross spectrum of the signals. The cross correlation function $R_{12}^V(\tau)$ will be

$$R_{12}^V(\tau) = \int_{-2\pi B}^{2\pi B} \bar{S}_1^{V*}(\omega_v) \bar{S}_2^V(\omega_v)\, e^{i\omega_v \tau}\, d\omega_v \quad ,$$

where $\bar{S}_1^V(\omega_v)$ and $\bar{S}_2^V(\omega_v)$ are used to denote either the upper or the lower sideband video spectrum. The complex conjugates of Equations (B.1.5.a and b) allow us to express the integral above in terms of the power spectra $|\bar{S}_1^u(\omega_v)|$ and $|\bar{S}_1^\ell(\omega_v)|$. For upper sideband data we have

$$R_{12}^u(\tau) = \int_0^{2\pi B} |\bar{S}_1^u(\omega_v)|^2 \exp\{-i[(\omega_v+\omega_1-\omega_o)\tau_g + \phi_t + (\omega_v+\omega_1)\dot{t}_p\Delta t$$

$$+ \Delta\phi^u(t)]\}\, e^{i\omega_v \tau}\, d\omega_v$$

$$+ \int_{-2\pi B}^0 |\bar{S}_1^u(\omega_v)|^2 \exp\{-i[(\omega_v-\omega_1+\omega_o)\tau_g - \phi_t + (\omega_v-\omega_1)\dot{t}_p\Delta t$$

$$- \Delta\phi^u(t)]\}\, e^{i\omega_v \tau}\, d\omega_v$$

where we have split the integration into separate integrals

for positive and negative frequencies because the expressions for the cross spectra are different in these regions. In Chapter 2 we saw that the power spectrum of the signals should be constant over a 2 MHz bandwidth. If we assume this is the case, then we can integrate the expression for $R_{12}^{u}(\tau)$. After some algebra we obtain

$$R_{12}^{u}(\tau) = \cos\{(\omega_1-\omega_0)\tau_g + \omega_1\dot{t}_p\Delta t + \phi_t + \Delta\phi^{u}(t) + \pi B(\tau_g+\dot{t}_p\Delta t - \tau)\}$$

$$\sin\{\pi B(\tau_g + t_p\Delta t - \tau)\}/[\pi B(\tau_g + t_p\Delta t - \tau)] \quad . \quad (B.1.7)$$

In examining this cross correlation function we note that the time lag τ is only multiplied by the bandwidth B whereas τ_g (Table B.1) is multiplied by both the bandwidth and the RF frequency. (Note that ϕ_t has an implicit dependence on frequency because it represents the phase of the cross spectrum at the reference frequency ω_0, i.e. $\phi_t = \omega_0\tau_p$). Consequently as the lag τ between the tapes is changed the arguments in Equation (B.1.7) will change slowly, whereas if either the group or phase delay changes these arguments will change rapidly.

It is instructive to compare Equation (B.1.7) with two other cross correlation functions. If the signals at RF had been recorded (at a sampling rate of π/ω_1 samples/sec), the cross correlation function would be

$$R_{12}^{r}(\tau) = \cos\{(\omega_1-\omega_0)\tau_g + \omega_1 t_p\Delta t + \phi_t - \omega_1\tau + \pi B(\tau_g + t_p\Delta t - \tau)\}$$

$$\sin\{\pi B(\tau_g + t_p\Delta t - \tau)\}/[\pi B(\tau_g + t_p\Delta t - \tau)] \quad .$$

While this expression looks very similar to Equation (B.1.7)

there is a major difference. The cross correlation function
of signals recorded at RF is a function of $\omega_1\tau$ and hence
changes very rapidly as the lag between the tapes is changed.

At the other extreme, if the signals from the radio
source in the frequency range 0-2 MHz were recorded, their
cross correlation function, $R^b_{12}(\tau)$, would be

$$R^b_{12}(\tau) = \cos\{\pi B(\tau_g + t_p\Delta t - \tau)\}$$

$$\sin\{\pi B(\tau_g + t_p\Delta t - \tau)\}/[\pi B(\tau_g + t_p\Delta t - \tau)] \quad .$$

In this case there are no rapid variations in the cross cor-
relation function for variations in either τ or τ_g.

The cross correlation functions given above are shown
graphically in Figure B.1.1. This figure highlights the
advantages of the heterodyned observations. We gain the
sensitivity of the RF recording to τ_g and retain the advan-
tages of slow sampling rate and slow variation of cross cor-
relation function of 0-2 MHz recordings. [The loss of cor-
relation in the heterodyned case is due to the difference in
the oscillator phases, which appears in the cosine term in
Equation (B.1.7). The oscillator phases add directly to the
cross spectral phases at all frequencies in the video band-
width, i.e. unlike a delay, their phase (in the video
spectrum) does not depend on a delay-like quantity. Conse-
quently these phases will change the shape of the video
signals at each site (if the oscillator phases are different),
hence decorrelating the signals. This loss does not occur if
the delay resolution function is used (see Rogers, 1970, and

Figure B.1.1 Cross correlation functions for three types of recording

a. Heterodyned signals

$$\text{modulus}\{(\omega-\omega_o)\tau_g + \omega_o\tau_p, 2\pi\}$$
$$= \pi/4$$
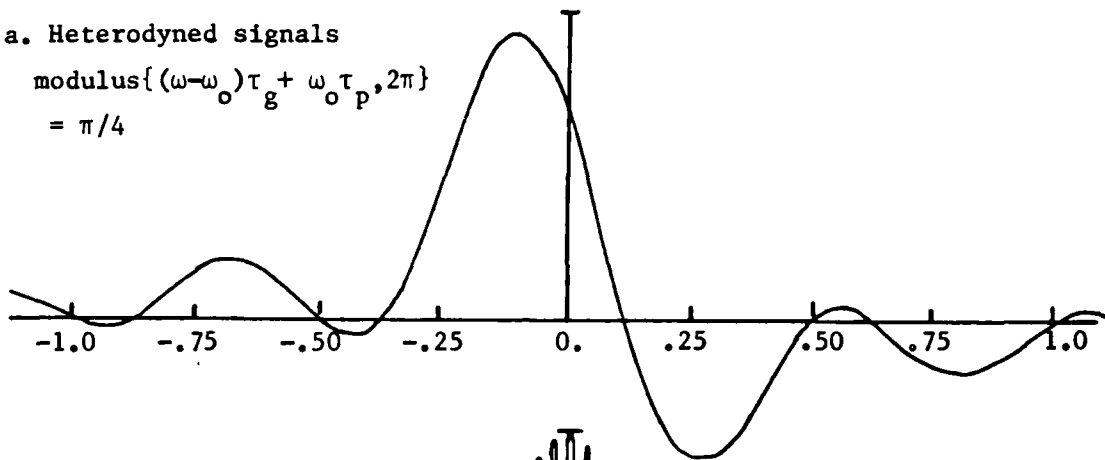
b. Signals recorded at RF
(2MHz bandwidth)
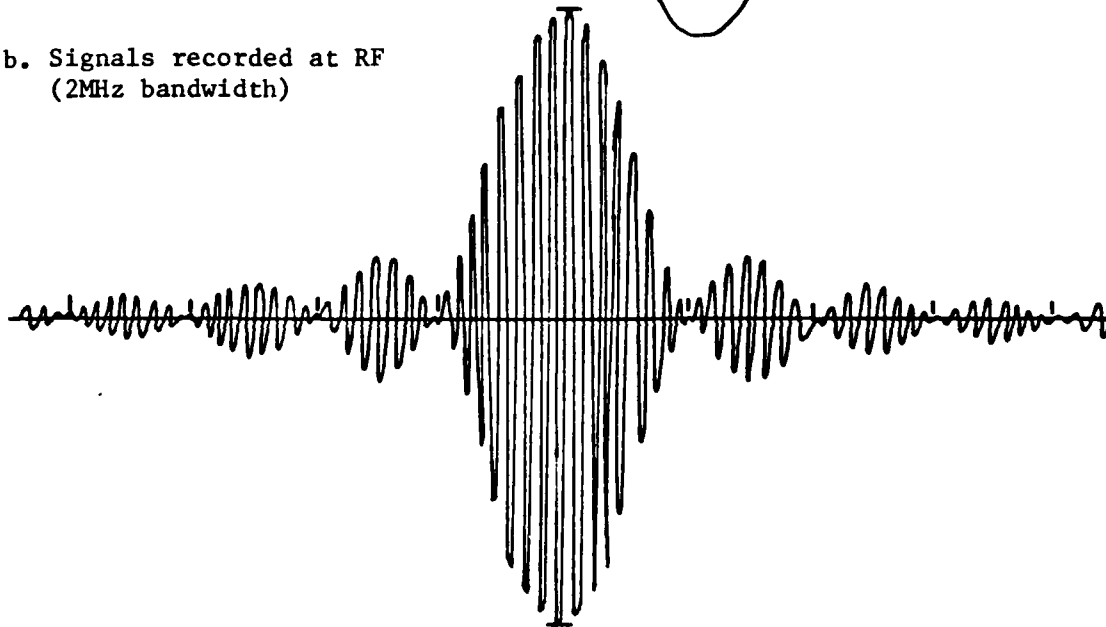
c. Signals recorded at DC
(2MHz bandwidth)

$$\tau_g-\tau \ (\mu sec)$$
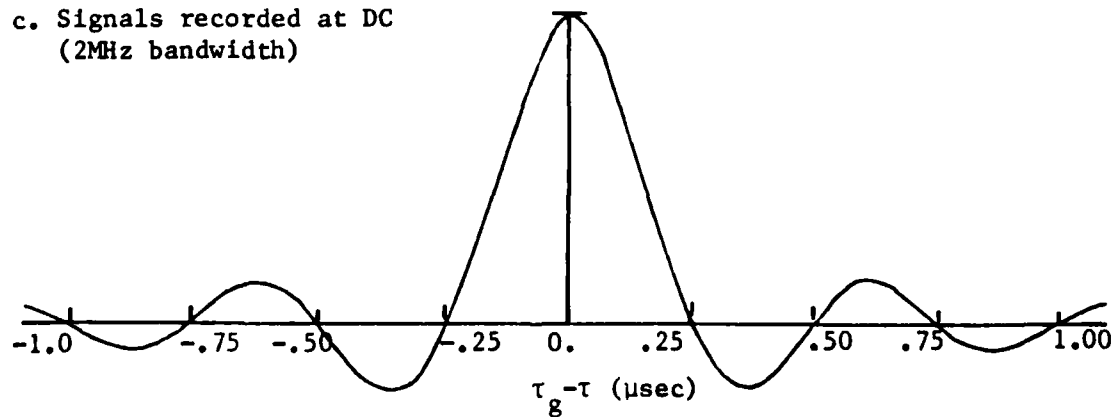
Section B.4).]

An examination of the cross correlation function shown in Figure B.1.1 indicates that the magnitude of the cross correlation function decays rapidly as $\tau_g + \dot{\tau}_p \Delta t - \tau$ deviates from zero. This rapid decay allows the cross spectrum to be computed from the cross correlation function estimated at a small number of lags about the <u>a priori</u> estimate of the delay. The Mark III correlator evaluates the cross correlation function at eight lags centered on the bit offset closest to the <u>a priori</u> estimate of the delay. The lags are spaced at the sampling interval (usually 0.25 μsec). (Later in this appendix we will discuss the effects of using only eight lags.)

B.2 Correlator algorithms

We will now derive the algorithms used by the Mark III correlator to estimate the correlation function from the data recorded on the tapes during an observation. Several approximations will be made while deriving these algorithms and we will investigate the effects of these approximations on the VLBI measurements in Section B.4.

The algorithms will be derived for USB data; the LSB algorithms may be derived using the same procedures. While deriving the correlator algorithms we will introduce many symbols. In Table B.2.1, we define each of the symbols we will be using frequently. (This table is meant as a convenient reference, more exact meanings (if necessary) of the

Table B.2.1   Summary of symbols

$x_1(t_j+\text{offset})$, $x_2(t_j+\text{offset})$ the USB signals recorded at sites
   1 and 2.

t_j the epoch of the jth, 4 µsec data segment.

t_q^c the epoch of the center of the qth accumulation
         period.

t_q^1 the epoch of the first bit in the qth accumulation
         period.

t_c the epoch of the corrections to the a priori group
         delay, phase delay rate and visibility phase (usual-
         ly near the center of the observation).

Δt_q the difference in the epoch of the qth accumulation
         period, t_q^c, and t_c.

Δt_s the sampling interval ($\Delta t_s = 1/(2B) = 0.25$ µsec).

$N(q)$ the number of bits correlated in the qth accumula-
         tion period.

T_s the duration of a short segment of data (4 µsec)

T_{ac} the duration of an accumulation period (typically 2
         sec).

T_{obs} the duration of the observation (100-400 sec).

τ the lag between the tapes during correlation (con-
         tinuous variable).

$\Delta\tau_i$ the discrete lag index (integral multiples of Δt_s);

Table B.2.1 Continued

τ_{ap} the actual lag between the tapes will be $\Delta\tau_i + \tau_{ap}^b$ (see below).

τ_{ap} the <u>a priori</u> value of τ_g (see Table B.1); in most cases τ_{ap} will be a function of time.

τ_{ap}^c the <u>a priori</u> delay at the center of the observation.

$(\tau_{ap}^b)_j$ the nearest bit shift to τ_{ap} for the jth segment of data (at the center of the segment).

τ_{ap}^b $(\tau_{ap}^b)_j$ when reference to a particular segment is not needed.

$(\dot{\tau}_{ap})_q$ the <u>a priori</u> value of the phase delay rate at the center of the qth accumulation period.

$R_u^c(\Delta\tau_i)$ the USB complex cross correlation function at lag $\Delta\tau_i$.

$R_\ell^c(\Delta\tau_i)$ the LSB complex cross correlation function at lag $\Delta\tau_i$.

$C_j(\omega_v)$ the video cross spectrum of the jth segment of data (ω_v takes on discrete values between 0.25 MHz and 1.75MHz). Strictly, this is the conjugate of the cross spectrum as defined in Table B.1; see discussion concerning Table B.1.

$C_{qr}^a(\omega_v)$ the video cross spectrum coherently averaged over the qth accumulation period, r represents the radio frequency (RF) of the channel from which the data were taken (see comment above).

Table B.2.1   Continued

ω_r the radio frequency of $\omega_v = 0$ (video DC) in channel $\underline{r}$[‡]

ω_o the radio frequency to which the visibility phase is
             referred[‡]

$\omega' = \omega_r + \pi B$ the radio frequency corresponding to the center
             frequency of channel $\underline{r}$[‡]

B   the recorded bandwidth (Hz).

$\delta\tau_g$ and $\delta\tilde{\tau}_p$ the trial values of the corrections to the <u>a</u>
             <u>priori</u> values of the multiband group delay and phase
             delay rate, respectively.

$\delta\tilde{\tau}_b$ the correction to the <u>a priori</u> value of the single
             band delay.

$\delta\tau_b'$ the single band delay index (16 values of $\delta\tau_b$ spaced
             at half the sampling interval, Δt_s).

$\delta\bar{\tau}_b'$ the value of $\delta\tau_b'$ which maximizes the multiband delay
             function.

$D_{qr}(\delta\tau_b')$ the single band delay function for the $\underline{q}$th accumula-
             tion period for channel $\underline{r}$; see Section B.2.

$\Delta C(\omega)$ general symbol to denote the integrated cross
             spectrum (exact meaning will depend on context).

$\Delta\omega_e$ general symbol to denote the error in the rate of
             counter rotation of the cross spectral components.

[‡] all frequencies denoted by ω are angular frequency measures
(rad/sec).

symbols will be given in the text.) We start deriving the
algorithms by carrying out the correlation of the data in
short segments (each of 4 μsec duration, i.e. 16 data
samples). Later we will see that, in practice, it is not
necessary to correlate the data in this fashion. By forming a
complex correlation function, we can correlate long segments
of data (of up to 2 sec duration). These long segments are
called accumulation periods. The 4-μsec-segment correlation
technique will, however, give us a convenient means of
deriving, and studying the limitations of, the complex
correlation function procedure.

We denote the USB data on each tape by $x_1(t_j+offset)$ and
$x_2(t_j+offset)$. The time arguments for the data are specified
in two parts, t_j is the epoch (at site 1) of the center of the
4 μsec segment to which a particular data sample is assigned,
and the offset gives the epoch of the data sample relative to
this central epoch. These time arguments are given in dis-
crete form because we have sampled the signals at discrete
times. Using the techniques we discussed in Appendix A, we
compute the cross correlation function at eight lags from 16
samples each of x_1 and x_2. The value of the cross correlation
function for each of the lags, $\Delta\tau_i$, evaluated using the jth
segment of data, will be denoted by $R_j(\Delta\tau_i)$. This cross
function coefficient is given by

$$R_j(\Delta\tau_i) = \sum_{k=-7}^{8} x_1(t_j+ k\Delta t_s) x_2(t_j+ (\tau_{ap}^b)_j + k\Delta t_s - \Delta\tau_i) \quad ,$$

$$(B.2.1)$$

where the $\Delta\tau_i$, the lag indices, are integral multiples of the sampling interval Δt_s, and $(\tau_{ap}^b)_j$ is the epoch of the bit nearest to the a priori estimate of the delay between the tapes at the center (k=0) of the jth segment of data. When $(\tau_{ap}^b)_j$ is converted to a offset of the tapes to compensate for this delay, the offset is called the "bit shift." The cross spectrum can now be computed from this cross correlation function. The cross spectral components $C_j(\omega_v)$ from the jth segment will be

$$C_j(\omega_v) = \sum_{i=-4}^{3} R_j(\Delta\tau_i) \ e^{-i\omega_v \Delta\tau_i} \quad , \qquad (B.2.2)$$

where ω_v are 7 video frequencies spaced at intervals of 0.25 MHz over the 2 MHz bandwidth commencing from 0.25 MHz. We now wish to sum together these cross spectral components from each data segment accounting for the changes in phase due to the phase delay rate. But we must remember the cross spectral components which we have computed at video frequency are the same spectral components which were originally at $\omega_v + \omega_r$, where ω_r is the radio frequency (i.e., ω_1 in earlier discussions) of the rth channel in which the data were taken, of the signal when the signal arrived at the radio telescopes. (We derived the expressions which relate the video cross spectrum to the RF cross spectrum earlier in this appendix (see Equations (B.1.6)). In those derivations we needed to distinguish between the sum of the LO frequencies at the two radio telescopes (hence the notation ω_1 and ω_2). We now need to distinguish between each of the channels and hence the change in

notation.) The phases of these cross spectral components will be changing at a rate given by the product of their original frequency $(\omega_v+\omega_r)$ and the phase delay rate. Hence the accumulated cross spectrum will be given by

$$C^a_{qr}(\omega_v) = \sum_{j=1}^{n(q)} C_j(\omega_v)\, e^{-i(\omega_v+\omega_r)(\dot{t}_{ap})_q(t_j-t_q^C)} \quad , \quad (B.2.3)$$

where n(q) is the number of data segments in the qth accumulation period, t_q^C is the epoch of the center of the qth accumulation period and $(\dot{t}_{ap})_q$ is the _a priori_ phase delay rate computed at the center of the acumulation period. We should note that this expression is only valid for $\omega_v>0$. For $\omega_v<0$ the sign of ω_r should be changed (see Equations (B.1.4)). The cross spectral components for $\omega_v<0$ are not correctly rotated. This error is not a problem because the delay resolution function only requires the cross spectral components with $\omega_v>0$ for USB data (see discussion associated with Equation (B.1.6)).

When we combine all of the previous formulas together we obtain

$$C^a_{qr}(\omega_v) = \sum_{j=1}^{n(q)} \sum_{i=-4}^{3} \sum_{k=-7}^{8} x_1(t_j+k\Delta t_s) x_2(t_j+(\tau^b_{ap})_j+k\Delta t_s-\Delta\tau_i)$$

$$\exp\{-i[\omega_v\Delta\tau_i + (\omega_v+\omega_r)(\dot{t}_{ap})_q(t_j-t_q^C)]\} \quad .$$

$$(B.2.4)$$

If we rearrange the terms in the above equation and substitute $\omega'=\omega_r+\pi B$ (ω' is the radio frequency corresponding to the

center of the video band), we obtain

$$
\begin{aligned}
C_{qr}^{a}(\omega_v) = \sum_{i=-4}^{3} \sum_{j=1}^{n(q)} \sum_{k=-7}^{8} & \{ x_1(t_j+k\Delta t_s) x_2(t_j+(\tau_{ap}^{b})_j+k\Delta t_s-\Delta\tau_i) \\
& \exp[-i\omega'(\dot{\tau}_{ap})_q(t_j-t_q^{c})]\} \; \exp[-i\omega_v\Delta\tau_i] \\
& \exp[-i(\omega_v-\pi B)(\dot{\tau}_{ap})_q(t_j-t_q^{c})] \quad . \qquad (B.2.5)
\end{aligned}
$$

The last term in the above expression represents the differ-ential rotation of the cross spectral components at ω_v rela-tive to the component at the center of the video band, due to the phase delay rate. If we neglect this term we see that the video frequency only appears in one other place (the exp $(-i\omega_v\Delta\tau_i)$ term). The exp $(-i\omega_v\Delta\tau_i)$ term is independent of $\underline{j}$ and $\underline{k}$ and hence we can perform the summation over $\underline{j}$ and $\underline{k}$ without specifying which ω_v will be used (we should note that ω_r does not depend on the choice of ω_v either). (If the last term is included in the summation, then ω_v must be specified in order to carry out the summation over $\underline{j}$). Also, if we re-examine the summations over $\underline{j}$ and $\underline{k}$, we see there is no reason to keep them as separate summations when the last term in Equation (B.2.5) is neglected. So we will combine these two summations together. To combine the summations we need to change the way we specify the epoch of a bit as a function of the summation indices. We will introduce a new summation index which is related to $\underline{j}$ and $\underline{k}$ by $p = 16j + k$ (remember that the $\underline{j}$ index steps through the segments, each of which is 16 bits long). The epoch of a bit at Site 1 will now be given by $t_q^{1}+p\Delta t_s$, where t_q^{1} is the epoch of the first bit (at Site

1), in the accumulation period (i.e, $t_q^1 + p\Delta t_s = t_j + k\Delta t_s$ when p = 16j+k). With the introduction of the summation index $\underline{p}$, Equation (B.2.5) becomes (when we neglect the $\exp(-i(\omega_v - \pi B)(\dot{\tau}_{ap})_q (t_j - t_q^C))$ term),

$$C_{qr}^a(\omega_v) = \sum_{i=-4}^{3} \left\{ \sum_{p=1}^{N(q)} x_1(t_q^1 + p\Delta t_s) x_2(t_q^1 + (\tau_{ap}^b)_p + p\Delta t_s - \Delta\tau_i) \right.$$
$$\left. \exp[-i\omega'(\dot{\tau}_{ap})_q(t_p - t_q^C)] \right\} e^{-i\omega_v \Delta\tau_i} \quad , \quad (B.2.6)$$

where N(q) = 16n(q) and equals the number of bits in the $\underline{q}$th accumulation period. The inner summation is called the complex cross correlation function, $R_u^C(\Delta\tau_i)$, which is evaluated at 8 complex lags (the complex lag is the combination of the lag $\Delta\tau_i$ and the rotation $\exp[-i\omega'(\dot{\tau}_{ap})_q(t_p - t_q^C)]$, i.e.

$$R_u^C(\Delta\tau_i) = \sum_{p=1}^{N(q)} x_1(t_q^1 + p\Delta t_s) x_2(t_q^1 + (\tau_{ap}^b)_p + p\Delta t_s - \Delta\tau_i)$$
$$\exp[-i\omega'(\dot{\tau}_{ap})_q(t_p - t_q^C)] \quad . \quad (B.2.7)$$

The cross spectrum is obtained by computing the Fourier transform of the complex cross correlation function. The Fourier transform is approximated by the summation over $\underline{i}$ in Equation (B.2.6). The use of 4 µsec segments of data is now no longer necessary; all the data in an accumulation period is summed directly into the complex cross correlation function.

B.3 "FRNGE" algorithms

Now that we have derived the correlator algorithms, we are ready to develop the algorithms used by FRNGE. The

correlator only computes the complex cross correlation function. The components of the cross spectrum are computed in FRNGE. Once these components have been computed, they can be used to evaluate the delay resolution function for trial values $\tilde{\tau}_g, \tilde{\dot{\tau}}_p$ of the group delay and phase delay rate.

The delay resolution function was given in Appendix A (Equation (A.10)). When the cross spectral components from all accumulation periods are used, the delay resolution function becomes

$$\sum_{q=1}^{Q} \sum_{r=1}^{R} \sum_{j=-7}^{7} C_{qr}^{a}(\omega_{vj}) \exp[-i(\omega_r + \omega_{vj} - \omega_o) \delta\tilde{\tau}_g - i(\omega_r + \omega_{vj}) \delta\tilde{\dot{\tau}}_p (t_q^c - t_c)]$$

(B.3.1)

where $\delta\tilde{\tau}_q$ and $\delta\tilde{\dot{\tau}}_p$ are, respectively, the corrections to the group delay and the phase delay rate at the epoch t_c of the observation (usually t_c is near the center of the observation). The summation over q represents a summation over time, Q being the number of accumulation periods; the r summation index runs over the number of frequency channels (typically <8 for geodetic observations); and the j index runs over the discrete video frequencies for upper and lower sidebands (in earlier equations we did not need to specify discrete video frequencies and hence we did not use a subscript). (The summation over j excludes the video DC term j = 0.) (In Equation (A.10) these three summations were expressed as a single summation (see comments made at that time).) If only one sideband is available the summation need only be evaluated for the video frequencies which are

available.

If we had an infinitely fast computer, or were planning to make only a few Mark III observations, we could directly sum Equation (B.3.1) using a series of trial values for $\delta\tilde{\tau}_g$ and $\delta\tilde{\tau}_p$ until we found the maximum value of Equation (B.3.1). (In a typical observation $\approx 10^5$ values for $\delta\tilde{\tau}_g$ and $\delta\tilde{\tau}_p$ are tried.) Unfortunately, computers work at finite rates and the number of Mark III observations is not small (between July, 1980, and January, 1982, over 60,000 Mark III observations were made in geodetic experiments). Consequently the actual calculation of Equation (B.3.1) is not by direct summation. A large saving in computation time can be made by adopting two group delay estimates rather than the single one ($\delta\tilde{\tau}_g$) shown in Equation (B.?.1). These two delays are called the multiband group delay and the singleband group delay. The multiband group delay is determined from the change of phase between the widely spaced frequencies, the singleband group delay is determined from the change of phase across the 2 MHz recorded bandwidth. (There are other distinctions between these delay types which will be discussed later.)

We may incorporate the singleband delay in Equation (B.3.1) by writing this equation as

$$\sum_{q=1}^{Q} \sum_{r=1}^{R} \sum_{j=-7}^{7} C_{qr}^{a}(\omega_{vj}) \exp[-i(\omega_r-\omega_o)\delta\tilde{\tau}_g - i(\omega_r+\omega_{vj})\delta\tilde{t}_p(t_q^c-t_c)$$

$$-i\omega_{vj}\delta\tilde{\tau}_b]$$

$$(B.3.2)$$

where $\delta\tilde{\tau}_b$ is called the single band delay. We don't seem to have saved any computer time by introducing the singleband delay. Actually, the maximization of Equation (B.3.2) would take longer than the maximization of Equation (B.3.1), because in the former case we need to search over three quantities $(\delta\tilde{\tau}_g, \delta\tilde{t}_p$ and $\delta\tilde{\tau}_b)$. The improvement in the computation time comes about by using an efficient method to search over the values of $\delta\tilde{\tau}_g, \delta\tilde{t}_p$ and $\delta\tilde{\tau}_b$.

The values of $\delta\tilde{\tau}_g, \delta\tilde{t}_p$ and $\delta\tilde{\tau}_b$ which maximize Equation (B.3.2) will be the maximum likelihood estimates (within the limits of the approximations made during the derivation; see Section B.4), of the multiband group delay, singleband group delay and phase delay rate. There are a number of techniques we could use to perform the maximization, e.g. "brute force" searching, using successively finer spacings of the trial values until a maximum is found, or we could perform a "coarse" search and use the least squares estimator with the coarse search estimates as the a priori values (see Appendix A). For the remainder of this appendix, we will discuss one particular implementation of the maximization of Equation (B.3.2), that carried out in the Haystack Observatory's program "FRNGE." This program has been used to process all of the data which is analysed in Chapters 3 through 8. The program "FRNGE", being software, is subject to change (especially if its algorithms are found to deficient). The algorithms which we will discuss are those used at this time (May, 1983) but they may change in the future (see, e.g., the

discussions in Section B.4 on the neglect of video rotation),
making obsolete some of the discussion which follows.

The efficent search methods are developed by carrying out
the summations in Equation (B.3.2) in stages. Firstly, the
summation over j is performed to generate the "singleband"
delay function, $D_{qr}(\delta\tau_b')$, (the remaining summations are
performed later),

$$D_{qr}(\delta\tau_b') = \sum_{j=-7}^{7} C_{qr}^{a}(\omega_{vj}) \exp[-i\omega_v \delta\tau_b'] \qquad (B.3.3)$$

Equation (B.3.3) is evaluated at 16 trial values of the cor-
rection, $\delta\tau_b'$, to the a priori delay, spaced at half the sam-
pling interval and centered on zero. Before proceeding, let
us consider the meaning of Equation (B.3.3). If we were to
maximize the magnitude of $D_{qr}(\delta\tau_b')$ by searching over many
values of $\delta\tau_b'$, what would be the results (assuming that $\delta\tau_b'$ is
no longer restricted to the 16 discrete values)? For dual
sideband data, the value of $\delta\tau_b'$ which maximizes the magnitude
of $D_{qr}(\delta\tau_b')$, would be a least squares estimate of the slope of
the cross spectral phases across the combined 4 MHz channel
pair. For single sideband data, the slope is across the 2 MHz
channel; see Appendix A. The phase of $D_{qr}(\delta\hat{\tau}_b)$, where $\delta\hat{\tau}_b$ is
the value of $\delta\tau_b'$ which maximizes the magnitude of $D_{qr}(\delta\tau_b')$,
would be the estimate of the intercept of the sloped line at
$\omega_{vj} = 0$, the DC edge of the video band, which corresponds to
the radio frequency ω_r (see Section B.1). We will refer to
this intercept as the "video DC phase". Consequently,
maximizing the singleband delay function can be viewed as

estimating, from the 14 cross spectral phases for dual sideband data, a slope across the recorded bandwidth (the singleband delay) and a phase at video DC (the video DC phase). If single sideband data are used, only seven cross spectral phases are used to estimate the slope and the intercept. These concepts of the nature of the singleband delay function will be important later on.

The uniform spacing of the values $\delta\tau_b'$ allows Equation (B.3.3) to be computed using a 32 point fast Fourier transform (FFT) routine.

After the computation of the singleband delay function, the remaining summations in Equation (B.3.2) are carried out for many trial values of the phase delay rate and multiband group delay . These summations are again carried out using FFT routines. To facilitate using the FFT routines the phase delay rate summation is computed for trial values of the "fringe rate" rather than for trail values of the phase delay rate. (The "fringe rate" is defined to be $\omega_r \bar{t}_p$, and represents the rate of change of phase with time, in contrast to the rate of change of delay with time, which is the phase delay rate. The fringe rate is used in doing the FFT's because standard routines (contained in firmware) which use the standard transformation between the "time" and "frequency" domains can be directly used.) The FFT's constructed using the fringe rates and the multiband delays are called the "fringe rate spectrum" and the "multiband delay function."

The maximum value of the multiband delay function will

occur at the trial values of the multiband group delay, the singleband group delay and the fringe rate which are closest to the values of these quantities which maximize the delay resolution function. The FFT's used to compute the multiband delay function can be computed very rapidly which allows the evaluation of the multiband delay function at a large number of trial values of $\tilde{\tau}_g, \tilde{\tau}_b$ and $\tilde{\dot{\tau}}_p$. (Usually between 128 and 256 values are used for $\tilde{\tau}_g$ and $\tilde{\dot{\tau}}_p$, 16 values of $\tilde{\tau}_b$ are used as mentioned earlier ($\tilde{\tau}_b$ will not be determined very accurately and hence does not require a fine spacing (see earlier discussions on the interpretation of the singleband delay)).) To determine the final values of τ_g, τ_b and $\dot{\tau}_p$ an iterative procedure is used in which the delay resolution function is summed directly using finer spaced values of $\tilde{\tau}_g$ and $\tilde{\dot{\tau}}_p$ which are centered on those values, found from the FFT's in the first iteration, or from the previous iteration in subsequent iterations. (These summations will only be carried out a relatively small number of times and hence the efficiency of the FFT's is not needed.)

The delay resolution function, $R(\delta\tilde{\tau}_g, \delta\tilde{\dot{\tau}}_p, \delta\tilde{\tau}_b)$, as implemented in FRNGE, is constructed from the singleband delay functions (rather than from the cross spectral components) by direct summation of these functions:

$$R(\delta\tilde{\tau}_g, \delta\tilde{\dot{\tau}}_p, \delta\tilde{\tau}_b) = \sum_{q=1}^{Q} \sum_{r=1}^{R} D_{qr}(\delta\tilde{\tau}_b') \exp[-i(\omega_r - \omega_o)\delta\tilde{\tau}_g$$

$$-i\omega_r \delta\tilde{\dot{\tau}}_p (t_q^c - t_c) - i\phi_r^{cal} - i\pi B\eta_{qr}(\delta\tilde{\tau}_b - \delta\tilde{\tau}_b')]$$

$$(B.3.4)$$

where ϕ_r^{cal} is the phase calibration phase for the $\underline{r}$th channel
(these phases are discussed in detail in Section 2.3); $\delta\bar{\tau}_b'$ is
the value of $\delta\tau_b'$ which maximized the multiband delay function
(remember that $\delta\tau_b'$ were 16 discrete values used to construct
the single band delay function). The final term,
$(-i\pi B\eta_{qr}(\delta\tilde{\tau}-\delta\tau_b'))$, in Equation (B.3.4) we will examine in some
detail. The variable η_{qr} is given by

$$\eta_{qr} = [N_\ell(q)-N_u(q)]/[N_\ell(q)+N_u(q)] \qquad (B.3.5)$$

where $N_u(q)$ and $N_\ell(q)$ are the number of bits correlated in the
upper and the lower sidebands, respectively, in the $\underline{q}$th
accumulation period. (Prior to this time we had not needed to
distinguish, in the summations, between upper and lower side-
bands and hence we have been using one symbol to denote the
total number of bits correlated in both sidebands, $N(q)$, ($N(q)$
$= N_u(q)+N_\ell(q)$; see Table B.2.1).) If we examine Equation
(B.3.5) we see that if only USB data are available $\eta_{qr}=-1$,
i.e. $N_\ell(q)=0$; if only LSB data is available, $\eta_{qr}=+1$, i.e.
$N_u(q) = 0$. We first consider these two limiting cases. In
Figure B.3.1 we show graphically the purpose of this last
term. In the upper plot we show the cross spectral phases
from USB data in two frequency channels (we have chosen just
two to simplify the figure). Because the singleband delay
index $\delta\tau_b'$ is discrete, the value of $\delta\tau_b$ which maximizes
$D(\delta\tilde{\tau}_g,\delta\tilde{\tau}_b,\delta\tilde{\tau}_p)$ will not necessarily correspond to one of the
discrete indices. The singleband delay is obtained from the
least squares fit of a group delay and a phase at ω_r (video

Figure B.3.1 Effects of the discrete singleband delay on the video DC phase estimates

a. USB only

Cross spectral phase

$\delta\tilde{\tau}_g$

multiband group delay estimate

nearest singleband delay index

$\delta\bar{\tilde{\tau}}_b'$

ϕ_1^{cal}

ϕ_2^{cal}

$\delta\bar{\tilde{\tau}}_b'$

$\delta\tilde{\tau}_b$

$-\pi B(\delta\tilde{\tau}_b - \delta\bar{\tilde{\tau}}_b')$

$\delta\tilde{\tau}_b$

B

RF

b. LSB only

$\delta\tilde{\tau}_g$

ϕ_2^{cal}

ϕ_1^{cal}

$\delta\bar{\tilde{\tau}}_b'$

$\delta\tilde{\tau}_b$

$\pi B(\delta\tilde{\tau}_b - \delta\bar{\tilde{\tau}}_b')$

$\delta\tilde{\tau}_b$

$\delta\bar{\tilde{\tau}}_b'$

B

RF

c. Dual side-band

$\delta\tilde{\tau}_g$

ϕ_2^{cal}

$\delta\bar{\tilde{\tau}}_b'$

ϕ_1^{cal}

$\delta\tilde{\tau}_b$

$\delta\tilde{\tau}_b$

$\delta\bar{\tilde{\tau}}_b'$

RF

Symbols- • USB cross spectral phases

■ LSB cross spectral phases

DC) to the cross spectral phases in the 2 MHz bandwidth channels (see discussion on Equation B.3.3). Because the singleband delay function (Equation (B.3.3)) is evaluated at discrete, equally spaced, values of $\delta\tau_b'$ (rather than at the value of $\delta\tilde{\tau}_b$ which will maximize the delay resolution), the phase at ω_r will be in error by $-\pi B(\delta\tilde{\tau}_b - \delta\tilde{\tau}_b')$ when only USB data used (also see discussion below). The final term in Equation (B.3.4) corrects for the phase error at video DC.

Figure B.3.1.b shows the effects of the discrete single-band delay on LSB data. From this figure we clearly see the same error as in the USB case except that the sign of the error has changed. Figures B.3.1.a and b should indicate why the term involving η_{qr} is included in Equation (B.3.4), but how do we correct the immediate cases when some data are available in both sidebands? We need to calculate the video DC phase error (see discussion on Equation (B.3.3) which will result, when we perform a least squares fit of a slope and intercept to unequally weighted data, when the slope is total-ly constrained to a specific value (see Appendix E for a discussion on constrained least squares estimation). The value of η_{qr} will then be calculated to correct for this error. (The values of the slope are constrained because the single band delay function has only be computed at discrete values of $\delta\tau_b'$.) This estimation problem is simplified by the weights, given to each cross spectral phase, being equal within a sidband. (This is not quite true because the response of the bandpass filters, in the video converters, is

not constant across the 2 MHz bandwidth of each sideband. See
Section B.4 for more details.) Since the weights, within a
sideband are approximately equal, we may replace the seven
cross spectral phases, from each sideband, with their mean
value at their central frequency (1 MHz for USB and -1 MHz for
LSB). Consequently, the phase error at video DC, when the two
sidebands are combined, will be the weighted mean of these
central frequency phases. The difference, $\Delta\phi_w$, between the
weighted mean and the algebraic mean, will be given by

$$\Delta\phi_w = \left[\left[w_\ell/(w_u+w_\ell) \right] (-\pi B) + \left[w_u/(w_u+w_\ell) \right] (\pi B) \right] (\delta\bar{\tau}_b - \delta\bar{\tau}_b')$$

(B.3.6)

where w_u and w_ℓ are the weights given to the phases from the
upper and the lower sidebands. In Appendix A, we saw that the
weights to be used during the least squares solution are
proportional to the product of the number of bits correlated
and the product of the antenna temperatures divided by the
system temperatures at the two sites in the interferometer
(see Equation (A.19)). The latter term in the weight (i.e.,
the product of the antenna temperatures divided by the system
temperatures), will probably remain constant during an obser-
vation (typically 100 sec) and hence the major variations in
the weights assigned to each sideband will be due to changes
in the number of bits correlated. Hence if we replace w_u and
w_ℓ with $N_u(q)$ and $N_\ell(q)$, in Equation (B.3.6), we obtain the
expression containing η_{qr} in Equation (B.3.4). The possible
errors which are introduced into the estimates of the multi-

DC) to the cross spectral phases in the 2 MHz bandwidth channels (see discussion on Equation B.3.3). Because the singleband delay function (Equation (B.3.3)) is evaluated at discrete, equally spaced, values of $\delta\tau_b'$ (rather than at the value of $\delta\bar{\tau}_b$ which will maximize the delay resolution), the phase at ω_r will be in error by $-\pi B(\delta\bar{\tau}_b-\delta\bar{\tau}_b')$ when only USB data used (also see discussion below). The final term in Equation (B.3.4) corrects for the phase error at video DC.

Figure B.3.1.b shows the effects of the discrete single-band delay on LSB data. From this figure we clearly see the same error as in the USB case except that the sign of the error has changed. Figures B.3.1.a and b should indicate why the term involving η_{qr} is included in Equation (B.3.4), but how do we correct the immediate cases when some data are available in both sidebands? We need to calculate the video DC phase error (see discussion on Equation (B.3.3) which will result, when we perform a least squares fit of a slope and intercept to unequally weighted data, when the slope is total-ly constrained to a specific value (see Appendix E for a discussion on constrained least squares estimation). The value of η_{qr} will then be calculated to correct for this error. (The values of the slope are constrained because the single band delay function has only be computed at discrete values of $\delta\tau_b'$.) This estimation problem is simplified by the weights, given to each cross spectral phase, being equal within a sidband. (This is not quite true because the response of the bandpass filters, in the video converters, is

not constant across the 2 MHz bandwidth of each sideband. See Section B.4 for more details.) Since the weights, within a sideband are approximately equal, we may replace the seven cross spectral phases, from each sideband, with their mean value at their central frequency (1 MHz for USB and -1 MHz for LSB). Consequently, the phase error at video DC, when the two sidebands are combined, will be the weighted mean of these central frequency phases. The difference, $\Delta\phi_w$, between the weighted mean and the algebraic mean, will be given by

$$\Delta\phi_w = \left[\, [w_\ell/(w_u+w_\ell)] \, (-\pi B) \; + \; [w_u/(w_u+w_\ell)] \, (\pi B) \, \right] \, (\delta\bar{\tau}_b - \delta\bar{\tau}_b')$$

$$(B.3.6)$$

where w_u and w_ℓ are the weights given to the phases from the upper and the lower sidebands. In Appendix A, we saw that the weights to be used during the least squares solution are proportional to the product of the number of bits correlated and the product of the antenna temperatures divided by the system temperatures at the two sites in the interferometer (see Equation (A.19)). The latter term in the weight (i.e., the product of the antenna temperatures divided by the system temperatures), will probably remain constant during an observation (typically 100 sec) and hence the major variations in the weights assigned to each sideband will be due to changes in the number of bits correlated. Hence if we replace w_u and w_ℓ with $N_u(q)$ and $N_\ell(q)$, in Equation (B.3.6), we obtain the expression containing η_{qr} in Equation (B.3.4). The possible errors which are introduced into the estimates of the multi-

band group delay by the use of the discrete singleband delay are discussed further in Section 3.1.

The need for this correction to the phase at ω_r could be eliminated if the singleband delay functions were recalculated from the cross-spectral components at $\delta\tau_b$. This approach would however lead to considerably more computations. If these additional computations were carried out (and consequently η_{qr} set to zero in all cases), Equation (B.3.4) would be the same as Equation (B.3.2), with the addition of the phase calibration phases. (There is one term neglected in going from Equation (B.3.2) to Equation (B.3.4): the term $-i\omega_{vj}\delta\tilde{\tau}_p(t_q^c-t_c)$. We will discuss the effect of this approximation when we are investigating a similar term which was neglected in deriving Equation (B.2.6), namely $-i\omega_{vj}\tilde{\tau}_p(t_j-t_q^c)$.)

A parabola is fitted through the magnitude of the delay resolution function evaluated at three points centered on and equally spaced about the estimates for the multiband group delay, singleband group delay and phase delay rate from either the FFT's or the previous iteration. Such a parabola is determined separately for each observable, i.e. the parabola is fitted in only one dimension. The phase delay rate is fitted first, followed by the multiband group delay and then the singleband delay is estimated. The fitting process is repeated twice more using the new estimates of the multiband delay, phase delay rate and the singleband delay. The final iteration is carried out with half the spacing of the trial

values for the multiband delay and phase delay rate. The
visibility phase is then obtained from the phase of the delay
resolution function evaluated at the final estimates of the
multiband delay, single band delay and phase delay rate.

Is this search technique assured of converging to the
actual maximum value of the delay resolution? If it were
iterated an "infinite" number of times, using successively
finer spacings for the one dimensional parabolic fits - yes.
To see why, for the delay resolution function, convergence is
assured (it isn't assured for all functions), we need to
consider several properties of the delay resolution function
and the initial search algorithms. We stated earlier that the
initial search over many trial values is carried out using
FFT's. We may use several properties of Fourier transforms to
deduce the properties of the delay resolution function. The
delay resolution function is "bandlimited" (see, e.g.,
Bracewell, 1978, for a discussion on properties of discrete
Fourier transforms), in both the delay and rate domains
because only a finite bandwidth is spanned (usually <400 MHz),
and observations are of a finite duration (usually between 100
and 400 sec). Because the spanned bandwidth and observation
duration are finite there are finite maximum frequencies which
can exist in the delay resolution function. (We are using
frequency here as a general term to indicate variations of the
magnitude of the delay resolution function with changes of $\delta\tau_g$
and $\delta\dot\tau_p$, rather than in its usual sense of variations with
time.) The highest frequency variations which can occur with

changes in $\delta\tilde{\tau}_g$ and $\delta\tilde{t}_p$ are the inverse of spanned bandwidth and the inverse of the duration of the observations, respectively. The initial trial values $\delta\tilde{\tau}_g$ and $\delta\tilde{t}_p$ are separated at most by a quarter and a half of the inverses of these highest frequency variations, respectively. (The numbers of trial values of $\delta\tilde{\tau}_g$ and $\delta\tilde{t}_p$ are selected to be the smallest powers of two (to insure efficient operation of the FFT's (Bracewell, 1978, Chapter 18)), which will sample the $\delta\tilde{\tau}_g$, $\delta\tilde{t}_p$ space with at least these frequencies.) For a 100 sec observation with a spanned frequency range of 400 MHz, the initial spacings of $\delta\tilde{\tau}_g$ and $\delta\tilde{t}_p$ will be at most 0.6 nsec and 0.6 psec/sec (at X-band). Because the fringe rate spectrum and multiband delay function are oversampled, we are assured that the initial search values, which yield the maximum value of the multiband delay function, are within half of the search value separation, of the peak of the delay resolution function, i.e. we know the peak cannot be located nearer to any other values because this would require higher frequencies in the delay resolution function than can exist.

Once these initial values are found, the maximization is carried out with one-dimensional fits, and re-computation of Equation (B.3.4) at the fitted values, repeated three times. The peak of the delay resolution function will be reached when $\partial R/\partial\tau_g=0$ and $\partial R/\partial\tilde{t}_p=0$ (these conditions are satisfied by the above technique), provided $\partial^2 R/\partial^2\tau_g+\partial^2 R/\partial^2\tilde{t}_p > \partial^2 R/(\partial\tau_g\partial\tilde{t}_p)$, (Hildebrand, 1976, pp. 356-357). This latter condition is satisfied by the delay resolution function and may be

confirmed by direct differentiation of the function.
Consequently, the converged values of the one dimensional fits
will be at the maximum of the delay resolution function.

When we introduced the singleband delay, the reason
stated was for computational efficiency. But, there is a
second reason for the introduction of this delay. Consider
the effects, on the cross spectral phases, of adding a piece
of (ron-dispersive) cable (or any delay introducing device),
between the first LO and the IF distributor (see Figure
2.1.2). The arrival time of the signal at the sampler (see
Section 2.1) would be later than in the case where the addi-
tional cable was not present. All of the cross spectral
phases would change by an amount given by the product of their
intermediate frequency (because the cable was inserted between
the first LO and the IF distributor) and the delay through the
cable. In addition, the phase calibration phases would also
change by the product of their IF frequency by the delay.
When we introduced the singleband delay function (Equation
(B.3.3)), we discussed that maximization of this function
essentially allows us to replace the fourteen or seven
(depending on whether dual or single sideband data are used),
cross spectral phases in each recorded channel with a slope
(the singleband delay) and an offset (the phase at video DC).
Clearly, this slope will be affected by the additional cable
and hence the singleband delay estimate will change. However,
the change in phase at video DC will be the same as the change
in the phase calibration phase at the correponding RF and

hence when the phase calibration is applied (see Equation(B.3.4)), the calibrated phases (at video DC) will be the same as these phases before the cable was inserted. (This "equality" is not exact because the phase calibration signals will have a frequency 10 kHz higher than that of the video DC. Provided the delay introduced by the cable is less than 277 nsec, the difference between the video DC phase and the phase calibration phase will be less than 1 degree.) Since the multiband delay is estimated from the changes in the video DC phases between channels, the multiband group delay will not be affected by this additional cable.

We can now extend these arguments to show that all delays in the receiver and cables, after the injection of the phase calibration pulses, are calibrated by the phase calibration system (if this system is functioning correctly; see Section 3.2) and the multiband delay will be unaffected by these delays. The multiband delay measures the difference in arrival times of the signals from the radio source at the injection point of the phase calibration pulses. On the other hand, the singleband delay measures all of the delays through the receiving system. The singleband delay measures the difference in arrival times of the signals at the samplers. Consequently, the singleband and multiband delays measure the difference in arrival times at different locations. If the delay through the receiver remains constant then there should be just a constant offset between the singleband and multiband delays.

The location in the receiver to which the singleband and multiband delays are referred could be made the same, assumming that the contribution of the receiver electronics to the cross-spectral phases varied linearly with frequency over the 2 MHz recorded bandwith, by extracting the phase of the calibration rail at the mid video frequency (i.e. 1.010 MHz). This technique has not been adopted in the design of the Mark III 'see Section 2.3 for discussion).

We should also note that the group delay will be determined predominantly from the changes in the phase between channels (which will span 100-400 MHz). The additional information about the group delay which would be supplied by the change in phase across only 2 MHz will only have a small effect on the knowledge of the group delay from the channel to channel variation in phase.

B.4 Analysis of the effects of the approximations made in implementing the delay resolution function

We have now completed the discussion on the algorithms used to estimate the group delay, the phase delay, and the phase delay rate. We did, however, make a number of approximations while deriving these algorithms. The approximations were :

    1) The use of only 8 lags in calculating the cross spectrum from the cross correlation function

    2) The use of 3 level sine/cosine functions to do

rotations.

3) The neglect of the rotation across the video band
while accumulating the complex cross correlation
function and while summing the delay resoution
function.

4) The accumulation of the cross correlation function for
the duration of the accumulation period (up to 2
seconds) using an _a priori_ phase delay rate which may
not be correct.

These approximations will be analyzed in the order
listed. We will be concerned with their effects on the multi-
band group delay, the phase delay and the phase delay rate.

In Figure B.1.1 we saw that the cross correlation func-
tion decayed rapidly as the offset betweeen the tapes was
moved away from the actual delay at the epoch of recording.
But is eight lags (0.25 μsec apart) enough to calculate ade-
quately the cross spectrum from the cross correlation func-
tion? For the initial analysis of this problem we will assume
that the power spectral density of the video signals are
independent of frequency over the 2 MHz bandwidth of each
channel. For this case we may calcuate the cross correlation
in analytic form and compute the cross spectrum from the cross
correlation function at the eight lags computed by the
correlator. We may then directly compare these estimates of
the cross spectrum components with the actual cross spectrum
to evaluate the errors introduced by using only eight lags.

This problem was analyzed by Whitney (1974), but he

considered the errors introduced by using seven lags (the number of lags used in the Mark I correlator). For his analysis he also assumed that the actual group and phase delays were zero. This assumption is not realistic because even if the delay were zero at one epoch in an observation it would be nonzero at other times because of the (nonzero) phase delay rate. For a phase delay rate of 2 μsec/sec, the delay would change by 200 μsec during a 100 second observation.

The cross correlation function of the USB video signals is given by Equation (B.1.7). For the analysis of the effects of using only eight lags to compute the cross spectrum we will simplify the form of Equation (B.1.7) by assuming that $\tau_g = \tau_p$ and $\Delta\phi^u(t) = 0$. We will also replace ω_1 by ω_r (the RF of the $\underline{r}$th channel from which the data were obtained), which is consistent with the notation of Equation (B.3.2), (see also Table B.2.1). (We will investigate the effects of relaxing these assumptions after we have carried out an initial analysis of the problem.) In its simplified form, Equation (B.1.7) becomes

$$R_{12}^u(\tau) = \cos(\omega_r \tau_g(t) + \pi B(\tau_g(t)-\tau)) \frac{\sin(\pi B(\tau_g(t)-\tau))}{\pi B(\tau_g(t)-\tau)}$$

(B.4.1)

where $\tau_g(t)$ represents the delay at time $\underline{t}$, i.e. we have incorporated the phase delay rate term into the delay $\tau_g(t)$.

To compute the cross spectrum, the correlation function was evaluated at eight lags about the bit shift closest to the $\underline{a\ priori}$ estimate of the delay. The cross spectrum, computed

for discrete values of video frequency between 0.25 MHz and
1.75 MHz, and spaced at 0.25 MHz intervals, was determined by
Fourier transforming the correlation function evaluated at the
eight lags. The phases and amplitudes of the cross spectral
components were then used to estimate the phase at video DC
(this phase should equal the phase at RF) and the group delay
across the 2 MHz bandwidth, i.e the singleband delay (see
discussions in this appendix concerning the maximization of
Equations (B.3.1) and (B.3.2)).

We will use this simple model to study some of the
properties of the errors introduced by using eight lags. We
will then study this problem using a correlation function
which more closely matches the actual properties of the
signals recorded on tape. The results obtained using this
second approach will be compared with the results from actual
correlation tests. All of these tests will be carried out
initially for USB data only. We may then easily deduce the
effects of this approximation (using only eight lags) on the
LSB data and dual sideband data.

Equation (B.4.1) has two important properties. Firstly,
the cosine expression contains a very rapidly changing term:
$\omega_r \tau_g(t)$. For $\omega_r = 8$ Ghz and $\dot{\tau}_p = 1 \times 10^{-6}$, this rapidly changing
term will rotate through a full cycle in 125 μsec, corres-
ponding to correlation of 500 bits from each tape (during a 2
sec accumulation period, 8×10^6 bits will be correlated).
Secondly, even if the a priori values of the group delay and
phase delay rate were known exactly, the difference $\tau_g(t) - \tau$

could not be kept constant because τ can only be changed at integral multiples of the sampling rate.

In Figure B.4.1, we show the bounds of the phase errors at video DC (see discussion of the singleband delay function (Equation (B.3.3)) for reasons for computing the phase error at video DC), for values of $\tau_g(t)-\tau_{ap}$, where τ_{ap} is the a priori estimate of $\tau_g(t)$, ranging between -0.8 and +0.8 μsec. These phase errors at video DC are a direct measure of the error in the visibility phase. In the insert to the figure we show for a very small section the actual behavior of these phase errors. Each cycle of the phase error will be sampled $\approx$500 times for 8.0 GHz data and $\approx$2000 times for 2.0 GHz data. The samples over each cycle will therefore average to the mean value of the bounds. The mean value curve is shown as the dashed curve in Figure B.4.1.

During correlation, $\tau_g(t)-\tau_{ap}$ will not remain fixed at one location on the mean value curve because $\tau_g(t)$ is changing continuously while τ_{ap}^b, the nearest bit to the a priori delay, can only be changed discretely, i.e. during correlation τ_{ap}^b is used not τ_{ap} (see Table B.2.1 and Section B.2). The bit offset between the tapes is kept within ±0.125 μsec of the a priori delay as the data are correlated. The final phase error will thus be an average of the phase errors over the inteval $\pm$ .125 μsec about the error in the a priori delay. The average values of the phase errors are shown in Figure B.4.2.

It should be noted that the phase errors are anti-

Figure B.4.1 Bounds on the phase errors introduced by using only
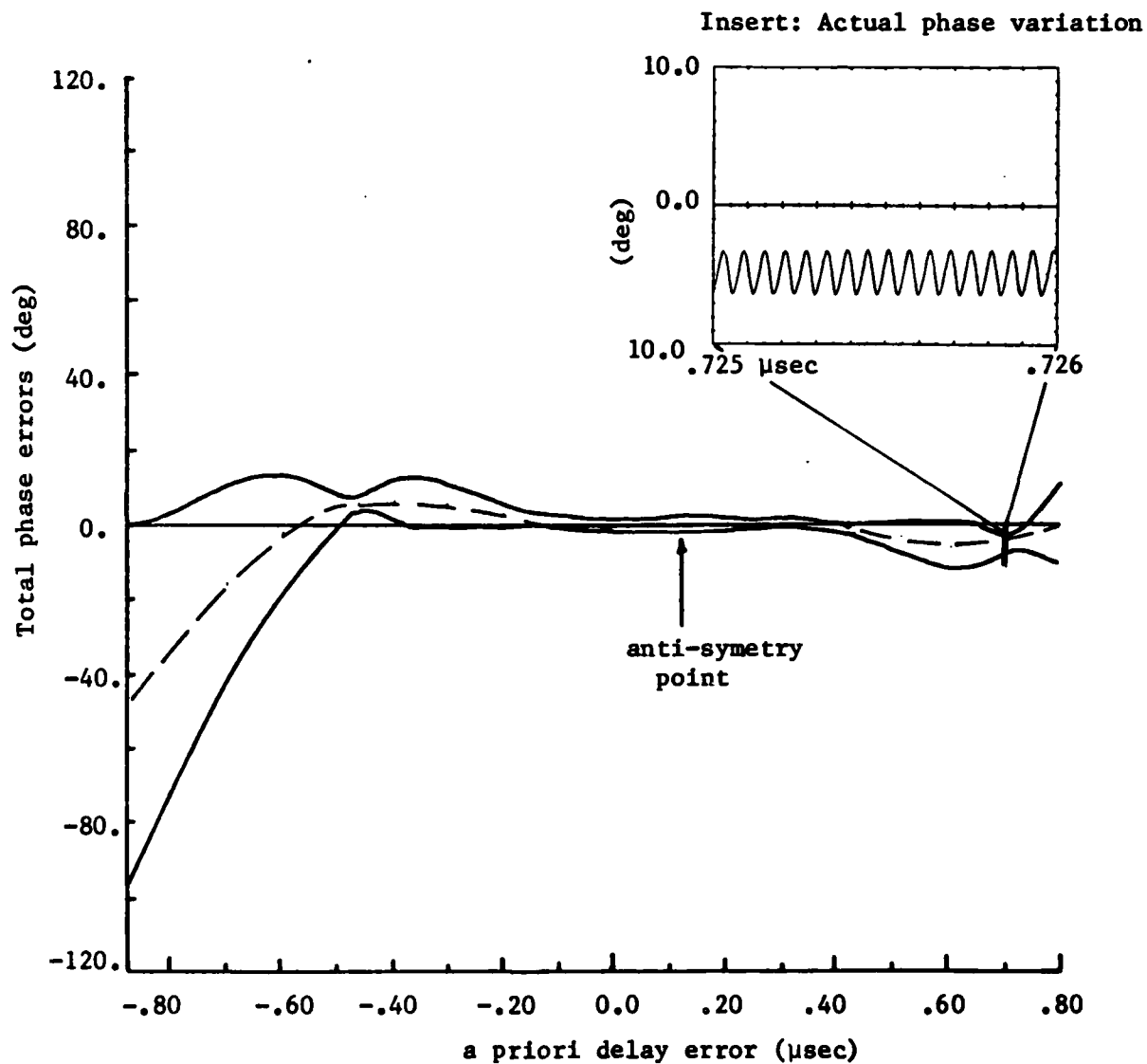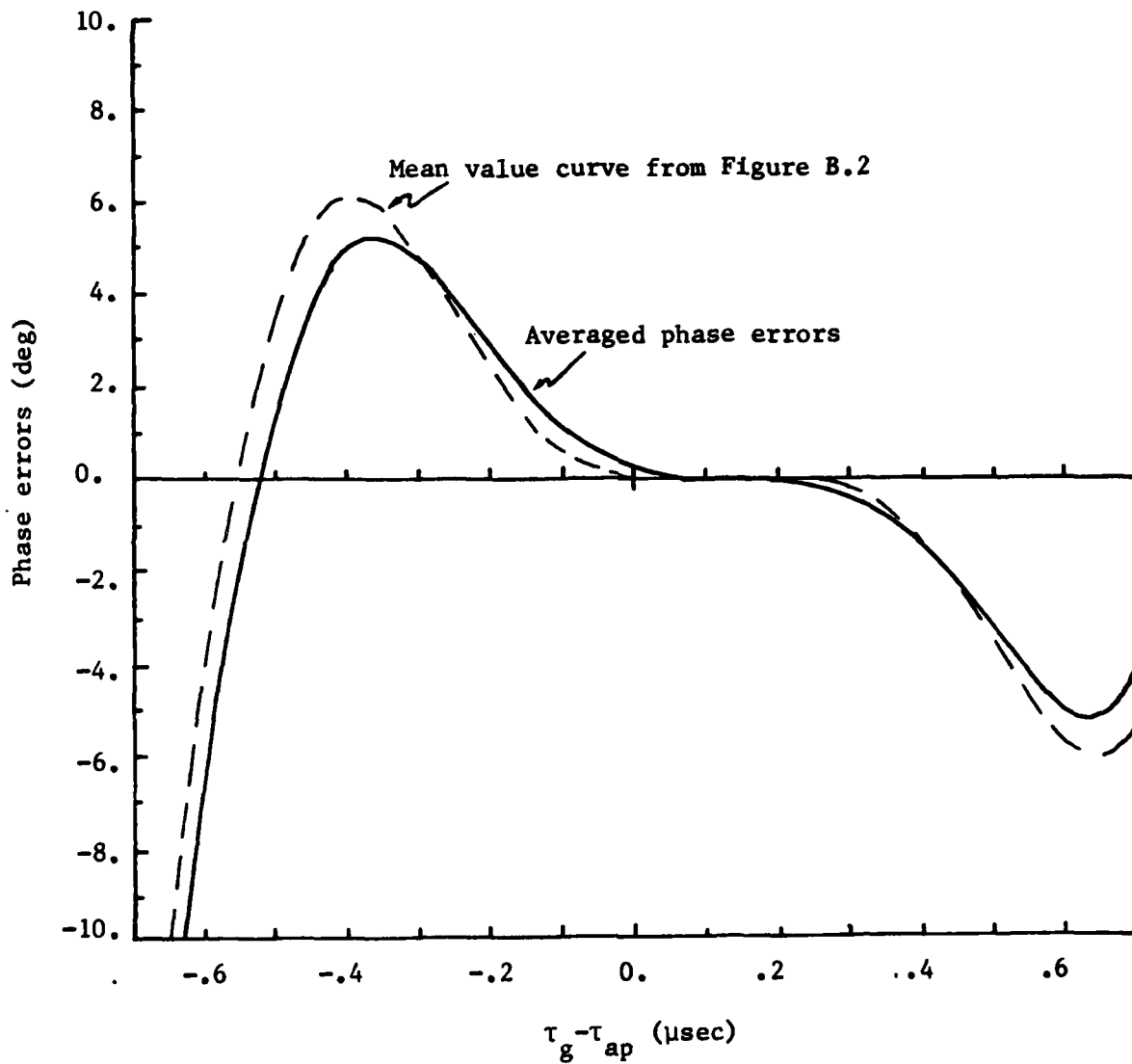8 lags to compute the cross spectrum.

Figure B.4.2 Averaged phase errors due to using 8 lags during correlation
For 2 MHz perfect band pass filter

symmetric about 0.125 μsec, not about zero. The asymmetry of the error about zero will lead to phase delay "closures errors" around triplets of baselines if the a priori delays, when referred to a common epoch, sum to zero around the triplet. (We discuss such delay closure errors in more detail in Chapter 3.) The asymmetry can be removed by using an odd number of correlator lags.

The phase errors shown in Figure B.4.2 should be easily detected by correlating a pair of tapes a number of times using different values for the a priori delay each time. This series of correlations was carried out by A.E.E. Rogers at the Haystack Observatory. Before we present the results of this test we should re-evaluate our model for the phase errors, trying to match our model to the real Mark III system as closely as possible.

The first improvement we can make to the model is to replace the perfect (i.e., "flat") 2 MHz bandpass filter with a more realistic filter. The Mark III video converters use a seven pole Butterworth filter (see, e.g., Mason and Zimmerman, 1960) to filter the output of video mixers before the signals are clipped and sampled. This filter will determine the properties of the power spectrum of the recorded signals. The transfer function for the Butterworth filter is given by (Mason and Zimmerman, 1960, p.426)

$$H(\omega) = \prod_{j=1}^{P} (i\nu - S_j)^{-1}$$

where $S_j = \exp[-i\pi(P+2j-1)/(2P)]$ and represents the poles of

the filter, P is the number of poles (seven in our case), $\nu = \omega/\omega_b$ and ω_b is the fullwidth at the 50% (-3db) power points for the filter output. For the Mark III video converters $\omega_b = 1.8$ MHz. The seven pole Butterworth filter power response is shown in Figure B.4.3. When the Butterworth filter is used the cross spectrum, $C_u^b(\omega_v)$, (the superscript $\underline{b}$ denotes that this is the cross spectrum which would be obtained when Butterworth filters are used), of the recorded data will be

$$C^b(\omega_v) = H_1(\omega_v)H_2^*(\omega_v) \exp[\ i(\omega_v \pm \omega_r\tau_g(t)]\qquad\text{(B.4.2)}$$

where $H_1(\omega_v)$ and $H_2^*(\omega_v)$ are the transfer functions of the Butterworth filters at the two sites. The above equation is simply an extension of Equation (B.1.5.a) with $|\bar{S}_1^u(\omega_v)|$, (the response of the "flat" band pass filter), replaced by the response of the Butterworth filters at the two sites. Because we are currently assuming that $\tau_g = \tau_p$, we have absorbed the phase delay rate into a time dependent delay (the group and phase delay are assumed equal (for this current discussion), hence the group delay rate will equal the phase delay rate). The upper sign is for $\omega_v > 0$ and the lower sign is for $\omega_v < 0$.

We can simplify the computations for this problem by remembering that when the correlator accumulates the complex cross correlation function, the rotations which are applied to compensate for the phase delay rate have the correct sign for $\omega_v > 0$ only, for USB. (For LSB data, the correct sign for $\omega_v < 0$ is used.) Consequently, the spectral components from USB data with $\omega_v < 0$ will not increase in magnitude proportional to the

Figure B.4.3 Seven pole Butterworth filter power response
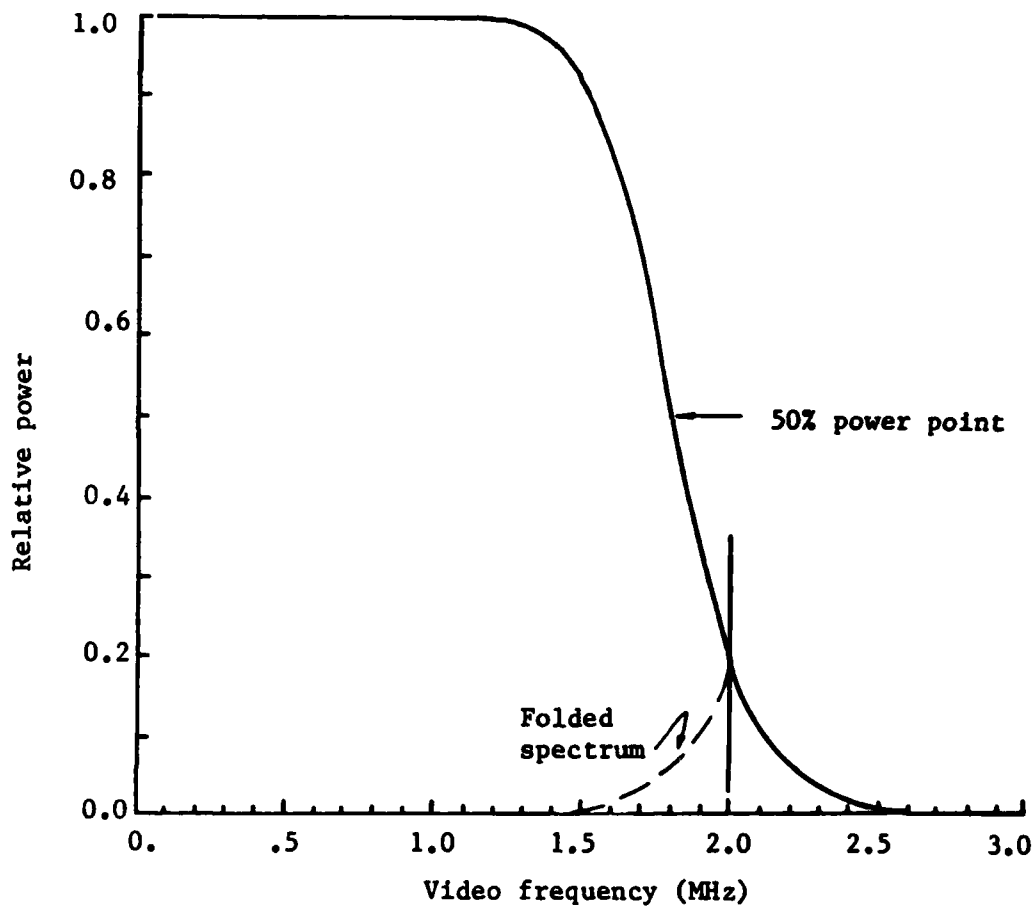
number of bits correlated (N(q); see Table B.2.1), as the data
are accumulated. The USB spectral components with $\omega_v < 0$ will
be rotated in the wrong direction. We will now investigate
the magnitude of these terms after the accumulation of the
complex cross correlation function. The accumulation of the
complex cross correlation function can be viewed as coherently
summing cross spectral components computed from 4 μsec
segments of data (see Appendix A for discussion). Many of the
errors introduced into the estimates of the group delay, phase
delay rate and visibility phase by the approximations made in
developing the estimation algorithms, can be studied by ana-
lyzing the effects of these approximations on the phase of the
summed cross spectral components. We will now develop a
simple formula which allows us to estimate the magnitude of
the cross spectral phase errors due to the rotation of a cross
spectral component at an incorrect rate. The coherently
averaged cross spectrum was given by (Equation (A.15)):

$$\bar{C}_q^a(\omega) = \sum_{j=1}^{n(q)} \bar{X}_{1j}^* \bar{X}_{2j} e^{\pm i(\alpha_{ap})_j}$$

where $(\alpha_{ap})_j$ is the $\underline{a~priori}$ expected change in the cross
spectral phase from the beginning of the accumulation period.
We may use the above formula to study the error in the
accumulation of the cross spectrum by replacing the summation
with an integration and replacing $(\alpha_{ap})_j$ by the error in the
rotation of the cross spectrum. These substitutions yield

$$\Delta C(\omega) = A \int_0^T ac~e^{+i\Delta\omega_e t}~dt \qquad (B.4.3)$$

where $\Delta C(\omega)$ is the accumulated cross spectrum, A is the amplitude of $\bar{X}_{1j}^{*}\bar{X}_{2j}$ (assumed constant), T_{ac} is duration of the accumulation period and $\Delta\omega_e t$ is the error in the rotation of the cross spectral components at time $\underline{t}$ (corresponding to index $\underline{j}$ in the summation). ($\Delta\omega_e$ will be used as a general term to denote rotation rate errors. We will use it in several different contexts in the remainder of this appendix.) The magnitude of $\Delta C(\omega)$ will be

$$|\Delta C(\omega)| = 2A \sin^2(\Delta\omega_e T_{ac}/2)/\Delta\omega_e \qquad (B.4.4)$$

We should notice that as $\Delta\omega_e$ approaches zero, $|\Delta C(\omega)|$ approaches AT_{ac} as expected (i.e., if the cross spectral components are rotated at the correct rate, then their coherent sum will equal their algebraic sum.)

We now return to the problem of the USB spectral components with $\omega_v < 0$. The error in the rotation of these cross spectral components, expressed as a frequency, will be $\Delta\omega_e = 2\omega'(\mathfrak{t}_{ap})_q$ (see Table B.2.1 for definitions). For a small phase delay rate of 100 psec/sec (this value will be exceeded most of the time for baselines >10 km long), $(\omega'/2\pi) \approx 2$ GHz, and $T_{ac} \approx 2$ sec (typical duration of an accumulation period), we have $|\Delta C(\omega)|/|\bar{C}_q^a(\omega)| = 8\times10^{-3}$, where $\bar{C}_q^a(\omega)$ is the corresponding USB cross spectral component with $\omega_v > 0$. This ratio is a direct measure of the expected phase error (in rad) of the USB cross spectrum with $\omega_v > 0$, due to the USB cross spectral components with $\omega_v < 0$.

Although the phase error given above is very small ($0.^{\circ}5$),

we will continue to analyze this problem because the analysis
will provide us with some techniques which will be useful
later in this appendix. So far, we have calculated the phase
error in a single accumulation period. But when the delay
resolution function is computed, we will sum together many
accumulation periods (see Equation (B.3.4). How do we compute
the errors after this summation? There are two techniques we
could adopt. We could compute the phase errors in each
accumulation period, at each cross spectral frequency, and
then estimate, by least squares, the errors introduced by
these phase errors into the estimates of the multiband group
delay, the singleband group delay, the phase delay rate and
the visibility phase. (See Appendix A and earlier discussions
in this appendix.) This technique is useful if we want to
correct the measurements for a known error (this technique was
used in the initial study of the errors introduced by using
only eight lags). However, if we only wish to bound the
effect of an error, this technique may be "too exact", i.e. it
may involve more effort than is warranted for determining a
bound on an error. We can adopt a simpler technique. To
demonstrate the application of this technique, we show, in
Figure B.4.4, the aggregate cross spectral component, ΔC_{obs},
(at a single frequency), after several accumulation periods
have been summed. This figure shows that the aggregate cross
spectrum will be

$$\Delta C_{obs}(\omega) = A \int_0^T{}^{obs} e^{+i\Delta\omega_e t} dt \qquad (B.4.5)$$

Figure B.4.4 Accumulation of cross spectral components as accumulation
periods are summed
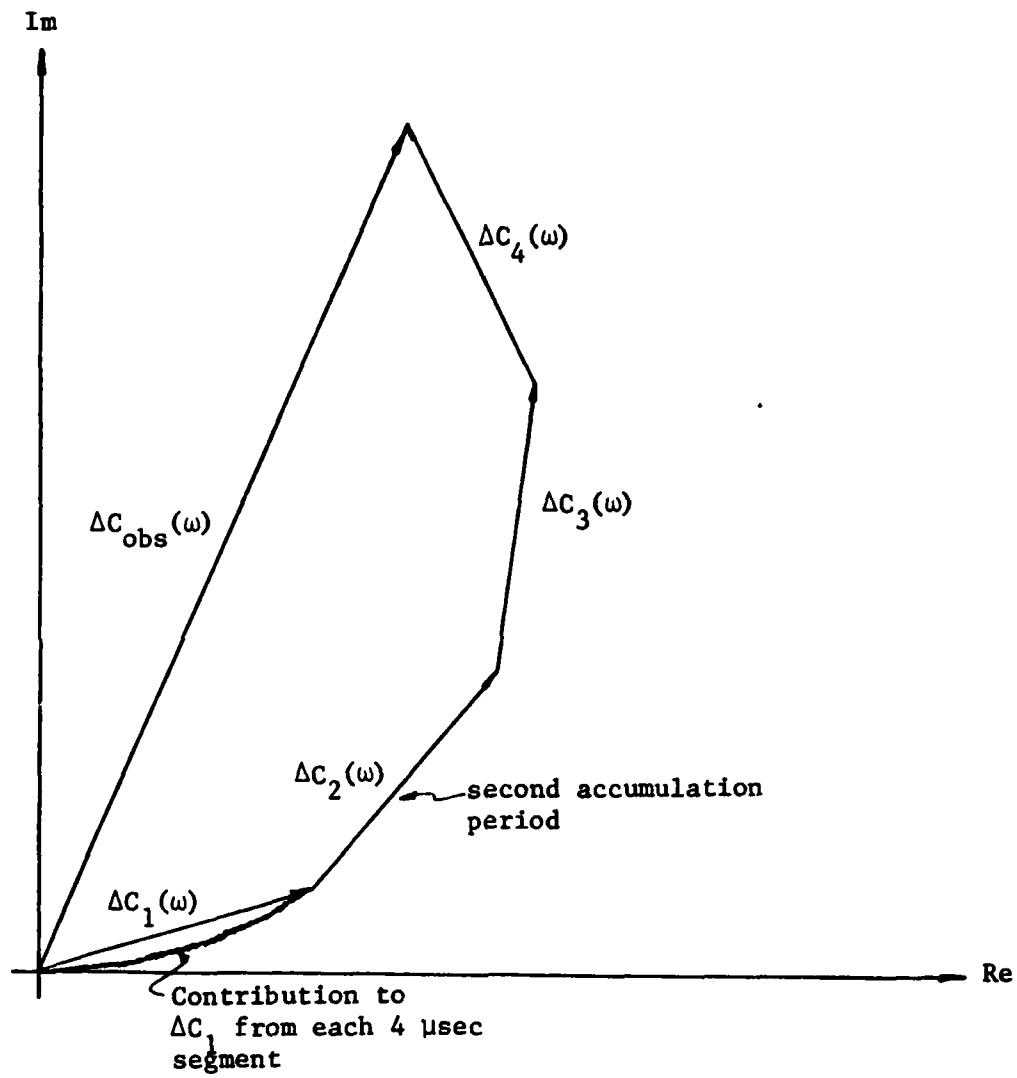
where T_{obs} is the duration of the observation. The magnitude of ΔC_{obs} will be $|\Delta C_{obs}(\omega)| = 2A \sin^2(\Delta\omega_e T_{obs}/2)/\Delta\omega_e$. We can now estimate that the magnitude of the USB spectral components with $\omega_v < 0$, relative to the USB spectral components with $\omega_v > 0$, after a 100 sec observation, will be 50 times smaller than the value quoted earlier. ($|\Delta C(\omega)|/|C_q^a(\omega)| = 1.6 \times 10^{-4}$ for $(\omega'/2\pi) = 2$ GHz and $t_{ap} = 100$ psec/sec.)

To form the complex cross correlation function we therefore only need to integrate $C_u^b(\omega_v)$ from zero to infinity.

The procedures used to evaluate the errors introduced by using only eight lags in conjunction with the seven pole Butterworth filter were the same as discussed for the perfect bandpass except that the complex cross correlation was directly evaluated by numerically integrating, rather than analytically integrating, the cross spectrum. When the complex cross correlation function is directly evaluated it is no longer necessary to average over a cycle of $\omega_r \tau_g(t)$. To see why this is so we need to return to the real (in the complex sense) cross correlation function (Equations (B.4.1)). The only difference between the calculations of the real and the complex cross correlation functions is the limits of the integration used in Fourier transforming the cross spectrum. The real cross correlation function is obtained by integrating the cross spectrum from $-\infty$ to ∞, and represents the expectation of the cross correlation function of a short segment of data (4 μsec duration) from the two sites. The complex cross correlation function is obtained by integrating the cross

spectrum from 0 to ∞ and represents the cross correlation function obtained from the coherently averaged cross spectrum. The phase term, $\omega_r\tau_g(t)$, appears in the cosine term in Equation (B.4.1) and changes the shape of the correlation function as $\tau_g(t)$ varies (see Figure B.1.1.a). These changes in the shape of the cross correlation function will affect the errors introduced by using only 8 lags (see insert in Figure B.4.1 and accompanying explanation). The changes in $\omega_r\tau_g(t)$ will also affect the complex cross correlation function, but in this case the shape of the correlation function will be unchanged as $\tau_g(t)$ changes. This constancy of sphape may be easily seen by writing the expression for the complex cross correlation function. $R_u^C(\Delta\tau_i)$ will be given by

$$R_u^C(\Delta\tau) = e^{i\omega_r\tau_g(t)} \int_0^\infty H_1(\omega_v)H_2^*(\omega_v)e^{i\omega_v\tau_g(t)}e^{-i\omega_v(\Delta\tau_i+\tau_{ap})}\, d\omega_v$$

(B.4.6)

where we have included τ_{ap} because the lag is taken about this value, i.e. $\tau = \tau_{ap}+\Delta\tau_i$ (see Table B.2.1 and Section B.2).

We should notice that $\omega_r\tau_g(t)$ appears outside of the integration and hence will only rotate (in the complex plane), the complex cross correlation function, and not change the shape of this function. (If we replace the Butterworth filter transfer functions in the above equation, with the transfer function of a perfect bandpass filter (i.e., $H_1(\omega_v)=1$ for $-B<\omega_v<B$ and zero elsewhere) we see that $R_u^C(\Delta\tau_i)$ reduces to

$$R_u^C(\Delta\tau) = e^{i\omega_r\tau_g(t)+i\pi B(\tau_g(t)-\Delta\tau_i-\tau_{ap})}$$

$$\text{sinc}(\pi B[\tau_g(t)-\Delta\tau_i-\tau_{ap}])$$

(B.4.7)

where $\text{sinc}(\alpha)=\sin(\alpha)/\alpha$. We see from the above equation that $R_u^c(\Delta\tau_i)$ is a sinc function rotated in the complex plane by $\omega_r\tau_g(t)+\pi B(\tau_g(t)-\Delta\tau_i-\tau_{ap})$.
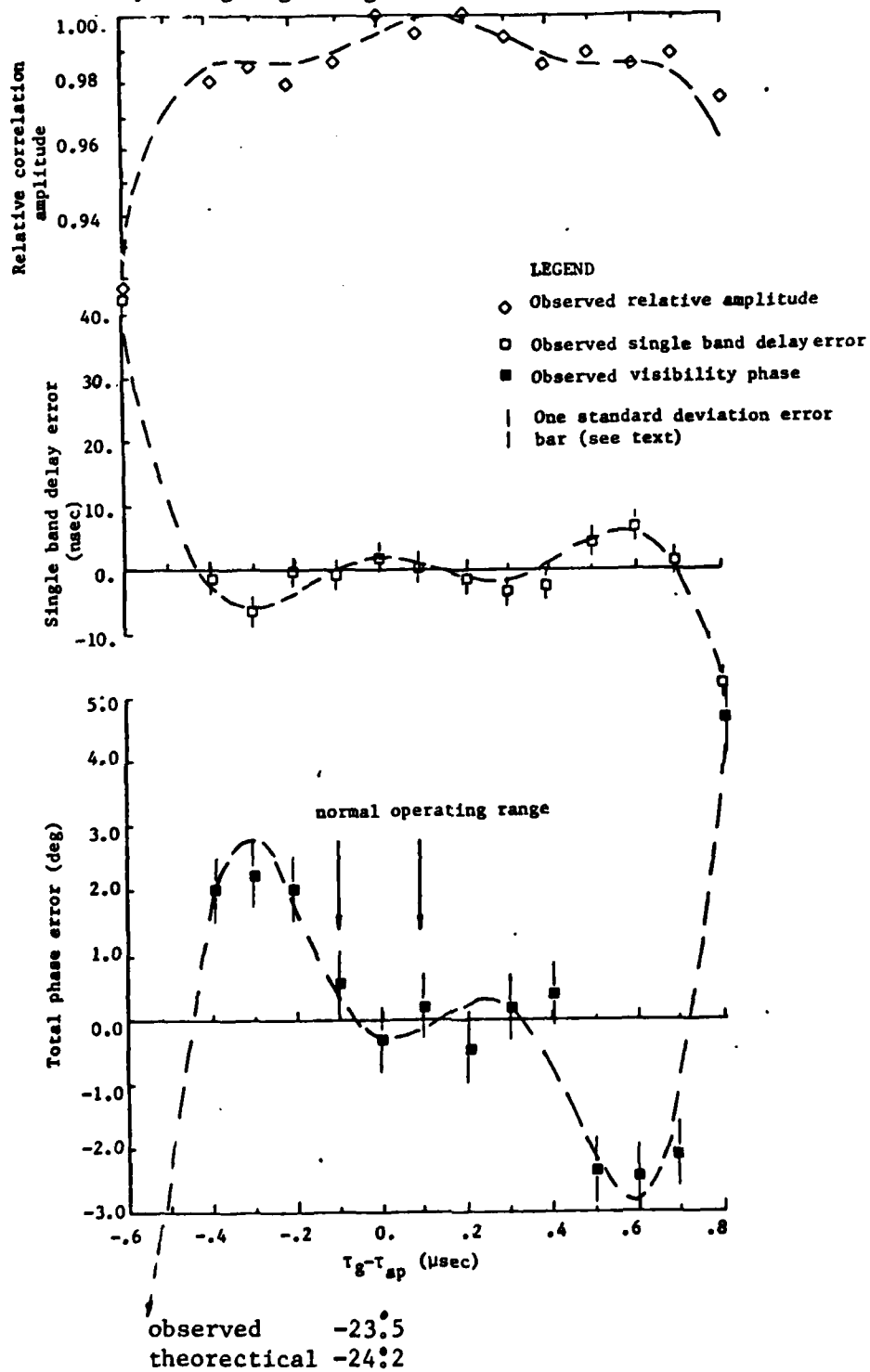
The cross spectral components which are generated from the complex cross correlation function, evaluated at eight lags, will only rotate as $\tau_g(t)$ is varied by small amounts, of order $2\pi/\omega_r$, i.e. each cross spectral component will be multiplied by $e^{-i\omega_r\tau_g(t)}$ independently of its video frequency because $e^{-i\omega_r\tau_g(t)}$ can be taken outside of the calculation. The amplitude and phase errors in the cross spectral components will remain the same over these small variations in τ_g because the shape of the cross correlation function will not change. (The shape of the correlation function will change with variations in τ_g of the order of $1/B$).

The results obtained using the complex cross correlation function should be exactly the same as those obtained from the real cross correlation function after we have averaged over a cycle of $\omega_r\tau_g(t)$.

In Figure B.4.5, the estimated errors in the relative correlation amplitude, the singleband delay and the visibility phase are shown for errors in the a priori delay that vary from -0.6 µsec to 0.8 µsec. In addition we also show the results obtained from the repeated correlations of a Westford-Fort Davis, USB, S-band observation of the radio source 3C 273B on July 7, 1982 at 02:34 UT. The experimental and theoretical results match reasonably well although there are some differences. (These differences will be discussed

Figure B.4.5 Comparison of the theorectical and observed errors introduced
by using eight lags

LEGEND
◇ Observed relative amplitude
□ Observed single band delay error
■ Observed visibility phase

| One standard deviation error
| bar (see text)

normal operating range

observed   -23°.5
theorectical -24°.2

after we have studied the multiband group delay results.)

For the geodetic applications of VLBI we are interested in the multiband group delay errors. What magnitude errors do we expect because of the use of 8 lags? We should expect zero errors because the phase errors in each channel are independent of the radio frequency. (The only term involving the radio frequency is $\omega_r \tau_g(t)$. As we have just seen this term can be taken outside of the calculations (with the exception of small fringe rate observations (see discussion associated with Equation (B.4.3)), and hence does not affect the errors introduced by using only eight lags.) In Figure B.4.6, we present the variations in the estimates of the multiband group delay results found from the repeated correlations. Clearly the observed variations are not zero. Over the normal range of ±0.1 µsec error in the a priori delay, the multiband delay changed by 50 psec.

To investigate these observed multiband group delay variations, we need to firstly look at the "performance" of the correlator, i.e. had we repeated these correlations, not changing the a priori delay, what variations would we have seen? The answer unfortunately is not zero, even though the correlator is a digital system and hence should be able to exactly reproduce results. The reproducibility problems seem not to be with the correlator (in general) but with reading the magnetic tapes. (In some cases, hardware problems do develop in the correlator modules causing nonreproducibility with a particular module with a hardware error. These cases

Figure B.4.6 Variations in the estimate of the multiband group delay as
a function of the _a priori_ delay error- S band
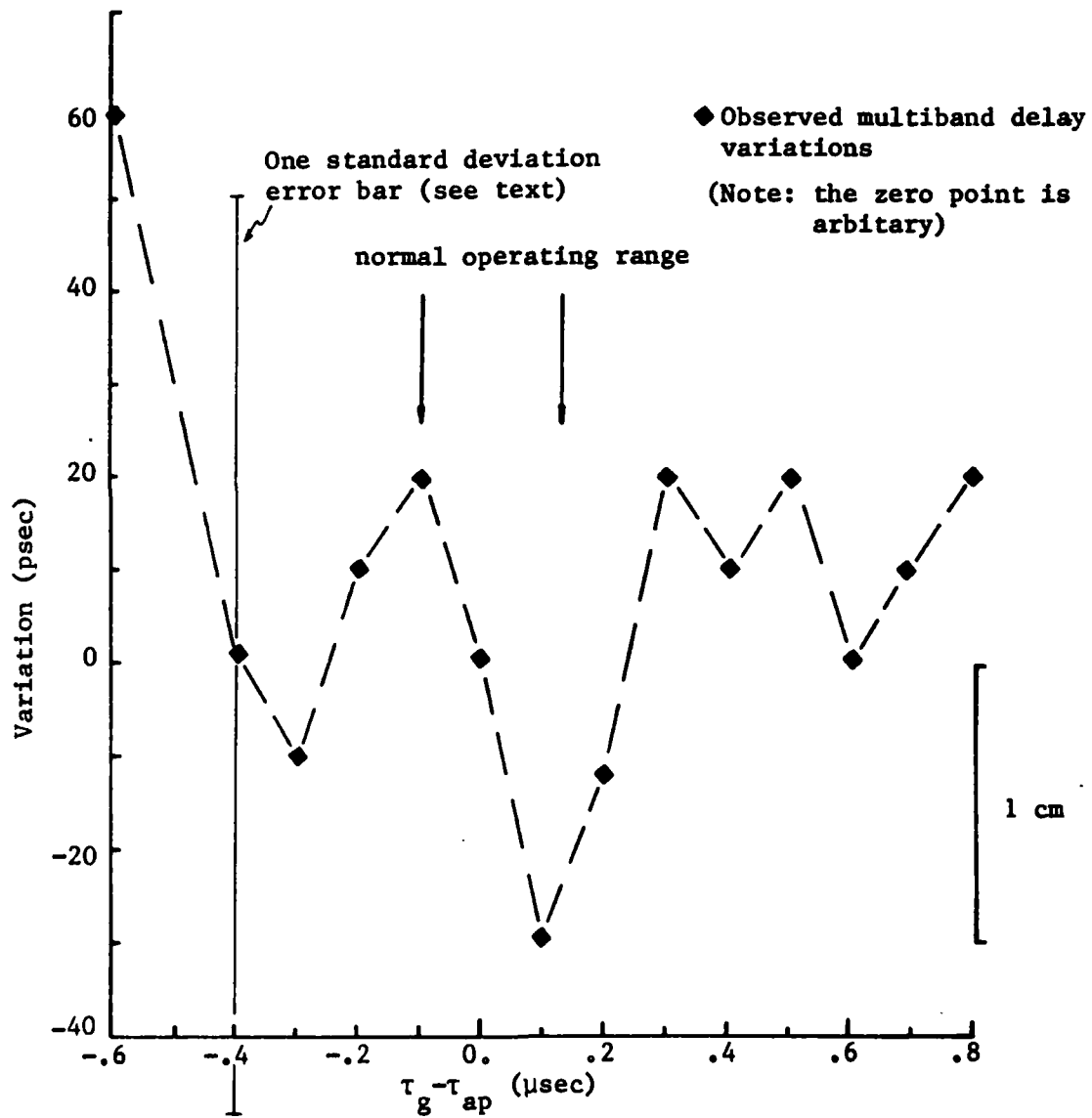
are rare and will not be discussed here.) To indicate the magnitude of nonreproducibility (i.e., the differences between measurements when tapes are recorrelated using the same a priori parameters), we show in Figure (B.4.7), a histogram of differences between the multiband group delays obtained from repeated correlations of 21 observations of the radio source pair 1038+52A and B (Marcaide, 1982). We have plotted the histogram with bins defined by the difference in estimates in the multiband group delay divided by their standard deviation. (This form was chosen because earlier reproducibility results had indicated that the nonreproducibility was a function of the standard deviations of the estimates.)

If we now reexamine Figure B.4.6 and compare the results with the possible variations in the multiband group delay with repeated correlations (Figure B.4.7), we see that the variations in Figure B.4.6 can be explained by correlator non-reproducibility. To further study this problem, the X band results from these repeated correlations (at different lags) were analyzed. These results are shown in Figure B.4.8. Again we see variations in the multiband group delays which are consistent with the nonreproducibility of the results. Also the X-band variations show no correlation with the S-band variations.

To confirm that the effects of errors in the a priori delay are zero (or very small), we could repeat these tests and average the results of the repeated correlations. To reduce the pseudo-noise due to correlator nonreproducibility,

Figure B.4.7 Histogram of the differences between the multiband group
delay estimates from repeated correlations

Obtained from pairs of correlations of the radio sources
1038+528A and 1038+528B, March, 1980.

Differences in standard deviations

Mean difference $-0.04 \, \sigma$

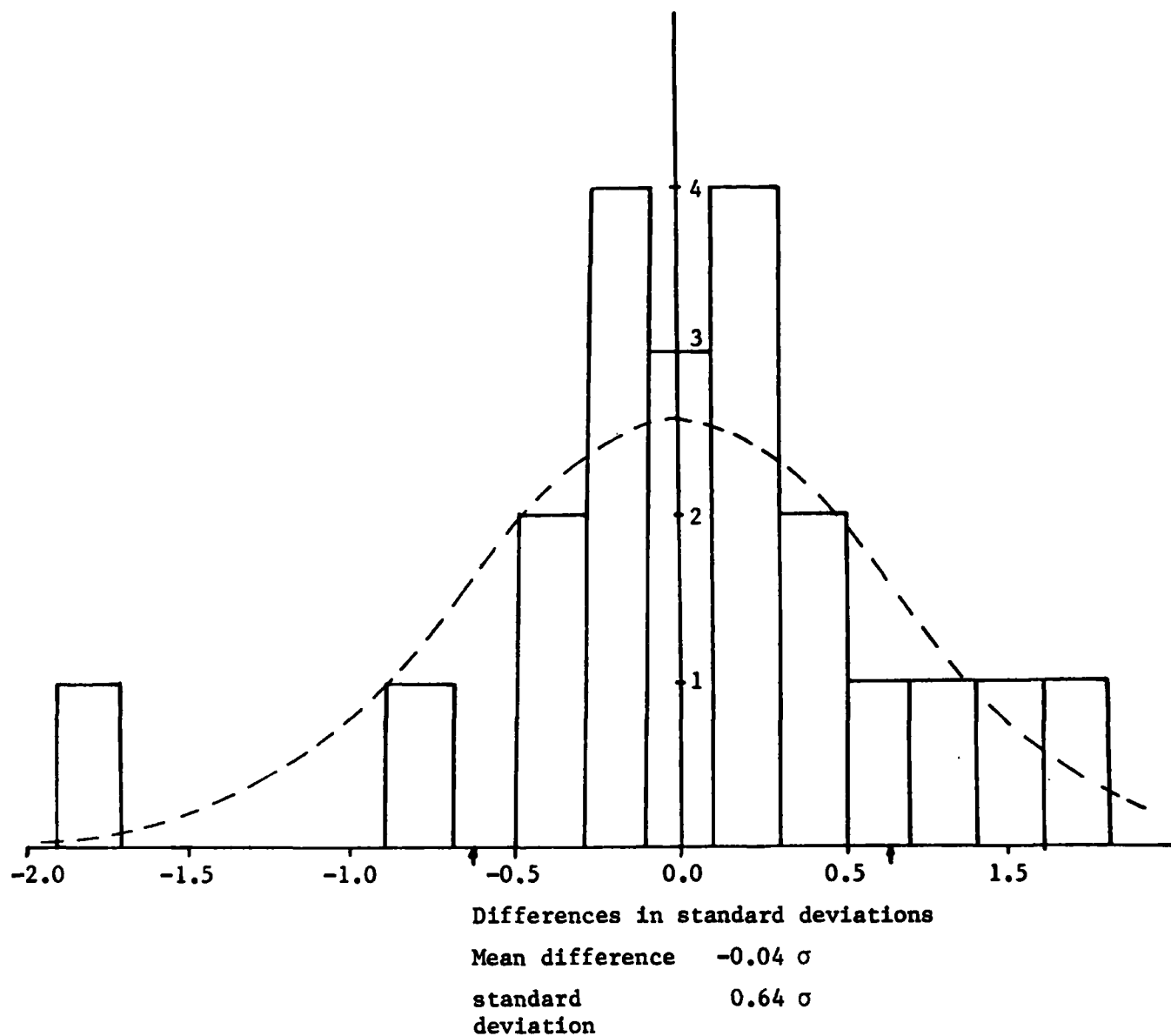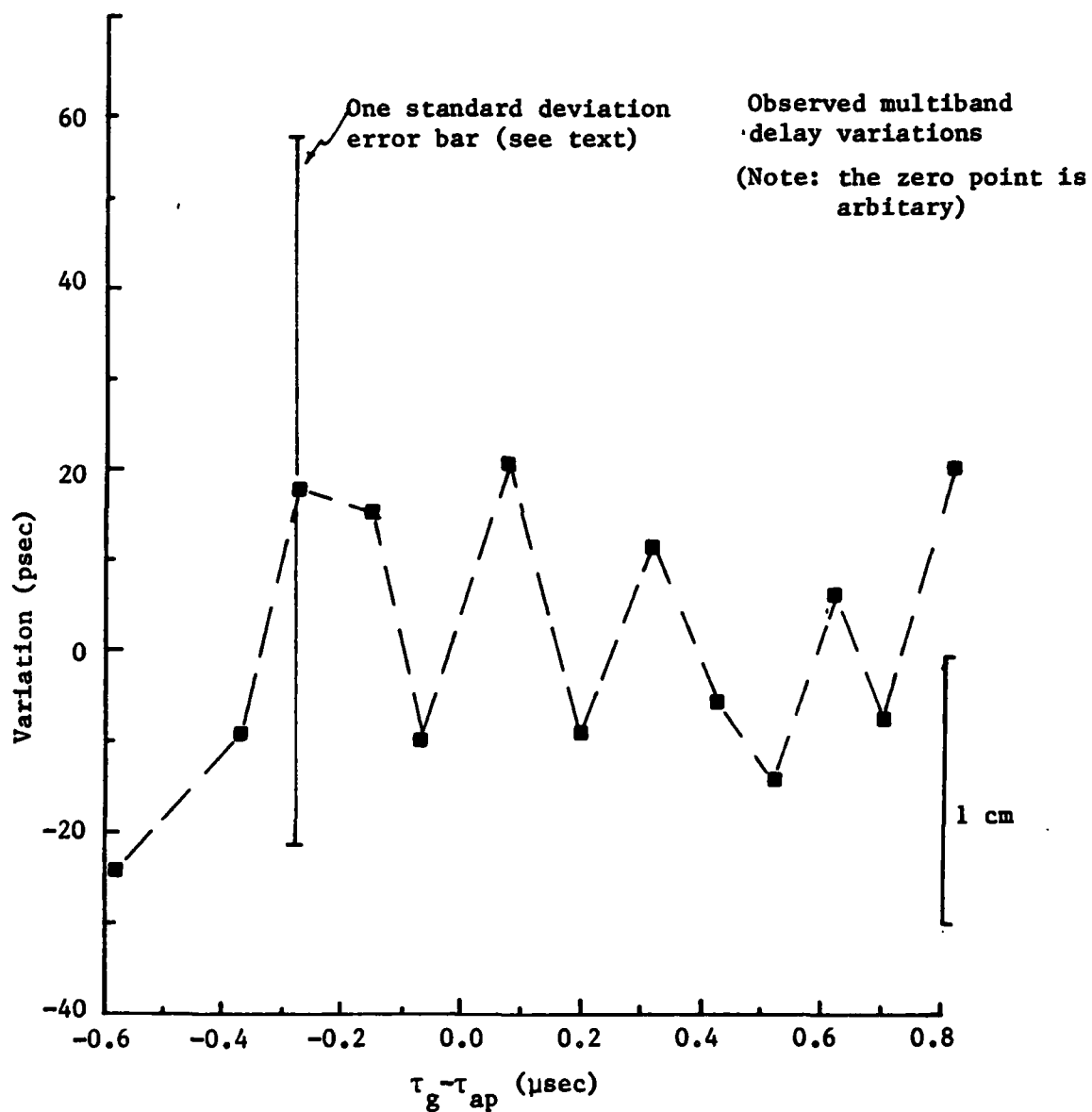standard $0.64 \, \sigma$
deviation

Figure B.4.8 Variations in the estimates of the multiband group delay as
a function of the a priori delay error - X band

to 10 psec, would require at least 25 repeated correlations at each of the 14 different values of the a priori delay. These 350 repeated correlations would be a large committment of (limited) resources.

There may be a better approach. In May, 1983 a geodetic experiment was carried out using the 100 m diameter antenna at Effelsburg in the Federal Republic of Germany. The multiband group delay estimates from baselines involving this antenna can have standard deviations of a few picoseconds. We should use these data to analyze the effects of errors in the a priori delay. This analysis will not be possible before the completion of this thesis.

The correlator nonreproducibility would also explain the differences between the observed variations and the predicted differences seen in Figure B.4.6. Again, the repeated correlations of baselines involving the Effelsburg antenna should confirm if this is the actual reason for these differences.

All of the results presented so far have assumed that only USB data were available. We will now consider the effects of using eight lags when only LSB data are available. For the simplified case we are currently studying (see earlier discussion), the lower sideband cross spectrum, $C_\ell^b(\omega v)$, is given by (see Equation (B.1.5.b))

$$C_\ell^b(\omega_v) = H_1(\omega_v) H_2^*(\omega_v) \ \exp[i(\omega_v \mp \omega_r) \tau_g(t)] \qquad (B.4.8)$$

where the upper signs refer to $\omega_v > 0$ and the lower signs to

$\omega_v<0$. We have written the cross spectrum in the same form as in Equation (B.4.2). Only the negative video frequencies from the LSB are used in the computation of the cross spectrum (see discussion concerning Equations (B.3.1) and (B.3.2)), and in the discussion on the complex cross correlation function associated with Equation (B.4.2). A review of the discussion associated with Figure 2.1.3. may also be useful at this time. For $\omega_v<0$ the cross spectrum becomes

$$C_\ell^b(\omega_v) = H_1(\omega_v)H_2^*(\omega_v) \; \exp[-i|\omega_v|+i\omega_r\tau_g(t)]$$

where we have explicitly shown the sign associated with ω_v by using its absolute value in the exponent. We would form the complex cross correlation function from this cross spectrum by

$$R_\ell^C(\Delta\tau) = e^{i\omega_r\tau_g(t)} \int_0^\infty H_1(-|\omega_v|)H_2^*(-|\omega_v|) \; e^{-i|\omega_v|\tau_g}$$

$$e^{i|\omega_v|(\Delta\tau_i+\tau_{ap})} \; d|\omega_v| \quad . \qquad (B.4.9)$$

If we compare this equation with the equation for the upper sideband complex cross correlation function (Equation B.4.6)), we see that they are very similar (as we should expect) except for some sign changes. To investigate the effects of these sign changes, we will analytically evaluate $R_\ell^C(\Delta\tau_i)$. To do this we will replace the Butterworth filter responses with a flat bandpass filter (this replacement will give us an equation which we can directly compare with (Equation B.4.7)).

With the flat bandpass filter, $R_\ell^C(\Delta\tau_i)$ becomes,

$$R_\ell^C(\Delta\tau_i) = e^{i\omega_r\tau_g(t)-i\pi B(\tau_g(t)-\Delta\tau_i-\tau_{ap})}$$

$$\text{sinc}[\pi B(\tau_g(t)-\Delta\tau_i-\tau_{ap})] \qquad (B.4.10)$$

In Figure B.4.9, we show a graphical comparison of $R_\ell^C(\Delta\tau_i)$ and $R_u^C(\Delta\tau_i)$, (Equation B.4.7). We should notice that the magnitudes of $R_\ell^C(\Delta\tau_i)$ and $R_u^C(\Delta\tau_i)$ as functions of $\Delta\tau_i$ are exactly the same; the arguments, however, have opposite sign slopes (as functions of $\Delta\tau_i$), intersecting $\Delta\tau_i=0$ at $-\omega_r\tau_g(t)+\pi B(\tau_g(t)-\tau_{ap})$ and $-\omega_r\tau_g(t)-\pi B(\tau_g(t)-\tau_{ap})$, respectively. (These intercepts may appear to have incorrect signs. This change of sign occurs because the Fourier transform of these correlation functions will generate the complex conjugates of the cross spectra (see discussion at the beginning of this appendix).)

When the estimates of the cross spectrum are calculated from the complex cross correlation functions, the contribution of the cross correlation function at each lag, (rotated by $e^{i\omega_v\Delta\tau_i}$), to the summation will be a constant rotation (independent of the video frequency being computed, i.e. $-\omega_r\tau_g(t)-\pi B(\tau_g(t)-\tau_{ap})$, for USB, and $-\omega_r\tau_g(t)+\pi B(\tau_g(t)-\tau_{ap})$, for LSB), plus a rotation which will have opposite signs for the upper and the lower sidebands ($[\omega_v-\pi B]\Delta\tau_i$ and $-[|\omega_v|-\pi B]\Delta\tau_i$, respectively). (Remember, only $\omega_v>0$ should be used with $R_u^C(\Delta\tau_i)$ and $\omega_v<0$ with $R_\ell^C(\Delta\tau_i)$.) Consequently, the phase errors of the cross spectral components (i.e., the difference between cross spectrum phase at ω_v and correct phase

Figure B.4.9 Magnitude and argument of the upper and lower sideband complex cross correlation functions
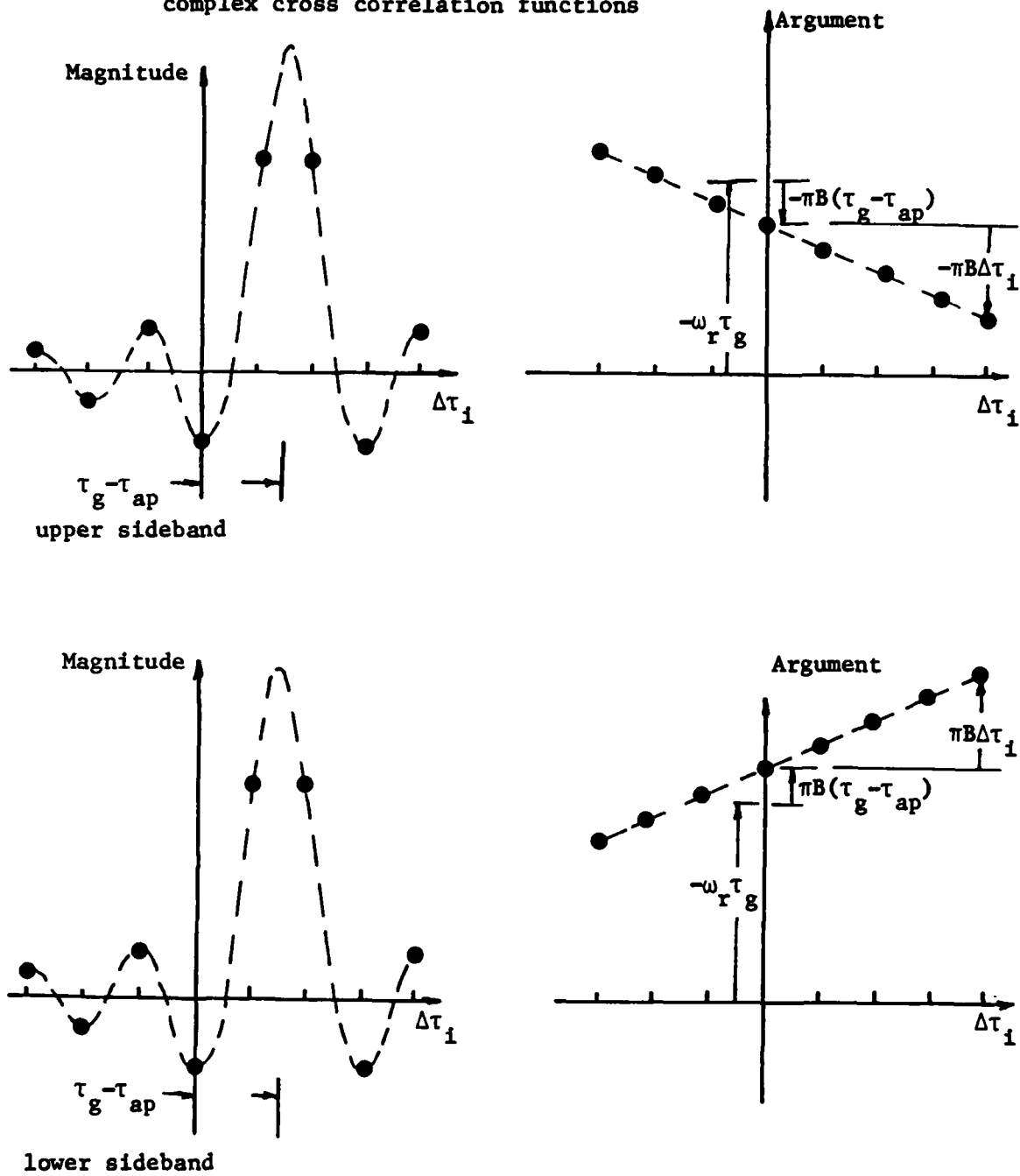
at ω_v (given by $\omega_r\tau_g(t)+\omega_v(\tau_g(t)-\tau_{ap})$), from upper and lower sidebands will have opposite signs (see Figure B.4.10).

Figure B.4.10 easily allows us to deduce, from the USB results, the effects of using only eight lags, when only LSB data are available. The singleband delay errors will be the same for each sideband. The phase error at video DC, which is the visibility phase error, will have opposite sign for upper and lower sideband data. When USB and LSB data are combined (in equal proportions), the visibility phase errors will be zero (this may be seen from Figure B.4.10; the weighted least squares fit of a slope and intercept through the cross spectral phases from both side bands will intercept $\omega_v=0$ at the correct visibilty phase). (Normally, the phase of the cross spectrum of heterodyned signals, at the $\omega_v=0$ boundary, is not continuous (i.e., the phase changes sign for real signals); this discontinuity is not the case in Figure B.4.10 because this figure does not represent the pure cross spectrum of a heterodyned signal (see discussion about Equation (B.1.1)).)

It should be noted that if the a priori delay is in error by more than 0.8 μsec, then the correlation amplitude will drop rapidly and the wrong peak in the delay resolution function could be selected when estimating the multiband group delay. If the wrong peak is selected then the visibility phase and the estimates of the quantities will be useless.

At the beginning of this discussion we assumed that $\tau_g=\tau_p$ and that $\Delta\phi^u(t)=0$. If these two assumptions are removed, will

Figure B.4.10 Cross spectral phases from upper and lower sidebands
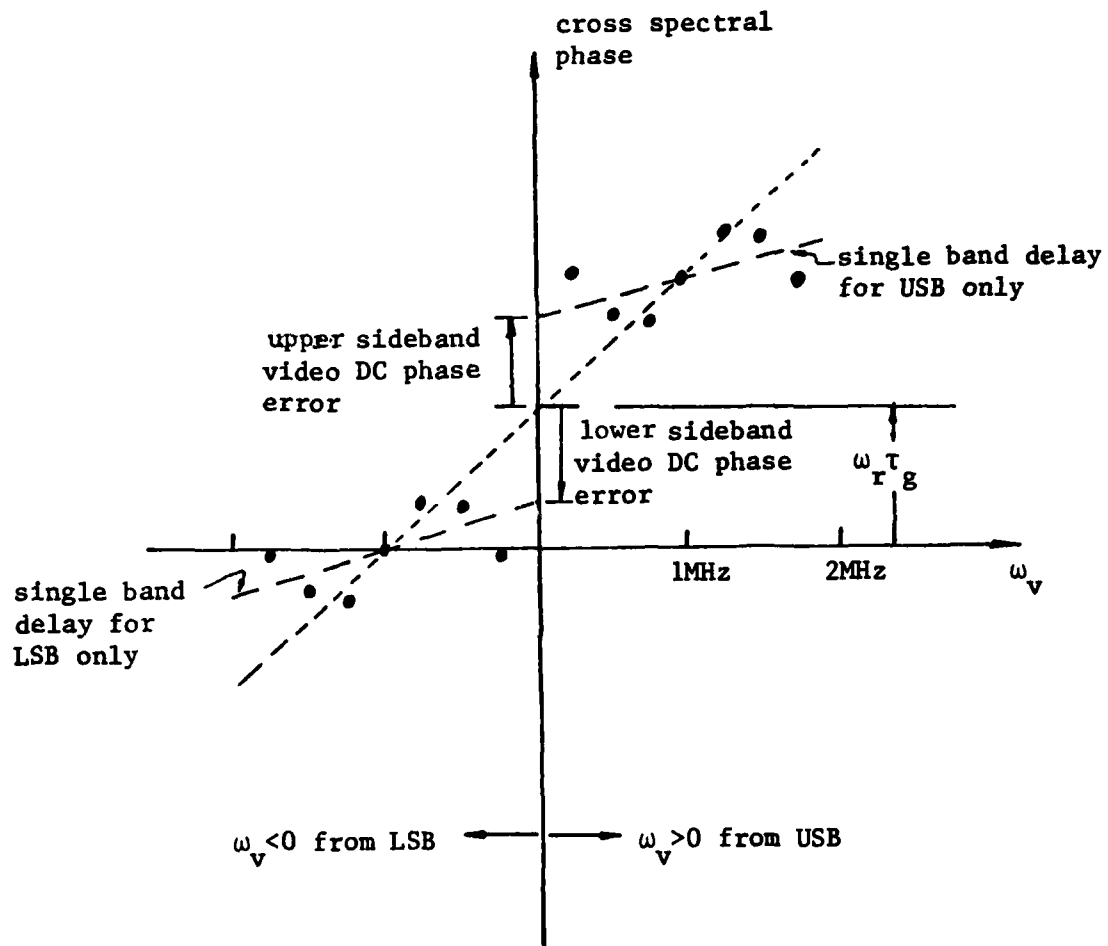
our conclusions be changed? No. The only effect will be to introduce an additional phase term to the exponent outside of the integration used to compute the complex cross correlation functions (see Equations (B.4.6) and (B.4.9)). Since these additional terms are independent of $\Delta\tau_i$ and ω_v, there presence will not affect the calculation of the errors.

Before finishing with the effects of using only eight lags we should stress that the estimated errors will only be valid if the phase delay rate is sufficiently large that $\omega_r \dot{\tau}_p \Delta t$ will rotate through several cycles during the observation (typically >100 seconds). For S-band data, the phase delay rate would need to be greater than 100 psec/sec to ensure 25 cycles of rotation of $\omega_r \dot{\tau}_p \Delta t$, in which case the visibility phase errors in Figure B.4.6 could be incorrect by 2.6 deg (see discussion on the complex cross correlation function (Equation (B.4.5)).

The next error we will discuss is the use of the 3 level sine/cosine function to carry out rotations (this function is shown in Figure B.4.11). This problem can be viewed very elegantly from a Fourier series approach. If we refer back to Equation (B.2.7) we see that the complex cross correlation function $R_u^C(\Delta\tau_i)$ was given by, (see Table B.2 for symbols)

$$R_u^C(\Delta\tau_i) = \sum_{p=1}^{N(q)} x_1(t_q^1 + p\Delta t_s) x_2(t_q^1 + (\tau_{ap}^b)_p + p\Delta t_s - \Delta\tau_i)$$

$$\exp[-i\omega'(\dot{\tau}_{ap})_q(t_p - t_q^C)]$$

Figure B.4.11   Three level sine/cosine approximation

$f_{ap} t$

$\pi/8$ π 2π 3π 4π 5π

1.   0.   -1.

$f_{ap} t$

$3\pi/8$ π 2π 3π 4π 5π

1.   0.   -1.
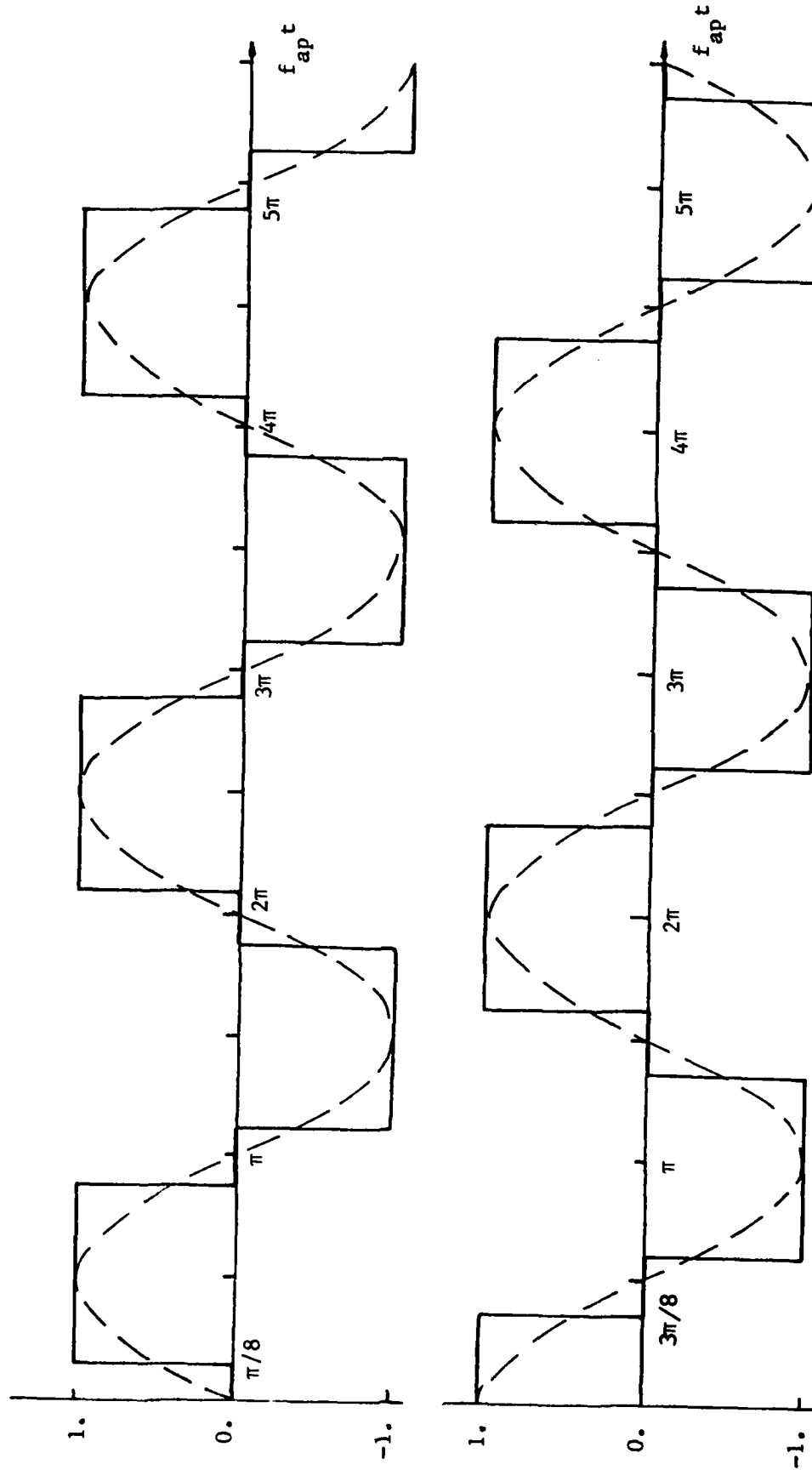
If we group terms in this expression by setting

$$r(p\Delta t_s, \Delta\tau_i) = x_1(t_q^1 + p\Delta t_s)x_2(t_q^1 + (\tau_{ap}^b)_p + p\Delta t_s - \Delta\tau_i)$$

and $f_{ap} = \omega'(t_{ap})_q$, then

$$R_u^C(\Delta\tau_i) = \sum_{p=1}^{N(q)} r(p\Delta t_s, \Delta\tau_i)\exp[-if_{ap}p\Delta t_s] \qquad (B.4.11)$$

Equation (B.4.11) should be recognized as the discrete Fourier series transform of $r(p\Delta t_s, \Delta\tau_i)$ at frequency f_{ap}. To determine the errors introduced by using the 3 level sine/cosine function we can use the convolution theorom of Fourier transforms (Bracewell, 1978, pp. 108-112). When the 3 level sine/cosine is used to evaluate Equation (B.4.11), the results obtained will be the product of the Fourier coefficient we want and the Fourier transform of the 3 level sine/cosine function. The 3 level sine/cosine function is periodic and hence can be expanded in Fourier series. The first five coefficients of the Fourier series for the 3 level sine/cosine functions are

Harmonic	coefficent		frequency
1	$(4/\pi)\cos(\pi/8)$	≈ 1.18	f_{ap}
2	0		$2f_{ap}$
3	$-(4/3\pi)\cos(3\pi/8)$	≈ -0.16	$3f_{ap}$
4	0		$4f_{ap}$
5	$(4/5\pi)\cos(5\pi/8)$	≈ -0.10	$5f_{ap}$

The first effect of using the 3 level sine/cosine function

will be to multiply the correlation amplitude by $(4/\pi)\cos(\pi/8)$ (i.e., an increase of $\approx 18\%$). This increase in amplitude is compensated after the summation (B.4.11) is completed by dividing the magnitude of the singleband delay function (Equation (B.3.3)) by $(4/\pi)\cos(\pi/8)$.

The increase in correlation amplitude is the major bias introduced by the 3 level sine/cosine approximation provided f_{ap} is sufficiently large that the $f_{ap}p\Delta t_s$ rotates through a large number of cycles during an observation. (See discussion below.)

For small fringe rates, the Fourier series coefficents of the 3 level sine function will be different from the coefficent of the 3 level cosine function if these functions have not rotated through an integral number of cycles. The difference in the coefficents will introduce errors in both the visibility phase and the correlation amplitude.

We can investigate the magnitude of these phase and amplitude errors by calculating the difference between the Fourier series coefficients of the 3 level sine and cosine functions. After the 3 level sin/cosine has cycled through N_{sc} cycles, i.e. integer value of $f_{ap}T_{obs}/(2\pi)$, where T_{obs} is the duration of the observation, the maximum difference between the sine and cosine Fourier series coefficients will occur when $f_{ap}T_{obs}$ equals $2\pi(N_{sc}+1/8)$, (see Figure B.4.12). For this value of the argument of the 3 level sine/cos function, the Fourier series coefficients will be $[(4N_{sc}+1)\cos(\pi/8) - \cos(\pi/4)]/[\pi(N_{sc}+1/8)]$ and

Figure B.4.12 Differences between $(4/\pi)\cos(\pi/8)$ and the 3 level sine/cosine Fourier series coefficents
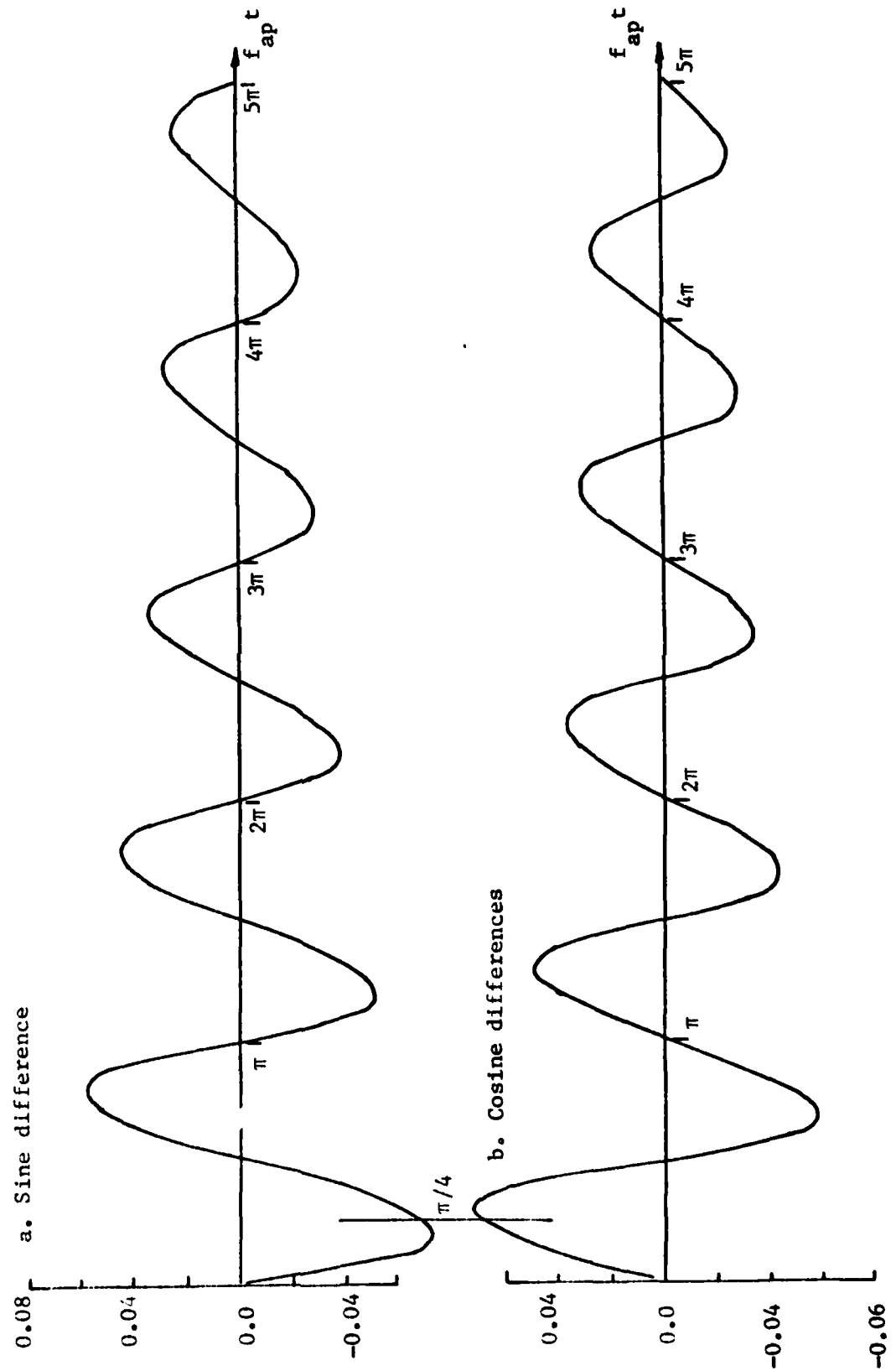
$[(4N_{sc}+1)\cos(\pi/8) + \sin(\pi/4)]/[\pi(N_{sc}+1/8)]$, for the 3 level sine and 3 level cosine functions, respectively. The maximum difference between the coefficients will be $[\cos(\pi/8)-2\sin(\pi/4)]/[\pi(N_{sc}+1/8)] \simeq -0.16/(N_{sc}+1/8)$. This difference could lead to cross spectral phase errors of up to 8° (0.14 rads), and correlation amplitude errors of 14%. But we also notice that the phase errors decay rapidly as $f_{ap}T_{obs}$ increases. Only on short baselines (<10 km length) will $f_{ap}T_{obs}$ be small for extended periods of time (more than one observation). For a phase delay rate of 100 psec/sec, ω_r=2.2 GHz, and T_{obs}=100 sec, the maximum cross spectral phase error due to the 3 level sin/cos approximation will be $0.^{\circ}4$.

The effects of the 3 level sine/cosine approximation on the multiband group delays may be calculated from the above results. The different fringe rates (i.e., f_{ap}), for each channel in the synthesized band will lead to different phase errors in each channel, and hence introduce multiband group delay errors. For the case given above and a synthesized bandwidth of 80 MHz (typical for S-band), the multiband group delay error could reach 0.030 nsec (again, only for short baselines will errors of this magnitude occur).

The third approximation we will study is the neglect of the video rotation term. While deriving the complex cross correlation function we neglected the term $\exp[-i\omega_{vj}t_p(t_p-t_q^c)]$, (see Equation (B.2.3)). This rotation represents changes in the phases of cross spectral components across the video frequency range due to the total phase delay

rate. Since ω' (see Table B.2.1), is chosen at the center of
the video band, the phase errors due to the neglect of the
video rotation will be zero for the frequency at the center of
the video band. For the other frequencies in the cross
spectrum there will, however, be errors introduced by neglect-
ing the video rotation.

The effects of the neglect of the video rotation term
were studied by Whitney (1974) for the Mark I VLBI system. A
sufficiently large number of the characteristics of the Mark I
system are different in the Mark III system for the effects of
this video rotation term to be re-evaluated. The maximum
error due to this term occurs in the video frequency
components near the edges of the bandwidth. The bandwidth of
the recording system increased from 360 KHz to 2 MHz as we
changed from Mark I to Mark III. This increase in bandwidth
will increase the error in the phase of the cross spectral
components near the edges of the bandwidth by a factor of 5.6,
i.e. by the radio of bandwidths. The length of an
accumulation period has also been increased by a factor of 10,
i.e. the accumulation period has changed from 0.2 sec to 2
sec. The combination of these two changes has increased the
error due to the neglect of video rotation by a factor of 56
over the values given by Whitney. If we use a maximum value
for the phase delay rate of 2×10^{-6} sec/sec and an accumulation
period of 2 sec, the error accumulated at the edge of the
video band (1 MHz from the center frequency) would be 4
cycles. This error is unacceptably large. Clearly the effect

of the error due to video rotation must be reduced.

To reduce this error we must first realize that the rotations applied to the data streams during correlation do not exactly compensate for a change in the arrival times of the signals at each site. Only the center frequency of the 2 MHz bandwidth has the correct rotation applied. The other frequencies in the cross spectrum are rotated by an angle which is error by $(\omega_v - \pi B) \dot{t}_{ap} (t_j - t_q^c)$. One method we could use to solve this problem would be to accumulate a cross correlation function for each of the video cross spectral frequencies whose components are be estimated. This approach however would require a sevenfold increase in the storage requirements in the correlator modules and in FRNGE. An alternative approach is to always keep the bit offset *between the tapes* within ±0.5 of the bit spacing (0.25 μsec). Each time a bit shift is applied there should be a 90° phase shift applied to compensate for the change in lag index. (See Equation (B.4.7); a change in τ_{ap} by 0.25 μsec will rotate $R_u^C(\Delta\tau_i)$ by $\pi B \Delta t_s$, which equals 90° (see Table (B.2.1)). The rotation is applied to compensate for this change. When we developed the correlator algorithms (Equation (B.2.7)) we included a term for changing the bit shift during correlation, $((\tau_{ap}^b)_j).$) The use of bit shifts does not eliminate the video rotation error.

The pattern of rotation errors of the cross spectral component at video frequency, ω_v, is shown in Figure B.4.13. Each discontinuity corresponds to a bit shift occurring, accompanied by the 90° phase rotation. The accumulated error

Figure B.4.13 Notation and geometry of the viseo rotation error

-374-
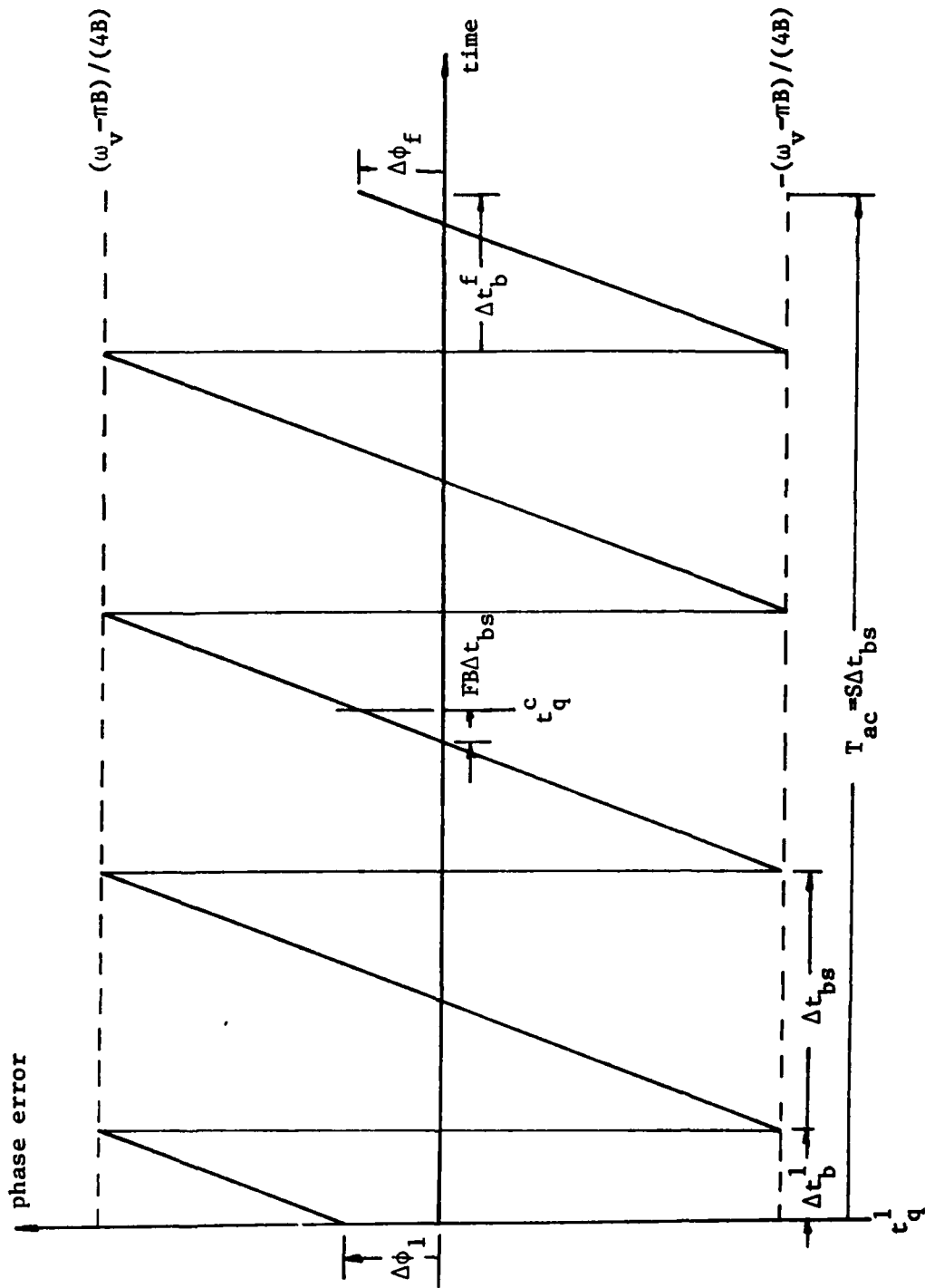
of the cross spectral component at ω_v will be the average of "saw tooth" phase errors shown in Figure B.4.13. This accumulated phase error, $\Delta\phi_{ac}$, will be given by

$$\Delta\phi_{ac} = \{[\frac{(\omega_v-\pi B)}{4B}+\Delta\phi_1]\Delta t_b^1/2 + [\frac{-(\omega_v-\pi B)}{4B} +\Delta\phi_f]\Delta t_b^f/2\} \ /T_{ac}$$

$$(B.4.12)$$

where $\Delta\phi_1$ and $\Delta\phi_f$ are the rotation errors at the beginning and the ends of the accumulation period, respectively, (see Figure B.4.13), Δt_b^1 is the interval of time between the beginning of the accumulation period and the first bit shift, and Δt_b^f is the interval of time between the last bit shift and the end of the accumulation period, B is the bandwidth and T_{ac} is the duration of an accumulation period (see Table B.2.1). Equation (B.4.12) can be reduced to a simpler form by noting that from Δt_b^1 and Δt_b^f, $\Delta\phi_1$ and $\Delta\phi_f$ can be calulcated. From geometry, in Figure B.4.13, we have

$$\Delta\phi_1 = (\omega_v-\pi B)/(4B) - (\omega_v-\pi B)\Delta t_b^1/(2B\Delta t_{bs})$$

and

$$\Delta\phi_f = -(\omega_v-\pi B)/(4B) + (\omega_v-\pi B)\Delta t_b^f/(2B\Delta t_{bs})$$

where Δt_{bs} is the interval of time between bit shifts (i.e., $\Delta t_{bs} = 1/(2t_{ap}B)$). After some simple algebra, Equation (B.4.12) reduces to

$$\Delta\phi_{ac} = \frac{\omega_v-\pi B}{4B} (\Delta t_b^1-\Delta t_b^f)[1-(\Delta t_b^1+\Delta t_b^f)/\Delta t_{bs}]/T_{ac} \qquad (B.4.13)$$

The above equation may be simplified further by expressing

Δt_b^1 and Δt_b^f as functions of the fractional bit shift at the center of the accumulation, FB, (FB= $\tau_{ap}^c/\Delta t_s - \| \tau_{ap}^c/\Delta t_s + .5 \|$), where τ_{ap}^c is the <u>a priori</u> delay at the center of the accumulation period, Δt_s is the sampling interval and $\| \ \|$ denotes the largest integer smaller than the quantity inside the symbols), and the "floating point" number of bit shifts in the accumulation period, S, (S = $(\dot{\tau}_{ap})_q T_{ac}/\Delta t_s$). (The actual number of bit shifts in an accumulation period will be $\|S\|$). From Figure B.4.13 we can deduce that

$$\Delta t_b^1 = (S/2 - N_{b1} - FB + 0.5)\ \Delta t_{bs} \qquad (B.4.14.a)$$

and

$$\Delta t_b^f = (S/2 - N_{b2} + FB + 0.5)\ \Delta t_{bs} \qquad (B.4.14.b)$$

where N_{b1} and N_{b2} are the numbers of bit shifts before and after the center of the accumulation period; $N_{b1} + N_{b2} = \|S\|$. By adding and subtracting the two Equations (B.4.14), we obtain

$$\Delta t_b^1 - \Delta t_b^f = (N_{b2} - N_{b1} - 2FB)\ \Delta t_{bs}$$

and

$$\Delta t_b^1 - \Delta t_b^f = (S - N_{b1} - N_{b2} + 1)\ \Delta t_{bs}.$$

Substitution of these results into Equation (B.4.13) yields

$$\Delta\phi_{ac} = \frac{\omega_v - \pi B}{2B}\ [FB - (N_{b2} - N_{b1})/2][S - (N_{b1} + N_{b2})]\ \Delta t_{bs}/T_{ac}$$

$$(B.4.15)$$

Finally, we can eliminate N_{b1} and N_{b2} from the above equation

by noting that

$$(N_{b2}-N_{b1})/2 = 0 \quad , \quad \text{if } |S| \text{ is even}$$
$$= 1/2 \ , \quad \text{if } |S| \text{ is odd and FB} < 0$$
$$= -1/2, \quad \text{if } |S| \text{ is odd and FB} > 0.$$

We will denote $(N_{b2}-N_{b1})/2$ by α_b, which yields for the error due to the neglect of video rotation

$$\Delta\phi_{ac} = \frac{\omega_v - \pi B}{2B} \left[FB - \alpha_b \right]\left[S - |S| \right) \right] \Delta t_{bs}/T_{ac} \quad . \qquad (B.4.16)$$

In Figure B.4.14, we show the error in the phase of the video DC cross spectral components which result from neglecting the video rotation as a function of the time measured in bit shifts, (one unit of bit shift time is Δt_{bs}). To generate this figure we assumed that $\Delta t_b^1 = 0$, (see earlier discussion). (The video DC frequency component is not computed from the cross correlation function but the error at this frequency will represent the visibility phase error after the singleband delay has been estimated, see discussion on Equation (B.3.3).)

As with earlier errors, we see that this error will only have a significant effect on observations with small phase delay rates. For a phase delay of 100 psec/sec, the visibility phase error, (i.e. the phase error, $\omega_v = 0$), could be, after a 100 sec observation, as high as $22\overset{\circ}{.}5$, (S=0.04 and hence it is unlikely that there will be any bit shifts in the observation). (We can extend the results from an accumulation period to the observation duration using the same arguments which were presented with Equations (B.4.3)-(B.4.5)). This is a very large error and steps have now been taken to eliminate

Figure B.4.14 Phase error at DC due to the neglect of video rotation
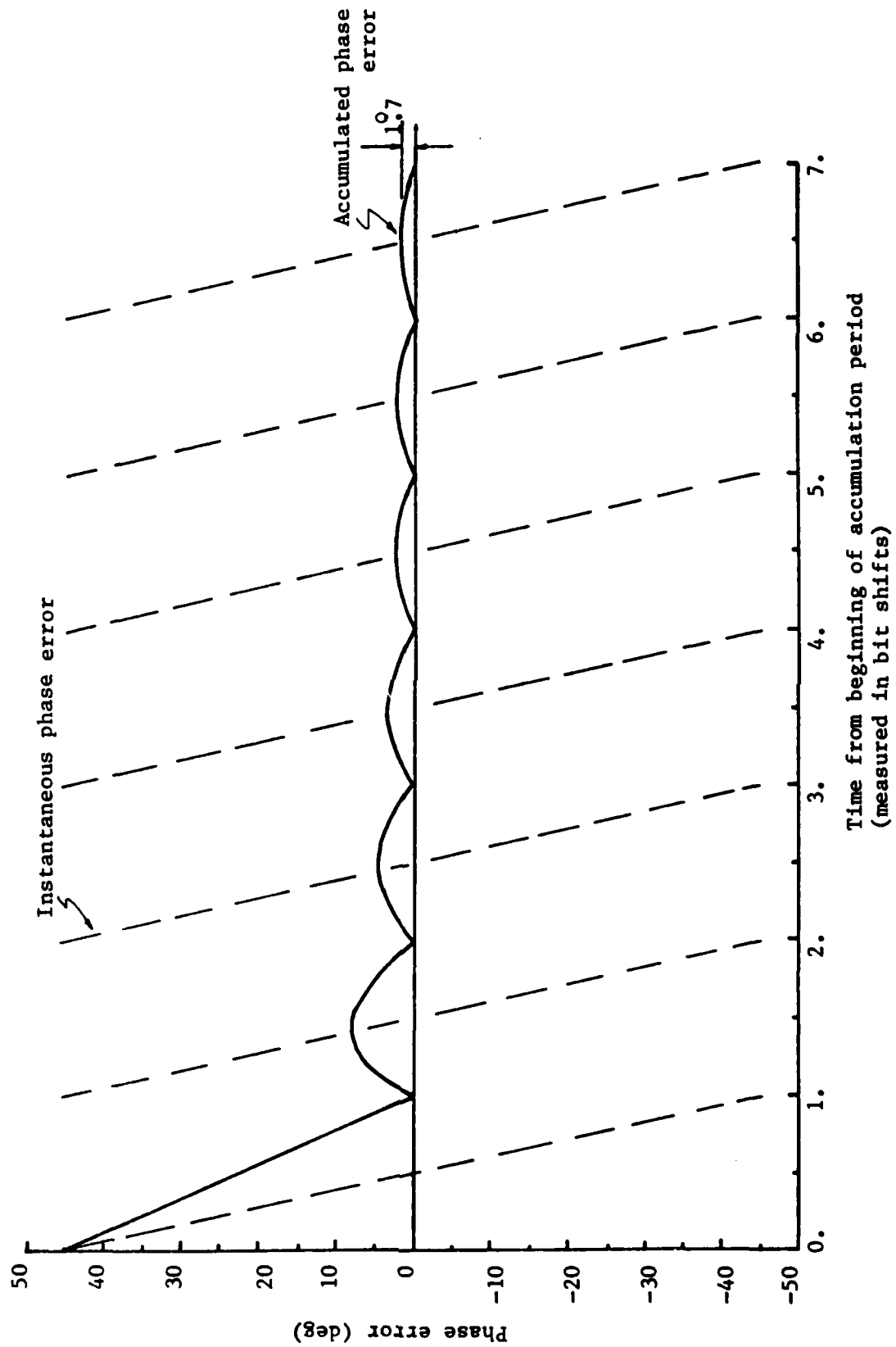
this error (see discussion later).

Dual sideband data will not be affected by this error, when the sidebands are combined in equal proportions, because the LSB error will have opposite sign to the USB error. (This sign change may be shown using arguments similar to those used when we were studying the effects of using eight lags; see Equations (B.4.8)-(B.4.10).) If the proportions of upper and lower sidebands are not equal, the error may be calculated from the weighted mean of the upper and lower sideband errors (see discussion associated with Equation (B.3.5)).

The multiband group delays will not be affected by this error, unless there are data losses in the observation, because the error does not depend on the RF frequency (see Equation (B.4.16)). A 10% data loss in one channel may introduce differences in the phase errors between channels of up to 2°, (assuming the data loss occurs only immediately after every bit shifts, which is unlikely), and hence may introduce errors in the multiband group delays. (This topic is explored in further detail in Chapter 3.)

The video rotation term will introduce a bias into the correlation amplitude. If the power spectral density of the recorded signals is constant over the 2 MHz bandwidth, the amplitude of the cross spectral components at frequencies different from that of the center of the video band will be reduced because of the incorrect phase rotation. The amplitude reduction of a cross spectrum component with frequency ω_v will be (if several bit shifts occur during the

observation so that we can ignore the contributions from Δt_b^1 and Δt_b^f (see Figure B.4.13 and the discussion associated with Figure B.4.4),

$$\left| \int_{-\theta_s}^{\theta_s} e^{-i\theta} d\theta \right| / 2\theta_s = 2 \sin(\theta_s)/[(\omega_v - \pi B)\Delta t_s] \qquad (B.4.17)$$

where $\theta_s = (\omega_v - \pi B)\Delta t_s/2$. For the 1.75 MHz and the 0.25 MHz video frequency components the reduction in amplitude will be 5.7%. The reduction in amplitude when the seven video frequencies are averaged will be 2.5%. This bias may be corrected by multiplying the observed correlation amplitude by ≈ 1.026. This type of correction is made in FRNGE but the factor used is $1/[1-\pi^2/288] \approx 1.035$. The correction applied in FRNGE is the amplitude correction which would be applicable if very closely spaced (<<0.25 MHz) frequencies in the video cross spectrum were evaluated (Whitney, 1974, pp. 164-166, i.e., Whitney integrated the bias over the video bandwidth rather than averaging the biases for each of the frequencies used).

After 10 August,1982, the error in total phase due to the neglect of the video rotation was eliminated by applying a (video) frequency dependent rotation correction, (Equation (B.4.16)), to the video cross spectral components computed in FRNGE.

We will now discuss the effects of the last approximation listed at the beginning of this discussion: possible biases caused by an error in the a priori phase delay rate. To analyze this problem it is convenient to visualize the accumu-

- 380 -

lation of the cross correlation function in terms of coher-
ently averaging the cross spectral components (this is the
method which we originally discussed in Appendix A). The
coherent average was formed by summing together the cross
spectral components from 4 μsec data segments. Each cross
spectral estimate was rotated by the expected change in the
phase of its components based on an a priori estimate of the
phase delay rate. Almost always the a priori phase delay rate
is known to within 10^{-11} sec/sec and hence the error in the
rotation of the cross spectral components is bounded by 58°
over a 2 second accumulaton period at an observing frequency
of 8 GHz. This error is not so severe as it first appears
because the estimates of the cross spectral components are
referred to the epoch at the center of the accumulation
period. The phase error due to an error in the a priori phase
delay rate will be zero for this epoch if all of the data
recorded during the accumulation period are used in computing
the estimates of the cross spectral components. All of the
data recorded during an accumulation period may not be used to
estimate the cross spectrum, for example because of tape
reading errors during correlation. The error in the phase of
the coherently averaged cross spectrum due to data losses will
depend upon the times during the accumulation period at which
data are lost. The worst possible error is 58° when only a
few bits of data from either the beginning or the end of the
accumulation period are used to compute the cross spectrum.
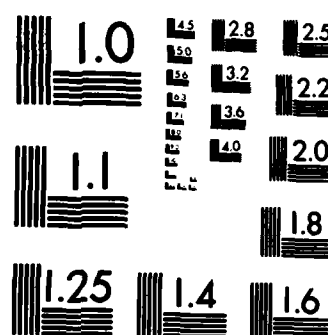This error could be positive or negative depending upon

END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

whether or not the segment is at the beginning or the end of the accumulation period. In general, the error in the phase of the cross spectral components could be positive or negative depending upon whether or not data are lost primarily before or after the temporal center of the accumulation period. A systematic bias could develop in the phase errors if data were lost systematically from before or after these center epochs. If this systematic loss of data occurred, the error in the phase delay would still be bounded by the product of the phase delay rate error and the duration of the accumulation period (for an error of 10^{-11} sec/sec in the phase delay rate and an accumulation period of 2 sec, the maximum error in the phase delay would be 10 psec ≈0.3 cm). If the same systematic data loss occurs in all frequency channels, then the maximum group delay error would also be 10 psec.

To evaluate the effects of systematic data loss in only one channel we can use the equivalence of the maximum likelihood and weighted least squares estimators (see Appendix A). There will be a compensating effect when only one channel loses data; the greater the data loss, the greater the phase error in that channel but the less weight this channel will have in the estimation of the group delay. In Appendix A, we saw that the weight given to the data from an accumulation period is directly proportional to the number of data segments used in evaluating the cross spectral components for that accumulation period. In Figure B.4.15, we show the errors in the estimated group delay which would result from systematic

Figure B.4.15 Errors in multiband group delay due to an error in the _a priori_ phase delay rate of 10^{-11} sec/sec
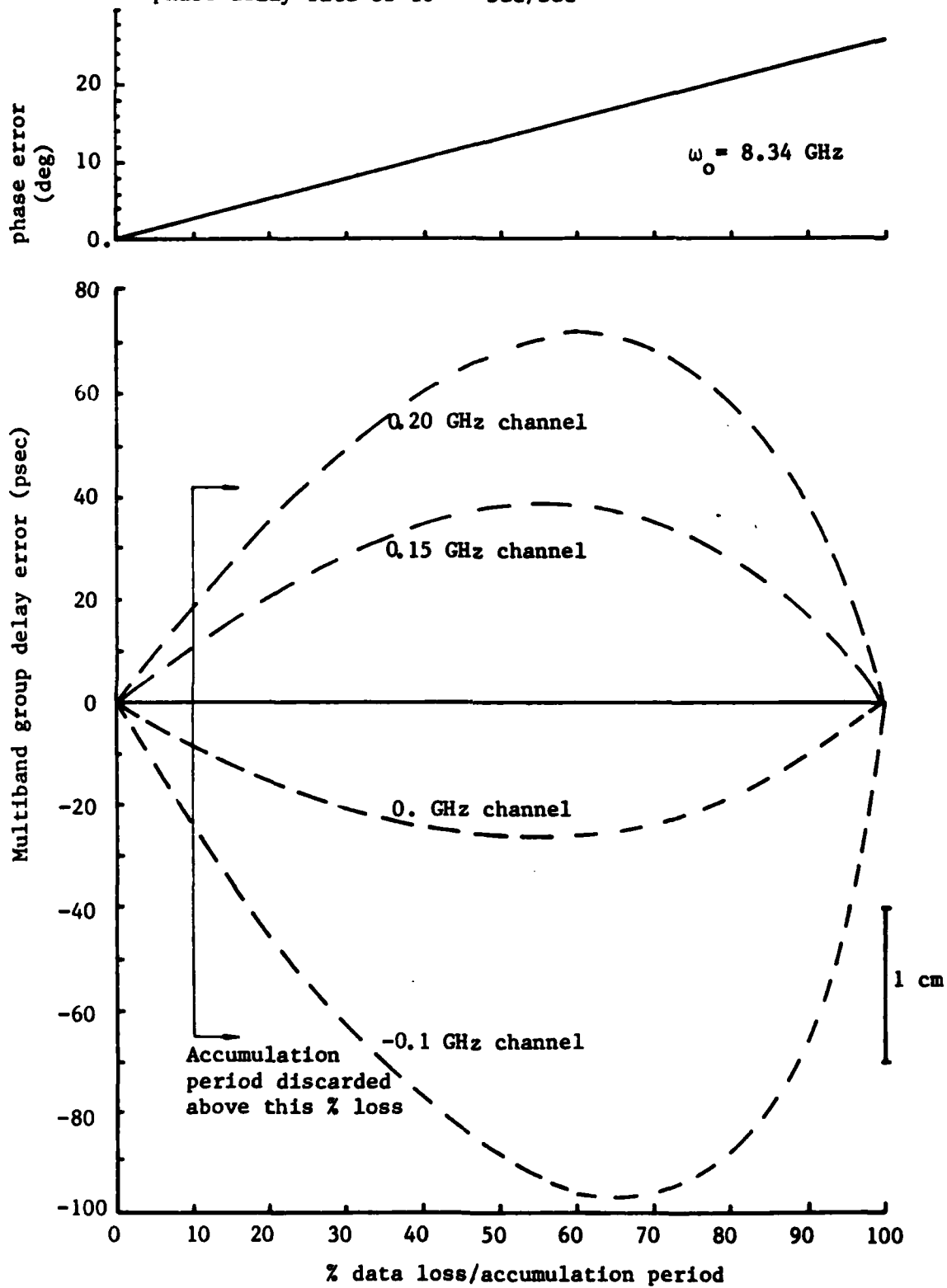
data losses in all accumulation periods for a given channel when the a priori phase delay rate is in error by 10^{-11} sec/sec. These results are extreme for several reasons. We have assumed the largest possible error in the cross spectral phase for a given percentage loss of data, i.e. the data are lost always from the beginning of the accumulation period, every accumulation period for a given channel was assumed to lose exactly the same amount of data at the same location, and the error in the a priori phase delay rate was extreme. More typical errors in the phase delay rate are 10^{-12} sec/sec. The curves will scale linearly with phase delay rate error. In practice if an accumulation period has discard rates of greater than 10%, it is discarded totally and will produce no error in the estimates of the group delays. The use of this cutoff in the amount of data loss in an accumulation limits the group delay error to 20 psec (≈ 0.7 cm) even in the extreme case studied.

The errors in the phase delays which resulted from these same data loses were all less than 2% of the group delay errors, i.e. <0.4 psec.

The use of upper and lower sideband data will not cancel the errors given because for such errors to develop requires loss of data. These data losses will, in general, be different for upper and lower sidebands and hence the error will not cancel.

We have now discussed some of the limitations of the Mark III processing software. There are other effects which could

be discussed which involve the equipment limitations, e.g. we have already seen how the bandpass filter in the video converters can affect the magnitude of an error. In Chapter 3 we discuss some of the equipment limitations along with the investigation of the quality of VLBI measurements.

APPENDIX C.   Correlations Between VLBI Observations.

When three or more radio telescopes simultaneously observe the same radio source, redundant information is obtained which can be used to check the quality of the observations (i.e. for a baseline triplet, the sum of the delays referred to a common epoch around the triplet should be zero if there is no contribution to the sum from the brightness distribution of the radio source). In this appendix, we will evaluate the magnitude of the correlation between the VLBI measurements from each baseline when two baselines have a common station and simultaneously observe the same radio source. The estimated correlations can be used for two purposes. Firstly, if the observations are correlated then we should account for this correlation when estimating geodetic parameters. Secondly, the correlation should be accounted for when we are computing the estimated uncertainty of the sum of the delays around a triplet of baselines.

We will first investigate the relationship between the cross spectra of the signals recorded at each site when three stations simultaneously observe the same radio source. The relationship between the cross spectra will allow us to deduce the correlation between the phases of the components of each cross spectrum. Then, using the relationship between the maximization of the delay resolution function and a least squares approach, established in

Appendix A, we will compute the correlation between the observed values of the group delays, the phase delay rates and the phase delays on two baselines which have a common station.

Three radio telescopes are assumed to be observing the same point radio source. At each station random signals from the source and random noise from the receiver and surroundings are being recorded. In the frequency domain, the spectrum of the signals recorded at each site will be

$$\bar{X}_1(\omega) = \sqrt{Ta_1}\ \bar{S}_1(\omega) + \sqrt{Ts_1}\ \bar{N}_1(\omega) \qquad\qquad (C.1.a)$$

$$\bar{X}_2(\omega) = \sqrt{Ta_2}\ \bar{S}_2(\omega) + \sqrt{Ts_2}\ \bar{N}_2(\omega) \qquad\qquad (C.1.b)$$

$$\bar{X}_3(\omega) = \sqrt{Ta_3}\ \bar{S}_3(\omega) + \sqrt{Ts_3}\ \bar{N}_3(\omega) \qquad\qquad (C.1.c)$$

where $\sqrt{Ta_1}\ \bar{S}_1(\omega), \sqrt{Ta_2}\ \bar{S}_2(\omega)$ and $\sqrt{Ta_3}\ \bar{S}_3(\omega)$ are the spectra of the signals from the radio source at the output of the receiver; $\sqrt{Ts_1}\ \bar{N}_1(\omega)$, $\sqrt{Ts_2}\ \bar{N}_2(\omega)$ and $\sqrt{Ts_3}\ \bar{N}_3(\omega)$ are the spectra of the random additive noise; and ω is the angular frequency (see Appendix A for details).

In Appendix A, we found that the spectra of the signals from the radio source are related by

$$\bar{S}_2(\omega) = \bar{S}_1(\omega)e^{-i\phi_{12}(\omega)} \qquad\qquad (C.2.a)$$

$$\bar{S}_3(\omega) = \bar{S}_1(\omega)e^{-i\phi_{13}(\omega)} \qquad\qquad (C.2.b)$$

where $\phi_{12}(\omega)$ and $\phi_{13}(\omega)$ are the pairwise phase differences between the spectra of the signals recorded at each of the two pairs of sites. These phase differences account for

the difference in arrival times of the signals at each of the sites. We can obtain estimates of these phase differences from the phases of the cross spectra of the data recorded at each site. These estimates will differ from the "true" phase differences due to the presence of the random noise $\sqrt{Ts_1}\ \bar{N}_1(\omega), \sqrt{Ts_2}\ \bar{N}_2(\omega)$ and $\sqrt{Ts_3}\ \bar{N}_3(\omega)$.

The cross spectra of the data from the interferometers formed by Sites 1 and 2 and Sites 1 and 3, $\bar{C}_{12}(\omega)$ and $\bar{C}_{13}(\omega)$, are

$$\bar{C}_{12}(\omega) = \bar{X}_1^*(\omega)\bar{X}_2(\omega)$$

and

$$\bar{C}_{13}(\omega) = \bar{X}_1^*(\omega)\bar{X}_3(\omega)$$

The expansion of these cross spectra in terms of the signal and noise spectra is

$$\bar{C}_{12}(\omega) = \sqrt{Ta_1 Ta_2}\ \bar{S}_1^*\bar{S}_1 e^{-i\phi}12 + \sqrt{Ta_1 Ts_2}\ \bar{S}_1^*\bar{N}_2$$
$$+ \sqrt{Ts_1 Ta_2}\ \bar{N}_1^*\bar{S}_1 e^{-i\phi}12 + \sqrt{Ts_1 Ts_2}\ \bar{N}_1^*\bar{N}_2 \qquad (C.3.a)$$

$$\bar{C}_{13}(\omega) = \sqrt{Ta_1 Ta_3}\ \bar{S}_1^*\bar{S}_1 e^{-i\phi}13 + \sqrt{Ta_1 Ts_3}\ \bar{S}_1^*\bar{N}_3$$
$$+ \sqrt{Ts_1 Ta_3}\ \bar{N}_1^*\bar{S}_1 e^{-i\phi}13 + \sqrt{Ts_1 Ts_3}\ \bar{N}_1^*\bar{N}_3 \qquad (C.3.b)$$

where we have directly substituted Equations (C.2) into the expressions for cross spectra. We have not shown the frequency dependence on the right hand side of the equations.

The first term in each of Equations (C.3) is the signal term. The remaining terms are noise terms. There

will be a correlation between the cross spectra if they
have common noise terms. There is one common noise term
which arises from the product of the signal spectra and the
noise spectra at the common site (in this case Site 1).
The noise spectra at Site 1, $\bar{N}_1$, also appears in another
term in each cross spectrum, $\sqrt{\overline{Ts_1 Ts_2}}\ \bar{N}_1^* \bar{N}_2$ and $\sqrt{\overline{Ts_1 Ts_3}}\ \bar{N}_1^* \bar{N}_3$,
but these terms will not introduce any correlation because
they are independent. This independence is easily shown by
using the expansion of the products of four zero mean
Gaussian variables (Davenport and Root, 1958):

$$\langle \bar{N}_1^* \bar{N}_2 (\bar{N}_1^* \bar{N}_3)^* \rangle = \langle \bar{N}_1^* \bar{N}_2 \rangle \langle \bar{N}_1 \bar{N}_3^* \rangle + \langle \bar{N}_1^* \bar{N}_1 \rangle \langle \bar{N}_2 \bar{N}_3^* \rangle + \langle \bar{N}_1 \bar{N}_3^* \rangle \langle \bar{N}_1^* \bar{N}_2 \rangle,$$

where <> denotes expectation. The expectation of each of
the products in the expansion is zero (at least one pair in
each term is composed of independent factors), and thus
there will be no correlation between the cross spectra due
to these terms. By the same arguments, the terms
containing the signal and the noise at the remote sites
will not introduce any correlation between the cross
spectral estimates.

Of the three noise terms in the cross spectra of the
signals recorded at each site, only one is common to cross
spectra from the two baselines. The common terms from the
two baselines are $\sqrt{\overline{Ts_1 Ta_2}}\ \bar{N}_1^* \bar{S}_1 e^{-i\phi_{12}}$ and $\sqrt{\overline{Ts_1 Ta_3}}$
$\bar{N}_1^* \bar{S}_1 e^{-i\phi_{13}}$. These two terms are deterministically related.
They differ in amplitude by $\sqrt{Ta_2 / Ta_3}$ and in argument by
$e^{-i(\phi_{12} - \phi_{13})}$. An interesting point and one which will be
very useful later, is that the angle between this common

term and the signal part of the cross spectrum is the same for both cross spectra. This angle is equal to the argument of $\bar{N}_1^* \bar{S}_1$.

At this stage we will group the two independent terms in each cross spectrum in Equations (C.3) together which yields

$$\bar{C}_{12}(\omega) = \sqrt{Ta_1 Ta_2}\, \bar{S}_1^* \bar{S}_1 e^{-i\phi}12 + \sqrt{Ts_1 Ta_2}\, \bar{N}_1^* \bar{S}_1 e^{-i\phi}12 + \bar{I}_{12}$$

$$(C.4.a)$$

and

$$\bar{C}_{13}(\omega) = \sqrt{Ta_1 Ta_3}\, \bar{S}_1^* \bar{S}_1 e^{-i\phi}13 + \sqrt{Ts_1 Ta_3}\, \bar{N}_1^* \bar{S}_1 e^{-i\phi}13 + \bar{I}_{13}$$

$$(C.4.b)$$

where

$$\bar{I}_{12} = \sqrt{Ta_1 Ts_2}\, \bar{S}_1^* \bar{N}_2 + \sqrt{Ts_1 Ts_2}\, \bar{N}_1^* \bar{N}_2$$

and

$$\bar{I}_{13} = \sqrt{Ta_1 Ts_3}\, \bar{S}_1^* \bar{N}_3 + \sqrt{Ts_1 Ts_3}\, \bar{N}_1^* \bar{N}_3$$

In Appendix A we saw that the individual estimates of the cross spectral components are not used to directly estimate the values of the VLBI observables. Coherent averages of the cross spectral estimates are used. The cross spectral components, $\bar{C}_{12}^a(\omega)$, averaged over an accumulation period, are computed from (see Appendix A, Equation (A.15) for details)

$$\bar{C}_{12}^a(\omega) = \sum_{j=1}^{n(q)} \bar{X}_{1j}^*(\omega) \bar{X}_{2j}(\omega) e^{i(\alpha_{ap})j}$$

where $(\alpha_{ap})_j$ is an estimate of the change in phase between the signal part of the cross spectral components at epoch j

Table C.1   Statistics of the sum of n(q) cross spectral component estimates

For baseline from Site k to Site m

Signal

$$\langle \bar{S}_k^{'*} \bar{S}_m^{'} \rangle = \sqrt{Ta_k Ta_m}\ n(q)$$

$$\text{Variance}\ (\bar{S}_k^{'*} \bar{S}_m^{'}) = Ta_k Ta_m\ n^2(q)/2$$

Independent noise

$$\langle \text{Re}\{\bar{I}_{km}^{'}\}\rangle = \langle \text{Im}\{\bar{I}_{km}^{'}\}\rangle = 0$$

$$\text{Variance}\ (\text{Re}\{\bar{I}_{km}^{'}\}) = \text{Variance}\ (\text{Im}\{\bar{I}_{km}^{'}\})$$

$$= (Ta_k Ts_m + Ts_k Ts_m)\ n(q)$$

Correlated noise

$$\langle \text{Re}\{\bar{N}_k^{'*} \bar{S}_m^{'} e^{-i\phi_{km}^{'}}\}\rangle = \langle \text{Im}\{\bar{N}_k^{'*} \bar{S}_m^{'} e^{-i\phi_{km}^{'}}\}\rangle = 0$$

$$\text{Variance}\ (\text{Re}\{\bar{N}_k^{'*} \bar{S}_m^{'} e^{-i\phi_{km}}\}) = \text{Variance}\ (\text{Im}\{\bar{N}_k^{'*} \bar{S}_m^{'} e^{-i\phi_{km}}\})$$

$$= Ta_m Ts_k\ n(q)$$

where the summed cross spectral components are defined by

$$\bar{S}_k^{'*} \bar{S}_m^{'} \equiv \sum_{j=1}^{n(q)} \sqrt{Ta_k Ta_m}\ (\bar{S}_k^* \bar{S}_m)_j\, e^{i(\alpha_{ap})_j}$$

$$\bar{I}_{km}^{'} \equiv \sum_{j=1}^{n(q)} (\bar{I}_{km})_j\, e^{i(\alpha_{ap})_j}$$

$$\bar{N}_k^{'*} \bar{S}_m^{'} \equiv \sum_{j=1}^{n(q)} \sqrt{Ts_k Ta_m}\ (\bar{N}_k^* \bar{S}_m)_j\, e^{i(\alpha_{ap})_j}$$

Figure C.1 Graphical representation of the components of the averaged cross spectra $\bar{C}_{12}^a(\omega)$ and $\bar{C}_{13}^a(\omega)$
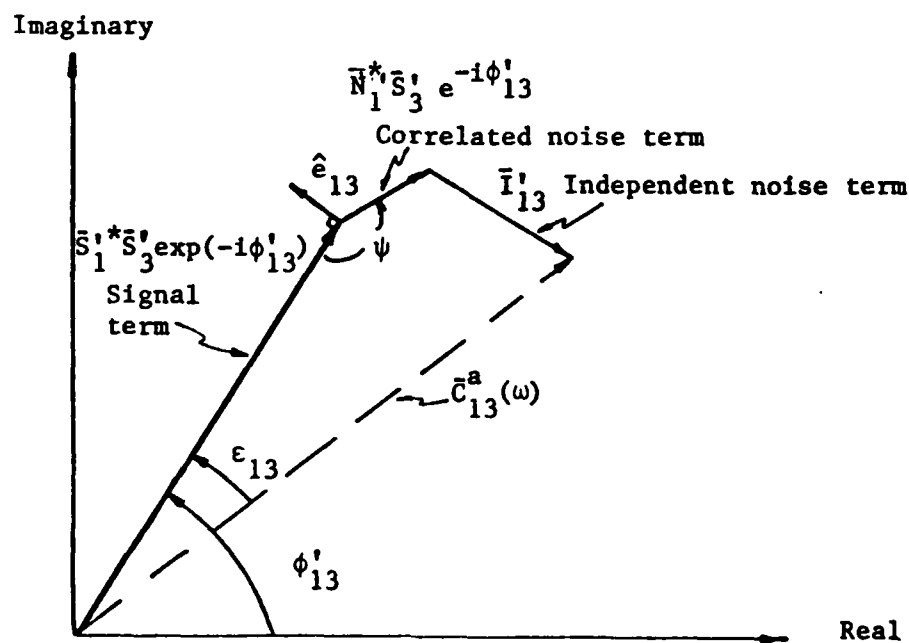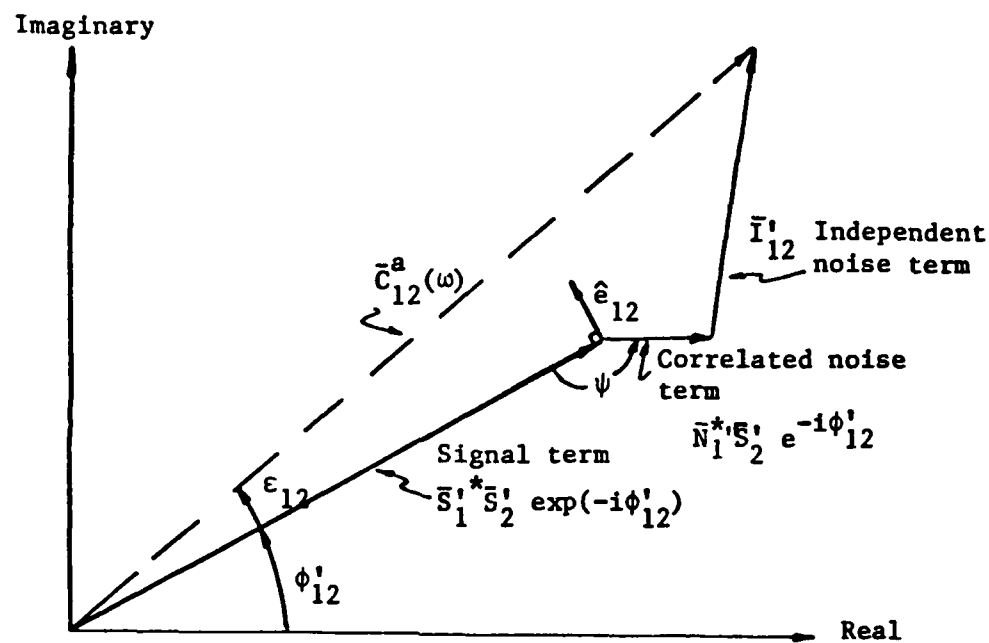
tral component. The phase errors for each baseline are
denoted in Figure (C.2) by ε_{12} and ε_{13}. By treating
complex numbers as vectors in the complex plane and assum-
ing that the noise terms are small compared to the signal
term, we can write these phase errors as

$$\varepsilon_{12} \simeq [\bar{I}'_{12} \cdot \hat{e}_{12} + (\bar{N}'_1 {}^* \bar{S}'_2 e^{-i\phi'}{}_{12}) \cdot \hat{e}_{12}]/|\bar{S}'_1 {}^* \bar{S}'_2| \qquad \text{(C.5.a)}$$

and

$$\varepsilon_{13} \simeq [\bar{I}'_{13} \cdot \hat{e}_{13} + (\bar{N}'_1 {}^* \bar{S}'_3 e^{-i\phi'}{}_{13}) \cdot \hat{e}_{13}]/|\bar{S}'_1 {}^* \bar{S}'_3| \qquad \text{(C.5.b)}$$

where $\hat{e}_{12}$ and $\hat{e}_{13}$ are unit vectors perpendicular to the
signal vector (see Figure (C.2)); ϕ'_{12} and ϕ'_{13} are the cross
spectral phases for the central accumulation period. These
phases account for the errors on the a priori models for
the total phase and the phase-delay rate, i.e. the phase of
the averaged cross spectral components will not be zero, in
general, even when there is no noise present because of
errors in the a priori models which are used to compute
$(\alpha_{ap})_j$.

The variances of ε_{12} and ε_{13} may now be calculated
from Equation (C.5) and the statistics given in Table C.1.
The variances will be given by

$$\langle \varepsilon^2_{12} \rangle = (Ts_1 Ts_2 + Ta_1 Ts_2 + Ta_2 Ts_1)/(Ta_1 Ta_2 n(q))$$

and

$$\langle \varepsilon^2_{13} \rangle = (Ts_1 Ts_3 + Ta_1 Ts_3 + Ta_3 Ts_1)/(Ta_1 Ta_3 n(q))$$

These expressions for the variances may be simplified
by realizing that for most geodetic experiments, the

antenna temperatures are much less than the system temperatures. If we assume that the product of the antenna and system temperatures is much less than the product of the system temperatures, the expressions for the variances reduce to

$$\langle \epsilon_{12}^2 \rangle \simeq Ts_1 Ts_2 / (Ta_1 Ta_2 n(q)) \qquad \text{(C.6.a)}$$

and $\langle \epsilon_{13}^2 \rangle \simeq Ts_1 Ts_3 / (Ta_1 Ta_3 n(q)) \qquad \text{(C.6.b)}$

The final expression we will need before computing the correlation between the errors in the estimates of the phase of the cross-spectral component is the covariance between the phase errors, $\langle \epsilon_{12} \epsilon_{13} \rangle$. This covariance will be given by

$$\langle \epsilon_{12} \epsilon_{13} \rangle =$$

$$\frac{\langle \bar{I}_{12}' \cdot \hat{e}_{12} \ \bar{I}_{13}' \cdot \hat{e}_{13} \rangle + \langle (\bar{N}_1'^* \bar{S}_2' e^{-i\phi}{}_{12} \cdot \hat{e}_{12})(\bar{N}_1'^* \bar{S}_3' e^{-i\phi}{}_{13} \cdot \hat{e}_{13}) \rangle}{|S_1'^* S_2'| |S_1'^* S_3'|}$$

The expectation of the product of the independent noise terms is zero. (This result was shown earlier in this Appendix.) The expectation of the second term can be evaluated by noting that the angles between $\bar{N}_1'^* \bar{S}_2' e^{-i\phi} 12$ and $\hat{e}_{12}$ and $\bar{N}_1'^* \bar{S}_3' e^{-i\phi} 13$ and $\hat{e}_{13}$ are identically equal. We will denote this angular difference with the symbol ψ (see Figure (C.2)). This equality allows the expectation of the product of the correlated noise terms to be reduced to

$$\langle (\bar{N}_1'^* \bar{S}_2' e^{-i\phi} 12 \cdot \hat{e}_{12})(\bar{N}_1'^* \bar{S}_3' e^{-i\phi} 13 \cdot \hat{e}_{13}) \rangle =$$

$$\langle \text{Re}\{ |\bar{N}_1'^* \bar{S}_2'| e^{i\psi} \} \ \text{Re}\{ |\bar{N}_1'^* \bar{S}_3'| e^{i\psi} \} \rangle$$

which may be evaluated using the statistics given in Table
C.1. The covariance between the phase errors will be

$$\langle \varepsilon_{12} \varepsilon_{13} \rangle = Ts_1 / \left[Ta_1 n(q) \right] \qquad (C.7)$$

The combination of Equations (C.6) and (C.7) allows the
correlation between the errors in the estimates of the
phases of the cross-spectral components to be evaluated.
This correlation will be

$$\rho_{\varepsilon} = \langle \varepsilon_{12} \varepsilon_{13} \rangle / \sqrt{\langle \varepsilon_{12}^2 \rangle \langle \varepsilon_{13}^2 \rangle} = \sqrt{Ta_2 Ta_3 / Ts_2 Ts_3} \qquad (C.8)$$

Equation (C.8) gives the correlation between the phase
errors of the summed cross spectral components for a single
accumulation period. The equivalence of maximizing the
delay resolution function and least squares can now be used
to determine the correlation between the errors in the
multiband group delay, phase-delay rate and phase delay
estimates for the pair of baselines.

We will estimate these correlations using a general
approach. If we treat the estimation of the values of the
VLBI observables as a least squares problem, we can express
the group delay, phase-delay rate and phase delay as a
linear combination of the phases of the cross-spectral
components from each accumulation period. We may write
these relationships for each baseline as

$$\beta_{12} = \sum_j \alpha_j (\phi_{12})_j \qquad (C.9.a)$$

and

$$\hat{p}_{13} = \sum_j \alpha_j (\phi_{13})_j \qquad\qquad\qquad (C.9.b)$$

where $\hat{p}_{12}$ and $\hat{p}_{13}$ are estimated values of the VLBI observables (i.e. multiband group delay or phase delay or phase delay rate), α_j are the linear coefficients which relate the coherently averaged cross-spectral phases to the parameter of interest, and $(\phi_{12})_j$ and $(\phi_{13})_j$ are cross-spectral phases from each accumulation period on each of the baselines. The summation will be over all accumulation periods.

The form of Equations (C.9) implies that exactly the same linear relationship is used to determine the values of the observables from the cross-spectral phases from each baseline. This is not exactly correct because some accumulation periods may not be used to estimate $\hat{p}_{12}$ and $\hat{p}_{13}$ on either one or both of the baselines due to tape reading errors. We will investigate the effects of this possibility shortly.

Equations (C.9) may be used to determine the correlation between the errors in the estimates of $\hat{p}_{12}$ and $\hat{p}_{13}$. This correlation will be given by

$$\rho_p = \langle \epsilon_{p_{12}} \epsilon_{p_{13}} \rangle / \sqrt{\langle \epsilon^2_{p_{12}} \rangle \langle \epsilon^2_{p_{13}} \rangle}$$

which reduces to

$$\rho_p = \sum_j \alpha^2_j \langle \epsilon_{12j} \epsilon_{13j} \rangle / \sqrt{(\sum_j \alpha^2_j \langle \epsilon^2_{12j} \rangle)(\sum_j \alpha^2_j \langle \epsilon^2_{13j} \rangle)} \qquad (C.10)$$

where we have assumed that the expectation of the product

terms from different accumulation periods is zero. (This assumption is consistent with the statistical assumptions made in Chapter 2 for the properties of the signal and noise recorded at each site.) If we assume that the expectations of the phase error products are independent of j, Equation (C.10) reduces to

$$\rho_p = \langle \epsilon_{12} \epsilon_{13} \rangle / \sqrt{\langle \epsilon_{12}^2 \rangle \langle \epsilon_{13}^2 \rangle} = \rho_\epsilon$$

Therefore, the correlations between the errors in the estimates of the group delays for a pair of baselines with a common station will be

$$\rho_{\tau_g} = \sqrt{Ta_2 Ta_3 / Ts_2 Ts_3} \qquad (C.11)$$

Similarly, the correlations between the errors in the estimates of the phase-delay rates and the phase delays from each baseline will be given by the same expression.

If the set of accumulation periods used to estimate the values of the observables is different for each of the baselines the correlation given by Equation (C.11) will be reduced. When some accumulation periods are not used the variance of the estimated parameter will increase while the covariance $(\Sigma \alpha_j^2 \langle (\epsilon_{12})_j (\epsilon_{13})_j \rangle)$ will decrease because there are less terms in the summation. Hence, Equation (C.11) places an upper bound on the correlation.

Before completing this appendix we will evaluate the effects of the correlation on the estimated standard deviation of a baseline triplet closure uncertainty. We will

consider a simple case for which the antenna and system temperatures are the same for each station in the triplet and exactly the same accumulation periods are used for each observation. For this case it can be easily shown that the variance of the closure quantity (group delay, phase-delay rate or phase delay) will be over-estimated by six times the covariance of the baseline quantity if the correlation between baseline quantities is not taken into account. (The variance of the sum of the delays around the triplet will be $\sigma_\Sigma^2 = \langle(\epsilon_{12}+\epsilon_{23}+\epsilon_{31})^2\rangle =$
$\langle\epsilon_{12}^2\rangle+\langle\epsilon_{31}^2\rangle+\langle\epsilon_{31}^2\rangle+2\langle\epsilon_{12}\epsilon_{23}\rangle+2\langle\epsilon_{23}\epsilon_{31}\rangle+2\langle\epsilon_{12}\epsilon_{31}\rangle$. For the case given the expectations of the cross terms would be equal and therefore $\sigma_\Sigma^2 = \sigma_{12}^2+\sigma_{13}^2+\sigma_{31}^2+6\langle\epsilon_{12}\epsilon_{23}\rangle$.)

TYpically values of the correlation between the errors in the values of the VLBI observables are less than 0.01, i.e. typically $\sqrt{Ta_2 Ta_3/Ts_2 Ts_3}$ is less than 0.01. The above result would indicate that neglecting correlations while computing closure statistics would introduce errors in these statistics of less than 1%.

The correlations between errors in the group delay, phase delay and phase delay rate estimates can be easily accounted for when closure statistics are computed. At this time, accounting for these correlations in the weighted least squares solutions for geodetic parameters does not seem warranted given the uncertainty in the appropriate statistics for the group delay and phase-delay rate measurements used in these weighted least squares solu-

tions. If antenna and system temperatures improve
considerably (such that correlations of 10% are likely), it
may appear that the correlation between measurements
discussed in this appendix could become important.
However, if this situation were to arise, the baseline
length measurements would be so accurate (standard
deviations of severval microns, probably), that we could
still neglect the correlations. [Remember, that for the
correlations, due to random noise, to be important, all
other noise sources in the analysis must be small compared
to this random noise. With the current receivers and
antennas, the standard deviations of baseline length deter-
minations would be a few millimeters, if these determina-
tions were signal-to-noise ratio limited. (This is with
less than 1% correlations between measurements.) If the
sytem temperatures were to improve tenfold, then baseline
lengths could be measured with submillimeter accuracy
(again if the analysis were random noise limited).
However, with these improvements in system temperatures, we
could use the group-delay measurements to predict the phase
delay ambiguities, thus decreasing the standard deviations
of the baseline length estimates by 40 fold (to the tens of
microns level). If the analysis was still signal-to-noise
ratio limited at this stage, the correlations could be
important. Of course, it is very unlikely we will ever
reach this stage.]

Appendix D. Feed rotation correction

In this appendix we will develop the algorithms for correcting the visibility phases for the effects of feed horn rotation. At first glance it is not obvious why the orientation of the feed horn should affect these phases. Externally, feed horns appear to be very symmetrical i.e., usually conical in shape. However, at the base of all feed horns there is a crossed dipole or an equivalent component. We will first show why the correction is necessary by considering the output of a pair of crossed dipole antennas in the presence of a propagating plane wave. We will then develop the algorithms for an antenna located on a rotating earth.

The electric field $\bar{E}$ of a monochromatic circularly polarized plane wave at position $\bar{x}$ and time t is given by (Jackson, 1975, pp. 273-278)

$$\bar{E}(\bar{x},t) = E_o(\hat{\epsilon}_1 \pm i\hat{\epsilon}_2) \ e^{i(\bar{k}\cdot\bar{x}-\omega t)}$$

where E_o is the amplitude of the wave (volts/m), $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ are orthogonal unit vectors (see Figure D.1), $\bar{k}$ is the wave vector of the wave (directed in the direction of propagation) and ω is the frequency (radians/sec) of the wave. The upper sign is for right circularly polarized (RCP) radiation (IEEE definition, i.e. for an observer looking in the direction of propagation, the $\bar{E}$ field vector at a fixed location appears to rotate clockwise); the lower sign is for left circularly polarization (LCP). (The use of complex numbers to represent

Figure D.1 Geometry of dipole and field unit vectors
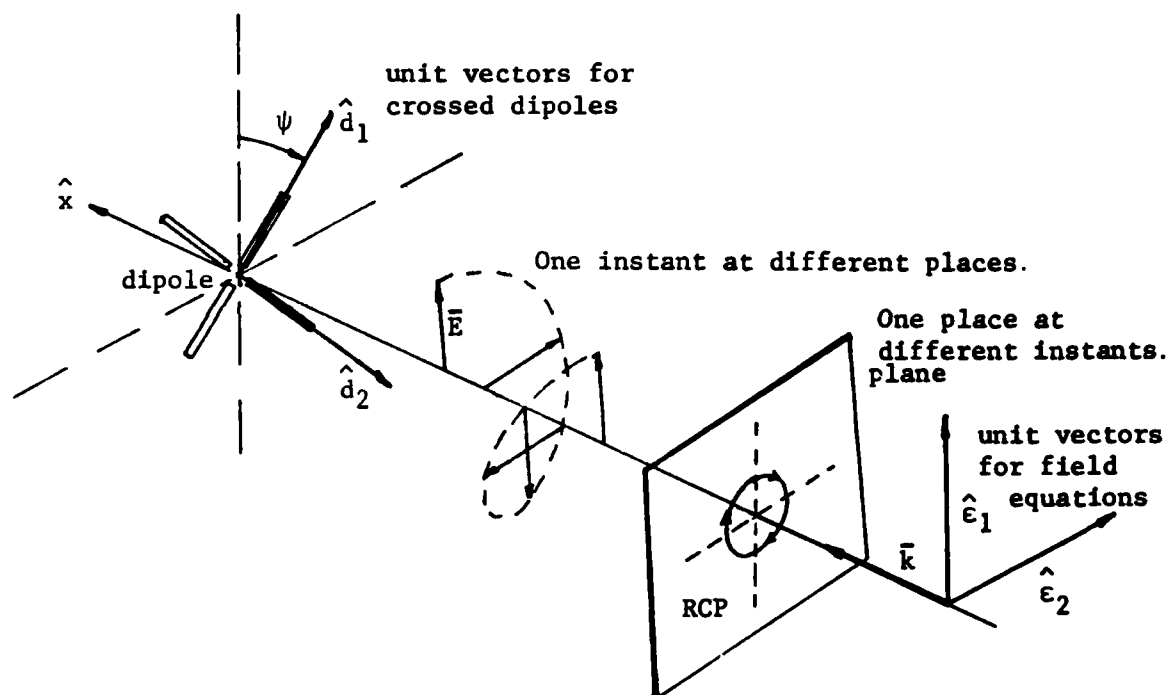
the electric field is a mathematical convenience. The actual field quantities can be obtained from the complex field quantities by taking their real part.)

To detect this wave we need a method for converting the spatially propagating wave into a guided wave propagating along a transmission line. A simple method to detect the propagating wave is to use a dipole antenna. The voltage difference between the terminals of the dipole, when the dipole is oriented in direction $\hat{d}_1$, will be (see Figure D.1)

$$V_1(t) = \mathrm{Re}\{\kappa E_o \hat{d}_1 \cdot (\ell_1 \pm i \ell_2) \; e^{i(\bar{k} \cdot \bar{x}_d - \omega t)}\}$$

where κ is a constant (with units of length) which will be a function of the dimensions of the dipole, the conductivity of the dipole and the shape of the $\bar{E}$ field in the region around the dipole antenna; and $\bar{x}_d$ is the position of the center of the dipole. (The fields in the region close to the dipole can not be plane, in general, because the tangential components of the electric field must be zero on the surface of the (perfectly conducting) dipole. For this discussion we do not need to know the exact functional relationship between these quantities and κ.) The above equation expresses that the component of the electric field parallel to the dipole is measured. (The equation is strictly valid only for dipoles whose length is small compared to the wavelength of the radiation (such that the dipole will not disturb the electric field away from its surface. In practice, dipoles can have lengths comparable to the wavelengths of the radiation

(Jackson, 1975, pp. 273-278).)

In order to detect only one circular polarization the output of a second dipole, $\hat{d}_2$, which is orthogonal to $\hat{d}_1$, is added to the output of $\hat{d}_1$ with a $\mp 90^\circ$ phase shift (Figure D.1). If only RCP is desired, the signal from $\hat{d}_2$ is rotated by -90°, and added to the signal from $\hat{d}_1$. Hence, the summed output will be

$$V(t) = \mathrm{Re}\{\kappa E_o[\hat{d}_1\cdot(\hat{\epsilon}_1\pm i\hat{\epsilon}_2)-i\hat{d}_2\cdot(\hat{\epsilon}_1\pm i\hat{\epsilon}_2)]e^{i(\overline{k}\cdot\overline{x}_d-\omega t)}\} \qquad (D.1)$$

which reduces to

$$V(t) = \mathrm{Re}\{\kappa E_o\hat{d}_1\cdot(\hat{\epsilon}_1+i\hat{\epsilon}_2)e^{i(\overline{k}\cdot\overline{x}_d-\omega t)}\} \qquad (D.2.a)$$

when $\overline{E}(\overline{x},t) = E_o(\hat{\epsilon}_1+i\epsilon_2)e^{i(\overline{k}\cdot\overline{x}-\omega t)}$,i.e. for RCP radiation,

and

$$V(t) = 0 \qquad (D.2.b)$$

when $\overline{E}(\overline{x},t) = E_o(\hat{\epsilon}_1-i\epsilon_2)e^{i(\overline{k}\cdot\overline{x}-\omega t)}$,i.e. for LCP radiation.

We may further simplify Equation (D.2.a) by substituting

$$e^{+i\psi} = \hat{d}_1\cdot(\hat{\epsilon}_1+i\hat{\epsilon}_2)$$

where ψ, the feed position (or parrallatic) angle, by definition equals $\tan^{-1}(\hat{d}_1\cdot\hat{\epsilon}_2/\hat{d}_1\cdot\hat{\epsilon}_1)$. With this substitution, Equation (D.2.a) becomes

$$V(t) = \mathrm{Re}\{\kappa E_o e^{i(\overline{k}\cdot\overline{x}_d+\psi-\omega t)}\} \quad . \qquad (D.3.a)$$

When the signal from $\hat{d}_2$ is added to the signal from $\hat{d}_1$ with a $+90^\circ$ phase shift and LCP radiation is observed, the output of

the crossed dipoles will be

$$V(t) = \text{Re}\{\kappa E_o e^{i(\bar{k}\cdot\bar{x}_d - \psi - \omega t)}\} \quad . \tag{D.3.b}$$

The sense of angle ψ is shown in Figure D.1. Equations (D.3) clearly indicate that the phase of V(t) is not only a function of the separation of the source and the dipole, but also a function of the orientation of the dipole.

We will now investigate a method for correcting the visibility phase for the effects of feed rotation. To be able to calculate the feed position angle ψ we need to express $\hat{d}_1$ and $\hat{\epsilon}_1$ in a common coordinate system. As with most problems the ease of finding a solution is a function of choosing an appropriate coordinate system and set of unit vectors. We will choose the $\hat{\epsilon}_1$ vector such that it lies in the plane containing the rotation axis of the earth and the radio source, and is pointed north (Figure D.2). (By definition, $\hat{\epsilon}_1$ is orthogonal to $\bar{k}$ and hence its orientation is uniquely determined except for sources at the pole.) For sources at the pole we can define $\hat{\epsilon}_1$ to point in any well defined direction, e.g. toward the equinox of date. We discuss this special case after we have formulated the general case.)

The $\hat{d}_1$ vector will be defined such that it lies in the plane containing the primary axis of the telescope and the source. The primary axis of a telescope is the axis whose direction always remains fixed with respect to a local (earth fixed) coordinate system. (For azimuth/elevation (az/el) mounts, this axis is in the direction of the local vertical,

Figure D.2 Definitions of the dipole and field unit vectors
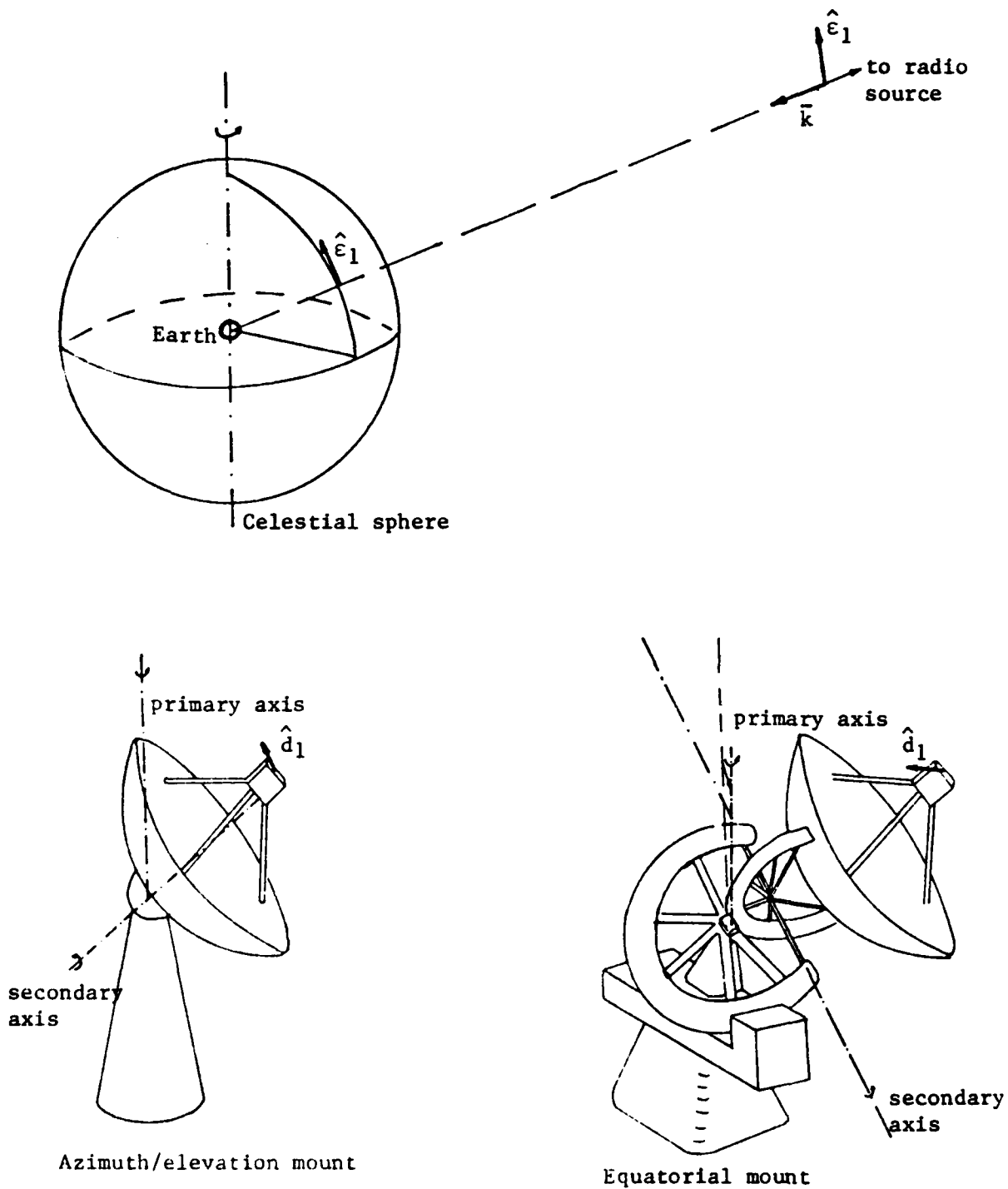
and for equatorial mounts this axis is parallel to the
rotation axis of the earth, if the telescope is located at the
latitude for which it was designed.) The orientation of $\hat{d}_1$ is
uniquely defined, to within a sign, by making it orthogonal to
the direction in which the telescope is pointing (assumed to
be in the direction of the radio source), (Figure D.2). (The
definition of $\hat{d}_1$ degenerates for az/el telescopes when the
radio source is at zenith. In this case, the feed rotation
correction cannot be calculated because the radio telescope
can be pointed in any azimuth and still observe the radio
source. In practice, this situation does not occur because
the antenna pointing programs also fail in this situation. In
all Mark III experiments that we will be analyzing there are
no observations within 2° of zenith at az/el telescopes. (The
Mark III scheduling program, SKED, will not allow observations
to be scheduled at az/el telescopes within this limit.)

We have now defined a convenient set of unit vectors, but
in different coordinate systems. To find the relationship
between the unit vectors we may project them onto the
celestial sphere. In Figure (D.3) we show such a projection
of the unit vectors $\hat{\varepsilon}_1$ and $\hat{d}_1$. The unit vector $\hat{\varepsilon}_1$ lies in the
plane of the rotation axis of the earth and the radio source,
and will project onto the great circle connecting the source
and the pole of the rotation axis of the earth. The unit
vector $\hat{d}_1$ lies in the plane of the principal axis of the
telescope and the source, and will project onto the great
circle between the source and the pole of the principal axis

Figure D.3 Projection of the source and feed horn unit vectors onto the celestial sphere
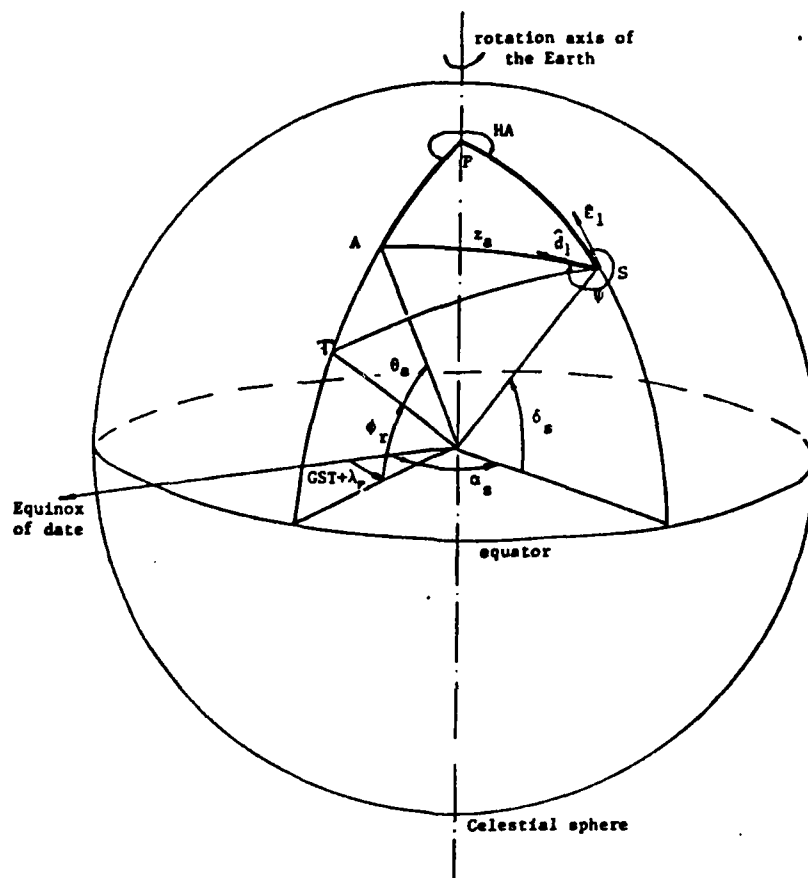
Legend

T   is the extension onto the celestial sphere of the line from the center of the earth to the radio telescope

S   is the position of the radio source on the celestial sphere

P   is the extension of the Earth's north pole onto the celestial sphere

A   is the extension of the radio telescope's principal axis onto the celestial sphere

λ_r, ϕ_r are the longitude and latitude of the radio telescope

α_s, δ_s are the right ascension and declination, of date, of the radio source

GST   is apparent Greenwich sidereal time

HA   is the local hour angle of the radio source (HA=GST+λ_r-α_s)

θ_a is the zenith angle of the primary axis of the radio telescope

z_a is the zenith distance of the radio source

ψ is the feed position angle

of the telescope. Figure (D.3) has been drawn for the case when the principal axis of the telescope lies in the north-south plane at the site and is inclined by angle θ_a with respect to zenith. If the telescope is an azimith/elevation mount, then $\theta_a = 0$. If the telescope is an equatorial mount, then $\theta_a = \pi/2 - \phi_r$, where ϕ_r is the latitude of the radio telescope. The case drawn corresponds to an equatorial mount that has been moved from the latitude for which it was designed. This situation will apply to the Richmond, Florida, antenna which will form part of the POLARIS (POLar motion Analysis by Radio Interferometry Surveying) network (Carter, 1979). (The primary axis of the Richmond antenna will be placed in the local meridian (Robertson, 1982; private communication), and we have drawn Figure D.3 assuming that this will be the case.)

The feed angle ψ can be easily computed from the spherical triangle PAS (see Figure D.3). The angle between the projection of the principal axis of the telescope and the source may be computed using the law of cosines for spherical triangles. Application of this law yields

$$\cos(z_a) = \sin(\phi_r + \theta_a)\sin\delta_s + \cos(\phi_r + \theta_a)\cos\delta_s\cos HA \qquad (D.4)$$

where the angles are defined in Figure (D.3). We could determine ψ using the law of sines but care would be needed in resolving the correct quadrant of the inverse sine. (This approach was adopted by Knight, 1981, Appendix 3, although the logic used to derive the correction was somewhat different from that presented in this appendix.) The quadrant problem

can be resolved by using the half angle formula (see, for example, Breyer, 1978; pp. 175-178). Thus,

$$\tan(\psi/2) = \text{sign(HA)} \left[\frac{\sin(s-\pi/2+\delta_s)\sin(s-z_a)}{\sin(s-\pi/2+\phi_r+\theta_a)\sin(s)} \right]^{1/2} \quad (D.5)$$

where sign(HA) is the sign of the hour angle expressed as an angle between $-\pi$ and π (the positive value of the square root should be taken), and $\underline{s}$ is half the sum of the sides of the spherical triangle PAS, i.e., $s = ((\pi/2-(\phi_r+\theta_a)) + (\pi/2-\delta_s) + z_a)/2$. Application of Equation (D.5) leads to a unique determination of ψ.

When $\theta_a=(\pi/2)-\phi_s$, the principal axis of the telescope is parallel to the rotation axis of the earth, i.e. A and P in Figure D.3 will coincide. In this case z_a will equal $(\pi/2)-\delta_s$ and $\underline{s}$ will reduce to $(\pi/2)-\delta_s$. Substitution of these values into Equation (D.5) shows ψ will be zero for all observations.

We may now write down the algorithm for the correction for feed rotation. The visibilty phase is the phase of the signal at Site 2 minus the phase at Site 1 at the same epoch (see Chapter 2 for discussion). From Equation (D.3.a) we deduce this difference in phase, ϕ_t, for RCP (the polarization used in the Mark III experiments that we will be studying):

$$\phi_t = (\bar{k}\cdot\bar{x}_2+\psi_2) - (\bar{k}\cdot\bar{x}_1+\psi_1)$$

where $\bar{x}_2$ and $\bar{x}_1$ are the coordinates of the dipole at Sites 2 and 1, respectively, and ψ_2 and ψ_1 are the corresponding feed angles at each site. The visibility phase corrected for feed rotation (ϕ_t^C) will be

$$\phi_t^C = \bar{k}\cdot\bar{x}_2 - \bar{k}\cdot\bar{x}_1$$

which can be generated from the observed visibility phase by subtracting the appropriate combinations of feed angles. For RCP, we have

$$\phi_t^C = \phi_t - (\psi_2 - \psi_1) \quad . \tag{D.6.a}$$

For LCP, the correction is

$$\phi_t^C = \phi_t + (\psi_2 - \psi_1) \quad . \tag{D.6.b}$$

We should note that ψ is independent of frequency and hence no correction need be applied to the group delays.

In Figure D.4 we plot the feed rotation correction, for an interferometer with one element being a telescope at 45° latitude, with an az/el mount, and the other, a telescope with an equatorial mount, at an arbitrary location. The correction is plotted as a function of the hour angle of the source at the az/el telescope. *The maximum observing time at the az/el telescope has been plotted for the source at different declinations, i.e. the limits on mutual visibility of the source from the two sites were ignored.*

In Figure D.4 we see one discontinuity in the correction (for $\delta_s = 45^\circ$, and zero hour angle). This discontinuity occurs as the source passes through zenith at the az/el telescope. When the source is at zenith, the radio telescope can be pointed in any azimuth and still observe the radio source. The feed rotation correction can have any value between $\pm 180^\circ$ depending on the azimuth of the telescope. In practice, this situation does not occur (see earlier discussion on the definition of $\hat{d}_1$).

We mentioned earlier in this appendix the degeneracy of

Figure D.4   Feed rotation correction to the visibility phase for an interferometer formed by two telescopes, one at 45° latitude with an az/el mount and the other with an equatorial mount at an arbitary location.

our definition of the direction of $\hat{e}_1$ when $\delta_s=\pi/2$. We may now study this case in more detail. If we substitute $\delta_s=\pi/2$ into Equation (D.4) we see that $z_a = \pi/2-(\phi_r+\theta_a)$ and hence s = $\pi/2-(\phi_r+\theta_a)$. If we now attempt to use Equation (D.5), we see that the equation is inderterminate, i.e. both $s-z_a$ and $s-\pi/2+\phi_r+\theta_a$ equal zero when $\delta_s=\pi/2$. We could determine the value of Equation (D.5) by finding the limit of $\sin(s-z_a)/\sin(s-\pi/2+\phi_r+\theta_a)$ as $\delta_s\to\pi/2$. However, a simpler approach is to use the law of sines to determine ψ. Thus,

$$\sin\psi = \sin(HA)\cos(\phi_r+\theta_a)/\sin(z_a).$$

When $\delta_s=\pi/2$, the law of sines yields $\sin\psi = \sin(HA)$ and hence $\psi = HA\pm n\pi$, where any integral value of $\underline{n}$ (including zero) will satisfy $\sin\psi = \sin(HA)$. The solution consistent with Equation (D.5) is $\psi = HA\pm\pi$. When the law of sines is used to solve the problem there does not appear to be any singularity, but $\hat{e}_1$ is undefined for sources at the pole. What happened? The answer is quite simple, for sources at the pole, the hour angle is undefined because the source may have any right ascension (see Figure D.3). We may arbitarily define the right ascension of the source and hence implicitly define the direction of $\hat{e}_1$. (If the right ascension is taken to be zero, $\hat{e}_1$ will be directed toward the equinox of date. This may be shown by considering the limit of the geometry in Figure D.3 as $\delta_s\to\pi/2$.) Since the correction to the interferometer visibility phase is the difference between the feed position angles at each site in the interferometer, any constant change in the feed position angle will cancel in the difference, except when

one element in the interferometer is an equatorial mount. For equatorial mounts (at their design latitudes), we saw that the feed position angle was zero (see earlier discussion). This statement is not correct if $\delta_s = \pi/2$, because in this case the radio telescope can be pointed at any right ascension and still observe the source. For reasons similar to those discussed for az/el telescopes pointing at zenith, this situation does not arise in practice and hence the singularities in the definitions of unit vectors $\hat{\epsilon}_1$ and $\hat{d}_1$ should cause no problems in practice.

There are several assumptions in the formulation presented in this appendix which are not strictly valid. We assumed that the $\hat{\epsilon}_1$ unit vector was fixed with respect to inertial space by defining its orientation as fixed with respect to the rotation axis of the earth. However, the rotation axis of the earth precesses and nutates in inertial space. In 13,000 years (half of the precession period) the error introduced by the motion of the rotation axis in inertial space will reach its maximum value of $\approx 46^\circ$. For the next several decades we can safely ignore this error. Similarly, the unit vector $\hat{d}_1$ was referred to the rotation axis of the earth which moves with respect to the crust of the earth. The motion of the rotation axis with respect to the crust of the earth will also introduce some error in the formulation. Both of these motions of the rotation axis will introduce errors of less than 1° in the corrections to the visibilty phases within the next several decades, with the major error contribution coming

from the precession of the rotation axis in inertial space.
If errors of this magnitude do become a problem, the formu-
lation of the feed rotation correction can be updated to
include the effects of precession and nutation.

Appendix E. Review of weighted least squares

In this appendix we review the weighted least squares estimation technique. We will not derive the basic least squares algorithm because this derivation can be found in nearly all books on estimation theory. We will instead concentrate on extensions of the basic formulation which allow large numbers of parameters to be estimated, and correlations to be accounted for (without the direct inversion of large matrices). We will also study some methods of interpreting the results of a least squares analysis.

The basic estimation problem which we wish to solve is, given n observations $\tilde{\underline{y}}$, with expectations $\underline{y}$, and noise contributions, $\underline{\varepsilon}$, such that $\tilde{\underline{y}} = \underline{y} + \underline{\varepsilon}$, (we will discuss the nature of $\underline{\varepsilon}$ shortly), we wish to estimate the values of m parameters, $\underline{x}$. These parameters are related to the observables by a mathematical model $\underline{y} = \underline{f}(\underline{x})$. In general, $\underline{f}(\underline{x})$ will be non-linear functions and we will first linearize the estimation problem by using a Taylor's series expansion of $\underline{f}(\underline{x})$ about a priori values of the parameters, $\underline{x}_o$.

With the first order Taylor's series expansion, the model becomes

$$\underline{y} = \underline{f}(\underline{x}_o) + \left(\partial \underline{f} / \partial \underline{x} \right) \Delta \underline{x} \qquad (E.1)$$

We can compress the above equation by defining the "pre-fit" residuals $\Delta \tilde{\underline{y}} = \tilde{\underline{y}} - f(\underline{x}_o)$ and the partials matrix $\mathbf{A} = \partial \underline{f} / \partial \underline{x}$,

which yields,

$$\Delta\tilde{\underline{y}} = A\Delta\underline{x} + \underline{\varepsilon} \quad . \tag{E.2}$$

For any choice of $\Delta\underline{x}$, $\Delta\tilde{\underline{x}}$, we can calculate the "post-fit" residuals, $\underline{v}$, given by

$$\underline{v} = \Delta\tilde{\underline{y}} - A\Delta\tilde{\underline{x}} \quad . \tag{E.3}$$

If we know the statistical properties of $\underline{\varepsilon}$, i.e. through a variance-covariance matrix, or simply covariance matrix, $V_{\varepsilon\varepsilon} = \langle\underline{\varepsilon}\underline{\varepsilon}^T\rangle$, then the weighted least squares estimator is defined by the estimates of $\Delta\underline{x}$ which minimize $\underline{v}^T V_{\varepsilon\varepsilon}^{-1}\underline{v}$. It may be easily shown that the weighted least squares estimate of the parameter corrections, $\Delta\hat{\underline{x}}$, is given by (see, e.g., Kaula, 1966, Chapter 5),

$$\Delta\hat{\underline{x}} = (A^T V_{\varepsilon\varepsilon}^{-1} A)^{-1} A^T V_{\varepsilon\varepsilon}^{-1}\Delta\tilde{\underline{y}} \quad . \tag{E.4}$$

In the remainder of this appendix, it will be convenient to write Equation (E.4) in two different forms. The first form uses the concept of generalized matrix inverses (see, e.g., Bjerhammar, 1973, Chapter 9; and Bjerhammar, 1975), and Equation (E.4) is written as

$$\Delta\hat{\underline{x}} = A^-\Delta\tilde{\underline{y}} \tag{E.5.a}$$

where A^- is the generalized inverse of A. (The generalized inverse of a matrix is defined by $A = AA^-A$. We should notice that if A is a non-singular square matrix, the generalized inverse is the usual inverse of matrix A). We should note that Equation (E.5.a) is a general form which will admit all linear estimators. The weighted least squares estimator is

just one form of the (non-unique) generalized inverse. We will discuss the interpretation of A^- in Section 3.1.

The second form in which we will write Equation (E.4) is its computational form. We introduce the normal equations, $N = A^T V_{\varepsilon\varepsilon}^{-1} A$, and a vector $\underline{\tilde{z}} = A^T V_{\varepsilon\varepsilon}^{-1} \Delta\underline{\tilde{y}}$. Equation (E.4) then reduces to

$$\Delta\underline{\hat{x}} = N^{-1}\underline{\tilde{z}} \quad . \tag{E.5.b}$$

Equations (E.5.a) and (E.5.b) look similar but the sizes of the matrices in these equations are very different, e.g., if we have 1000 observations, and 100 parameters to be estimated, A^- is a 100x1000 matrix, (requiring 390 kbytes of storage), whereas N^{-1} is a 100x100 (symmetric) matrix requiring only 20 kbytes of storage. Computationally, Equation (E.5.b) is much easier to implement than Equation (E.5.a). The advantages of Equation (E.5.a) will be discussed shortly.

We now return to the investigation of the differences between the pre-fit, $\underline{\tilde{\varepsilon}}$, and post-fit residuals, $\underline{v}$. If we substitute Equation (E.5.a) into Equation (E.3), and use Equation (E.2) to eliminate $\Delta\underline{\tilde{y}}$ from the resulting equation, we obtain

$$\underline{v} = (I-AA^-) \underline{\tilde{\varepsilon}} \quad . \tag{E.6}$$

This equation gives us a relationship between the pre-fit and post-fit residuals. It also allows us to compute, for an error $\underline{\tilde{\varepsilon}}$ in the observations (e.g., due to inadequate models or instrumental errors), the form of the post-fit residuals. When the form of $\underline{\tilde{\varepsilon}}$ is known, Equation (E.6) provides a con-

venient means of detecting $\underline{\varepsilon}$ in the post-fit residuals.

It would appear from Equation (E.6) that we could esti-
mate the errors in the observations by

$$\tilde{\underline{\varepsilon}} = (I-AA^-)^{-1} \underline{v} \quad .$$ (E.7)

Unfortunately, the matrix $I-AA^-$ is singular and hence does not
have a unique inverse. The solutions of Equation (E.7) are of
the form

$$\tilde{\underline{\varepsilon}} = \underline{v} + A\delta\underline{x}$$

where $\delta\underline{x}$ are arbitrary corrections to the parameters. It is
interesting to note that if we can bound $\tilde{\underline{\varepsilon}}$, then we could
bound $\delta\underline{x}$, and hence place an upper bound on the effects of
systematic errors.

We can estimate from Equation (E.6) the covariance matrix
of the post-fit residuals, V_{vv},

$$V_{vv} = \langle \underline{v}\underline{v}^T \rangle = (I-AA^-)^T V_{\varepsilon\varepsilon}(I-AA^-)$$ (E.8)

which after some matrix manipulation reduces to

$$V_{vv} = (I-AN^{-1}A^T) V_{\varepsilon\varepsilon}.$$ (E.9)

Ideally the matrix $AN^{-1}A^T$ should have all small elements, so
that the post-fit residuals have statistics which are similar
to those of $\tilde{\underline{\varepsilon}}$, but this will not always be the case. For
large numbers of observations, $I-AN^{-1}A^T$ is a very large matrix
and we currently have no software which will compute this
whole matrix. (For 1000 observations, $I-AN^{-1}A^T$ would require
nearly 2 Mbytes of storage.)

It should also be noted from Equation (E.9) that, in

general, the off diagonal elements in $I - AN^{-1}A^T$ will be non-zero and hence correlations between the post-fit residuals do not necessarily mean that the observation noise was correlated (see, e.g., Section 3.1 for discussion of correlations among residual phases).

The estimation technique we have just discussed uses a parameterization of the observables. There is another formulation of estimation problems which is normally not discussed in estimation texts because it is usually difficult to apply. This technique is estimation by conditions. If $\underline{m}$ observables are functions of $\underline{n}$ parameters, then there must by m-n conditions which can be imposed on the observations, e.g., if height differences between points are measured around closed loops, then the conditions would be that the sum of height differences between points around each closed loop should equal zero. (For this example, we could also use a parameter estimation technique by expressing the height differences as functions of the heights of each point.) For the height difference measurements, the conditions are easily formulated, but in many cases, finding the conditions can be very difficult. We will develop this method because some problems can be solved very easily with this approach.

We express the (linearized) conditions between the observables as

$$U\underline{y} = \underline{U}_o \tag{E.10}$$

where U is a matrix of the condition partial derivatives and $\underline{U}o$ is a vector of constants. (For the height difference case U

would be a $(m-n) \times m$ matrix containing ones, minus ones and zeroes, and the $\underline{U}_o$ an $(m-n)$ null vector). As in the case of the parameter estimation, the observations are related to their expectations by $\underline{\tilde{y}} = \underline{y} + \underline{\tilde{\varepsilon}}$. We now want to find estimates of the expectations which satisfy the conditions, subject to $\underline{v}^T V_{\varepsilon\varepsilon}^{-1} \underline{v}$ being a minimum, where $\underline{v}^T = \underline{\tilde{y}} - \underline{\hat{y}}$. To solve this problem we use Lagrange multipliers i.e. we minimize the function, E, given by

$$E = \underline{v}^T V_{\varepsilon\varepsilon}^{-1} \underline{v} - 2\underline{K}^T (U\underline{v} - (\underline{U}_o - U\underline{\tilde{y}})) \qquad (E.11)$$

where $2\underline{K}^T$ are the Lagrange multipliers.

A stationary value of E will occur when $\partial E/\partial \underline{v} = 2V_{\varepsilon\varepsilon}^{-1} \underline{v} - 2U^T \underline{K} = 0$, and therefore E will have a stationary value when

$$\underline{v} = V_{\varepsilon\varepsilon} U^T \underline{K} \quad . \qquad (E.12)$$

The Lagrange multipliers can be determined be substituting $U\underline{v} = \underline{U}_o - U\underline{\tilde{y}}$ into the above equation which yields

$$\underline{K} = (U V_{\varepsilon\varepsilon} U^T)^{-1} (\underline{U}_o - U\underline{\tilde{y}}) \quad . \qquad (E.13)$$

We may caluculate the post-fit residuals by substituting Equation (E.13) into Equation (E.12). Hence,

$$\underline{v} = V_{\varepsilon\varepsilon} U^T (U V_{\varepsilon\varepsilon} U^T)^{-1} (\underline{U}_o - U\underline{\tilde{y}}) \quad . \qquad (E.14)$$

The covariance matrix of $\underline{v}$ will be given by $\langle \underline{v}\underline{v}^T \rangle$. Hence,

$$\langle \underline{v}\underline{v}^T \rangle = \langle V_{\varepsilon\varepsilon} U^T (U V_{\varepsilon\varepsilon} U^T)^{-1} (\underline{U}_o - U\underline{\tilde{y}})(\underline{U}_o - U\underline{\tilde{y}})^T (U V_{\varepsilon\varepsilon} U^T)^{-1} U V_{\varepsilon\varepsilon} \rangle$$

which reduces to

$$V_{vv} = V_{\epsilon\epsilon} - V_{\epsilon\epsilon}U^T(UV_{\epsilon\epsilon}U^T)^{-1}UV_{\epsilon\epsilon} \qquad (E.15)$$

assuming that $\langle\underline{\epsilon}\rangle = 0$, and that $\underline{U}_o$ is not a random vector.

We should note that the estimation using either parameters or conditions will yield identical results, provided that the parameterization is consistent with the conditions. The only distinction between the two techniques is the ease of implementation.

Equation (E.14) shows both the advantages and disadvantages of estimation using conditions. The matrix to be inverted in Equation (E.14) has the dimensions of the number of conditions. If the number of conditions is small, then Equation (E.14) can be easily solved. If the number of conditions is large, then the inversion of the matrix in Equation (E.14) is formidable.

One particular application of Equation (E.14) which we will investigate is "parameter suppression." Quite often after parameters have been estimated from a set of observations, we need to determine the changes that would occur to the parameter estimates if a subset of the parameters had not been estimated. We could, of course, repeat the estimation with these parameters not estimated, but this could be very time consuming. Estimations by condition provides an ideal solution to this problem.

We denote the estimated parameters by $(\Delta\underline{\hat{x}}|\Delta\underline{\hat{x}}_s)^T$ where we have partitioned the parameter estimates into the subsets; $\Delta\underline{\hat{x}}_s$

is the subset of parameters we wish to suppress, and $\Delta\underline{\hat{x}}$ are the remaining parameters. Clearly, had $\Delta\underline{\hat{x}}_s$ not been estimated then $\Delta\underline{\hat{x}}_s$ would equal a null vector. These are our conditions. The condition equation, (Equation E.10), becomes

$$(0 \; I) \left| \begin{array}{c} \Delta\underline{\hat{x}} \\ \Delta\underline{\hat{x}}_s \end{array} \right| = \underline{0}$$

where I is a (square) unit matrix with dimensions of the number of parameters to be suppressed. If we now substitute these conditions into Equation (E.14) with the covariance matrix of the parameter estimates partitioned as

$$V_{\epsilon\epsilon} = \left| \begin{array}{cc} V_{xx} & V_{xs} \\ V_{xs}^T & V_{ss} \end{array} \right|$$

we obtain

$$\left| \begin{array}{c} \Delta\underline{\hat{x}}' \\ \Delta\underline{\hat{x}}_s' \end{array} \right| = \left| \begin{array}{c} -V_{xs}V_{ss}^{-1}\Delta\underline{\hat{x}}_s \\ -\Delta\underline{\hat{x}}_s \end{array} \right| \qquad (E.16)$$

where $\Delta\underline{\hat{x}}'$ and $\Delta\underline{\hat{x}}_s'$ are the corrections to estimates of the parameter corrections which were obtained when all of the parameters were estimated. (Note, that when $\Delta\underline{\hat{x}}_s'$ are added to the original initial parameter estimates $(\Delta\underline{\hat{x}}_s)$, the net correction to the supressed parameters will be zero.)

The covariance matrix of $(\Delta\underline{\hat{x}}'|\Delta\underline{\hat{x}}_s')^T$ may be calculated from Equation (E.15). The result is

$$V_{x's'} = \left| \begin{array}{cc} V_{xx}-V_{xs}V_{ss}^{-1}V_{xs}^T & 0 \\ 0 & 0 \end{array} \right| \qquad (E.17)$$

where we note that $\Delta\underline{\hat{x}}_s'$ has zero variance because we assumed

these parameters were perfectly known when their estimation
was supressed.

If only one parameter is supressed, Equations (E.16) and
(E.17) reduce to

$$\Delta\hat{x}_j^! = -(\rho_{js}\sigma_j/\sigma_s) \ \Delta\hat{x}_s \qquad \text{(E.18.a)}$$

and

$$(\sigma_j^!)^2 = (1-\rho_{js}^2) \ \sigma_j^2 \qquad \text{(E.18.b)}$$

where $\Delta\underline{\hat{x}}_j^!$ is the estimate of the $\underline{j}$th parameter correction when
parameter $\underline{s}$ is suppressed, ρ_{js} is the correlation between the
parameter estimates, σ_j^2 and σ_s^2 are the variances of the para-
meter estimates, and $(\sigma_j^!)^2$ is the variance of the new estimate
of the $\underline{j}$ th parameter. These simple formulas are very useful
for studying the interaction between parameters.

We now return to Equation (E.6) to study methods which
can be used to solve this equation when the normal equations
become too large to be handled in the main memory of the
computer. The technique we will develop is called sequential
least squares (see, e.g., Kaula, 1966; Morrison, 1969). We
wish to find an equivalent solution to Equation (E.6) which
does not involve inverting the complete normal equations at
one time. We can find such a solution provided that none of
the observations depend on all the parameters. We will
subdivide the parameters into two types: global parameters,
which affect large numbers of observations (e.g., station
positions and radio source positions); and local parameters,
which affect a relatively small number of observations (e.g.

clock and atmosphere polynomial coefficients).

There are many techniques which we could use to develop such a solution. One method which is particularly straight-forward is to treat the parameters as observations. (We have already been doing this; remember, the group delay measure-ments are themselves estimates of parameters obtained from estimating the intercept and slope of the video DC phases; see Apperdix B.)

We commence by extending Equation (E.2) to include the parameters as observations, which yields

$$\begin{vmatrix} \Delta \tilde{y} \\ \Delta \tilde{x} \end{vmatrix} = \begin{vmatrix} A \\ I \end{vmatrix} \Delta \underline{x} + \begin{vmatrix} \underline{\varepsilon} \\ \underline{\varepsilon}_x \end{vmatrix} \qquad\qquad (E.19)$$

where $\Delta \tilde{x}$ are a priori estimates of the parameter corrections with errors $\underline{\varepsilon}_x$ whose covariance matrix is N_g^{-1}. If we now apply the estimator Equation (E.4) to the above equation we find

$$\Delta \hat{x} = \begin{bmatrix} A^T & I \end{bmatrix} \begin{vmatrix} V_{\varepsilon\varepsilon}^{-1} & 0 \\ 0 & N_g \end{vmatrix} \begin{vmatrix} A \\ I \end{vmatrix} \begin{bmatrix} A^T & I \end{bmatrix} \begin{vmatrix} V_{\varepsilon\varepsilon}^{-1} & 0 \\ 0 & N_g \end{vmatrix} \begin{vmatrix} \Delta \tilde{y} \\ \Delta \tilde{x} \end{vmatrix}$$

which reduces to

$$\Delta \hat{x} = (A^T V_{\varepsilon\varepsilon}^{-1} A + N_g)^{-1} (A^T V_{\varepsilon\varepsilon}^{-1} \Delta \tilde{y} + N_g \Delta \tilde{x}) \qquad (E.20)$$

Equation (E.20) is the sequential least squares estimator. This equation is very easily implemented into a standard least squares program because the normal equations are formed exactly the same way as they normally would be formed $(A^T V_{\varepsilon\varepsilon}^{-1} A)$ and then, before inversion, the inverse of the a

priori covariance matrix is added to them. Similarly, the correction vector $(\mathbf{A}^T V_{\varepsilon\varepsilon}^{-1} \Delta \tilde{y})$ is calculated as it normally would be calculated, and then the a priori correction vector $(\mathbf{N}_g \Delta \tilde{x})$ is added. When the solution given by Equation (E.20) is completed, the covariance matrix of the global parameters are "stripped" from the inverse of the normal equations to yield an updated global parameter covariance matrix which can be used when the next set of data is to be included in the solution.

Once all of the data have been used in the sequential or global solution, the estimates of the global parameters are the same as those estimates which would have been obtained from a standard least squares solution using all of the data simultaneously. The estimates of the local parameters for each data set, except the last, which were obtained as the sequential solution was carried out are, however, not correct (because their estimation did not use the information about the global parameters which would be obtained from future data sets).

To recover the estimates of the local parameters a back solution is run. To construct the back solution we use the property of sequential least squares solutions, that the local parameters for the last data arc used are recovered correctly (because all of the information about the global parameters has been included in solution). To use this property in the back solution we essentially remove the contribution of a data arc from the (full) global solution and then add it back in as

if it were the last arc in the solution. This may sound like a sleight of hand, but the technique should work.

The implementation of this technique is relatively straight-forward but is more complicated than the forward solution.

To see how the back solution is carried out we will run through the matrix manipulations needed to process one arc during the back solution. The relationship between the observations and the parameters (separated into local and global parameters, Δx_ℓ and Δx_g, respectively) will be (from Equation E.2)

$$\Delta \tilde{y} = [A_g \; A_\ell] \begin{vmatrix} \Delta x_g \\ \Delta x_\ell \end{vmatrix} + \tilde{\epsilon} \tag{E.21}$$

The normal equations for the solution will be given by (including the <u>a priori</u> covariances for global parameters),

$$\begin{vmatrix} A_g^T V_{\epsilon\epsilon}^{-1} A_g + (N_g^{f-1})^{-1} & A_g^T V_{\epsilon\epsilon}^{-1} A_\ell \\ A_g V_{\epsilon\epsilon}^{-1} A_\ell^T & A_\ell V_{\epsilon\epsilon}^{-1} A_\ell^T \end{vmatrix} \begin{vmatrix} A_g^T V_{\epsilon\epsilon}^{-1} \Delta\tilde{y} + (N_g^{f-1})^{-1} \Delta\tilde{x}_g^{f-1} \\ A_\ell^T V_{\epsilon\epsilon}^{-1} \Delta\tilde{y} \end{vmatrix} = \begin{vmatrix} \Delta\hat{x}_g \\ \Delta\hat{x}_\ell \end{vmatrix} \tag{E.22}$$

where $(N_g^{f-1})^{-1}$ is the covariance matrix of the global parameters which would be obtained from all of the data excepting the arc which is being processed (in general this matrix is not known). In order to reduce the number of symbols needed in future equations, we will write Equation (E.22) as

$$\begin{vmatrix} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{vmatrix} \begin{vmatrix} z_g \\ z_\ell \end{vmatrix} = \begin{vmatrix} \hat{x}_g \\ \hat{x}_\ell \end{vmatrix} \tag{E.23}$$

where $B_{11} = A_g^T V_{\varepsilon\varepsilon}^{-1} A_g + (N_g^{f-1})^{-1}$, $B_{12} = A_g^T V_{\varepsilon\varepsilon}^{-1} A_\ell$, $B_{22} = A_\ell^T V_{\varepsilon\varepsilon}^{-1} A_\ell$, $\tilde{z}_g = A_g^T V_{\varepsilon\varepsilon}^{-1} \Delta\tilde{y} + (N_g^{f-1})^{-1} \Delta x_g^{f-1}$, and $\tilde{z}_\ell = A_\ell^T V_{\varepsilon\varepsilon}^{-1} \Delta\tilde{y}$. We should note the matrices B_{12} and B_{22}, and the vector $\tilde{z}_\ell$ are known, but the matrix, B_{11}, and the vector, $\tilde{z}_g$, are not known (unless this arc was the last arc to be used during the forward solution). The matrix B_{11} and the vector $\tilde{z}_g$ may be calculated by treating this arc as if it were the last arc in the solution.

If Equation (E.23) represented the last arc in the solution, then the inverse of the normal equations would be

$$\begin{vmatrix} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{vmatrix}^{-1} = \begin{vmatrix} (N_g^f)^{-1} & C_{12} \\ C_{12}^T & C_{22} \end{vmatrix} \tag{E.24}$$

where $(N_g^f)^{-1}$ is the final global parameter covariance matrix. (The other elements in the inverse C_{12} and C_{22} do not interest us at this time.) We may find an expression for $(N_g^f)^{-1}$ in terms of B_{11}, B_{12} and B_{22}, by performing a partitioned inverse of the normal equations. This partitioned inverse will be given by (see, e.g., Beyer, 1978 p.11)

$$(N_g^f)^{-1} = (B_{11} - B_{12} B_{22}^{-1} B_{12}^T)^{-1}.$$

By re-arranging the above equation, we can obtain an expression for unknown matrix B_{11}, i.e.,

$$B_{11} = N_g^f + B_{12} B_{22}^{-1} B_{12}^T. \tag{E.25}$$

To find the unknown vector $\tilde{z}_g$, we use a similar procedure.

From Equations (E.22) and (E.24), we have

$$\begin{vmatrix} (N_g^f)^{-1} & C_{12} \\ C_{12}^T & C_{22} \end{vmatrix} \begin{vmatrix} z_g \\ z_\ell \end{vmatrix} = \begin{vmatrix} \hat{x}_g \\ \hat{x}_\ell \end{vmatrix}$$

and the estimates of the global parameters will be given by

$$(N_g^f)^{-1} z_g + C_{12} z_\ell = \hat{x}_g \tag{E.26}$$

Again using the partitioned inverse of the left hand side of Equation (E.24), we can express (the unknown) C_{12} matrix in terms of the (known) matrices B_{11} (Equation B.25), B_{22} and $B1_2$, i.e.,

$$C_{12} = -(N_g^f)^{-1} B_{12} B_{22}^{-1} .$$

Therefore

$$\hat{x}_g = (N_g^f)^{-1} z_g - (N_g^f)^{-1} B_{12} B_{22}^{-1} z_\ell$$

which may be re-arranged to yield

$$z_g = (N_g^f)[\hat{x}_g + B_{12} B_{22}^{-1} z_\ell] \tag{E.27}$$

Now we can substitute the calculated matrix B_{11} and vector z_g into the normal equations and the back solution can be completed using the standard SOLVE software.

Global solution capability was added to SOLVE by adding three programs to the SOLVE suite of programs. Program INITL will save the global parameter adjustments and covariance matrix from any SOLVE solution. Program APCOV (APply COVariances) will add the covariances saved by INITL to the

normal equations (this program is automatically scheduled by
SOLVE) and APBAK will carry out the matrix multiplications for
the back solutions (again this program is automatically
scheduled).

The global solution software has some limitations
(imposed by the computer size). Currently, the maximum number
of parameters which can be estimated during any single arc is
95. As data from many experiments are combined the number of
global parameters (source and station positions) can get
sufficiently large that the 95 parameter limit for an arc can
be exceeded. There are two ways to overcome this problem. The
simplest method is to suppress some of the global parameters.
This could be done for the positions of radio sources which
are only used in a single experiment. The utility of this
technique is limited, however. An alternative technique,
which has not yet been implemented in the SOLVE programs, is
to combine separate global solutions together, i.e., if the
number of global parameters becomes too large, a second global
solution is started. When this second solution is completed,
the two covariance matrices are combined to yield the total
global so ition.

The mathematical model which would be used to combine the
solutions is

$$\left|\begin{matrix} I \\ \theta \end{matrix}\right| \left|\begin{matrix} \hat{x}_g^{(1)} \\ \hat{x}_g^{(2)} \end{matrix}\right| = \left|\begin{matrix} \hat{x}_g^f \\ \hat{r} \end{matrix}\right| \tag{E.28}$$

where θ is a rotation matrix and $\underline{r}$ is vector of rotation

parameters which accounts for rotations of the crust-fixed and
"inertial" coordinate systems (see Chapter 5) between the two
global solutions. Equation (E.28) can be solved using the
techniques which have already been described.

Many of the techniques introduced in this appendix are
used in other sections of this thesis and the practical
implementation of these techniques may be clearer when they
are discussed in these sections.

Appendix F. Effects of polarization leakage

The Mark III VLBI system nominally observes right-circularly polarized (RCP) radiation. However, since perfect feed horns can not be made some amount of the left-circularly polarized (LCP) radiation from the radio source will "leak" into the output of the nominally RCP feed horn. In this appendix we consider the effects of such a leakage on the group- and phase-delay measurements.

In order to examine the effects of polarization leakage, we commence by writing the expressions for the spectra of the signals at the outputs of the feed horns of the antennas which form an interferometer. We do this by writing a more general-ized *form of Equation (A.5)*. The spectra of the signals, $\bar{X}_1(\omega)$ and $\bar{X}_2(\omega)$, at the outputs of the feed horns will be given by (see Appendix A for introductory discussions):

$$\bar{X}_1(\omega) = \bar{R} + \bar{\beta}\bar{L} + \bar{N}_1' \tag{F.1.a}$$

$$\bar{X}_2(\omega) = K\bar{R}' + K\bar{\gamma}\bar{L}' + \bar{N}_2' \tag{F.1.b}$$

where $\bar{R}$ and $\bar{R}'$ are the RCP components of the signal; $\bar{L}$ and $\bar{L}'$ are the LCP components of the signal; $\bar{N}_1'$ and $\bar{N}_2'$ are noise components; K is a constant which represents the difference in sensitivity of the two antennas (see Equation (A.5)); $\bar{\beta}$ and $\bar{\gamma}$ are complex constants which express the amount of LCP radia-tion leaking into the outputs of the feed horns. These constants are complex to allow for phase shifts of the LCP signals relative to RCP signals after these signals have

propagated through the feed horns. Equations (F.1) are derived from Equation (A.5) by explictly writing the signals $\bar{S}$ and $\bar{S}'$ in terms of their RCP and LCP components. For perfect RCP feed horns $\bar{\beta}$ and $\bar{\gamma}$ would both be zero. We now proceed in a manner similar to the analysis which was used in examining Equation (A.5). The RCP components $\bar{R}$ and $\bar{R}'$ are related to each other by (see Appendix A for details):

$$\bar{R}'(\omega) = \bar{R}(\omega) \ e^{-i\alpha \ + \ \Delta\psi} \qquad\qquad (F.2.a)$$

where $\alpha = (\omega-\omega_o)\tau_g + \phi_t + \omega t_p\Delta t$; ω_o is the reference (angular) frequency to which the visibility phase, ϕ_t, is referred; τ_g is the group delay; t_p is the phase delay rate; Δt is the time interval between the epoch to which the delay measurements are referred and the temporal center of the data used to compute the spectra (see Equation (A.5) for more detailed discussions); and $\Delta\psi$ is the difference of the feed horn position angles at the two antennas (see Appendix D for details). The difference of the feed horn position angles was implicity included in ϕ_t in the derivation of Equation (A.5). We explicity include it here because its effect on the LCP signals is opposite signed to its effect on the RCP signals. We may write an expression similar to Equation (F.2.a) for the LCP signals:

$$\bar{L}'(\omega) = \bar{L}(\omega) \ e^{-i\alpha \ - \ \Delta\psi} \qquad\qquad (F.2.b)$$

where we note that the feed position angle, $\Delta\psi$, affects the phase of $\bar{L}'(\omega)$ with opposite sign of its effect on $\bar{R}'(\omega)$.

The cross spectrum of the signals at the output of the

feed horns will be

$$\bar{X}_1\bar{X}_2^* = K\bar{R}\bar{R}'^* + K\bar{\gamma}^*\bar{R}\bar{L}'^* + \bar{R}\bar{N}_2'^* + K\bar{\beta}\bar{L}\bar{R}'^* + K\bar{\beta}\bar{\gamma}^*\bar{L}\bar{L}'^*$$

$$+ \bar{\beta}\bar{L}\bar{N}_2'^* + K\bar{N}_1\bar{R}'^* + K\bar{\gamma}^*\bar{N}_1'\bar{L}'^* + \bar{N}_1'\bar{N}_2'^*$$

and the expectation of this cross sprectrum is

$$\langle\bar{X}_1\bar{X}_2^*\rangle = K\bar{R}^2 e^{-i\alpha+\Delta\psi} + K\bar{\beta}\bar{\gamma}^*\bar{L}^2 e^{-i\alpha-\Delta\psi} \qquad (F.3)$$

where we have assumed that the radiation from the radio source is unpolarized (i.e., $\langle\bar{L}\bar{R}^*\rangle = 0$), and we have substituted the relationships between the RCP and the LCP signals received at the two antennas (Equations (F.2)). This cross spectrum is shown graphically in phasor form in Figure F.1. We see from this figure that the angle between the LCP and RCP phasors is $2\Delta\psi+\arg(\bar{\beta}\bar{\gamma}^*)$. If the LCP signal is small compared to the RCP signal, i.e., $\mathrm{mag}(\bar{\beta}\bar{\gamma}^*)<<1$, we may express the error in the visibilty phase, $\Delta\phi_\ell$, due to the leakage of the LCP radiation as

$$\Delta\phi_\ell \simeq \mathrm{mag}(\bar{\beta}\bar{\gamma}^*)\ \sin(2\Delta\psi + \arg(\bar{\beta}\bar{\gamma}^*)). \qquad (F.4)$$

The group delay error due to the polariztion leakage can be obtained by differentiating $\Delta\phi_\ell$ with respect to (angular) frequency:

$$\Delta\tau_\ell = \mathrm{mag}\left(\ \partial(\bar{\beta}\bar{\gamma}^*)/\partial\omega\ \right)\ \sin(2\Delta\psi + \arg\left(\ \partial(\bar{\beta}\bar{\gamma}^*)/\partial\omega\ \right)$$

which can written in the form

$$\Delta\tau_\ell = \Delta\tau_0^p\ \sin(2\Delta\psi + \xi_0^p) \qquad (F.5)$$

where $\Delta\tau_0^p = \mathrm{mag}\left(\ \partial(\bar{\beta}\bar{\gamma}^*)/\partial\omega\ \right)$ and $\xi_0^p = \arg\left(\ \partial(\bar{\beta}\bar{\gamma}^*)/\partial\omega\ \right)$.

We should note that the polarization leakage will

Figure F.1 Phasor diagram of the effects of polarization leakage

Figure F.1 Phasor diagram of the effects of polarization leakage
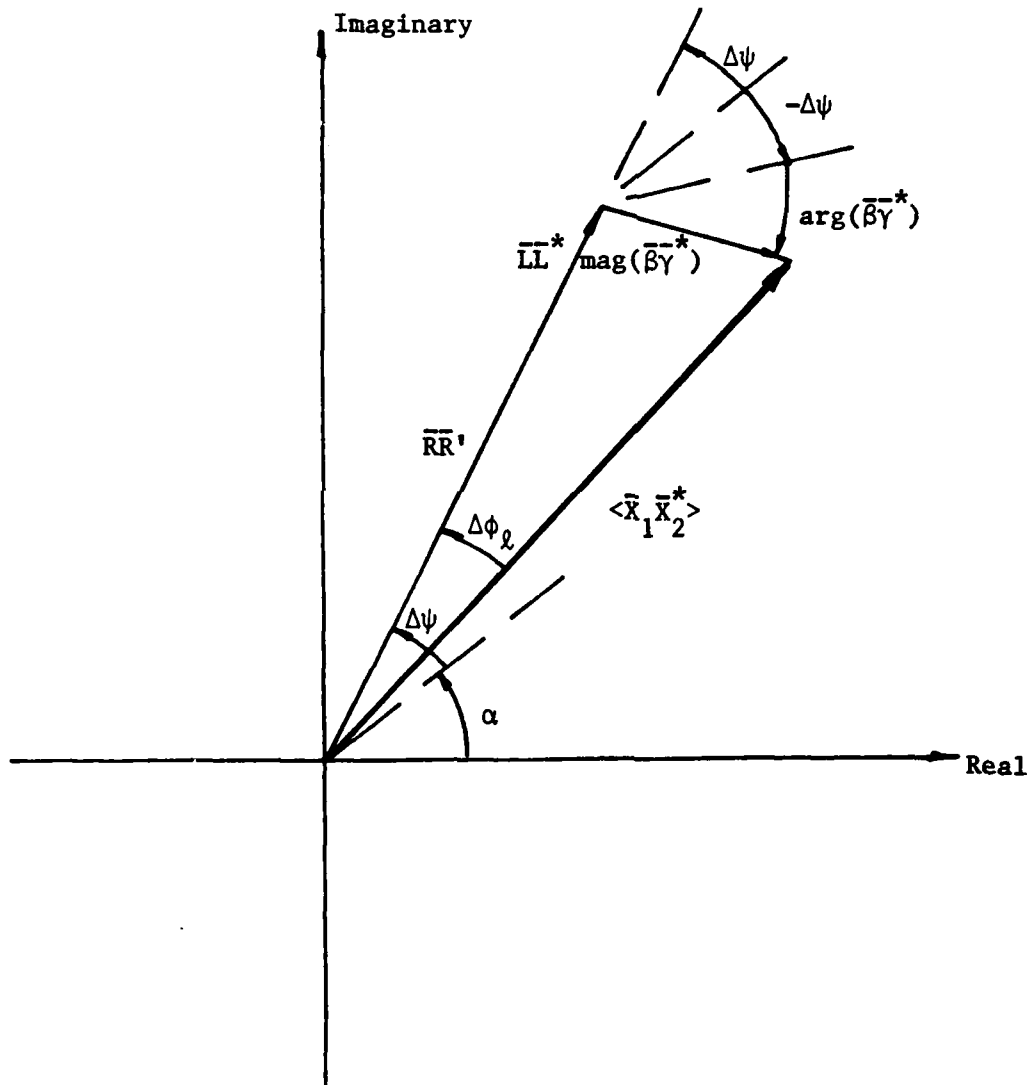
introduce errors in the group delay and the visibility phase measurements only if the feed horns at both antennas in an interferometer are leaking LCP signals, i.e., both $\bar{\beta}$ and $\bar{\gamma}$ must be non-zero. For this reason, polarization leakage can cause closure errors when two antennas in a triplet of of antennas have defective feed horns. Only the measurements with the interferometer which contains the two antennas with defective feed horns will be affected by the LCP signals. The closure errors which are believed to be due to polarization leakage are discussed in Section 3.3.

References

Altschuler, J. F., C. Wardle, Observations of the flux density and linear polarization of compact extragalatic radio sources at 3.7 and 11.1 cm wavelength, Mon. Not. R. Astron. Soc., 179, 153-178, 1977.

Beyer, W.H., CRC Stand Mathematical Tables, CRC Press Inc., West Palm Beach Fla., 613,1978.

Bjerhammar, A. Theory of errors and generalized matrix inverses, Elsevier Scientific Publishing Company, Amsterdam, 420, 1973.

Bjerhammar, A. Generalized matrix inverses, Methoden und Verfahren der mathematischen physics/ Band 14, Mathematical Geodesy, Part III, Bibliographisches Institut Mannheim, Wein, Zurich, 47-82, 1975.

Bogusch, R.L., F.W. Guigliano, D.L. Knepp, and A.H. Michelet, Frequency Selective Propagation Effects on Spread-Spectrum Receiver Tracking, Proc. IEEE, 69, #7, 787-796, 1981.

Booker, H.G., The role of the magnetosphere in satellite and radio star scintillation, J. Atmos. and Ter. Phys., 37, 1089-1098, 1975.

Bracewell, R.N., The Fourier Transfrom and its Applications, McGraw Hill, New York, 144, 1978.

Breidenthal, J. Master's Thesis, MIT, 1982.

Carter, W.E., Project Polaris: A Status Report, Radio
Interferometry Techniques for Geodesy, <u>NASA Conference
Publication 2115</u>, 455-460, 1980.

Clark, T.A., C.C. Counselman, P.G. Ford, L.B. Hinteregger,
W.J. Klepczynski, C.A. Knight, D.S. Robertson, A.E.E.
Rogers, J.W. Ryan, I.I. Shapiro and A.R. Whitney,
Synchronization of Clocks by Very-Long-Baseline Inter-
feromtetry, <u>IEEE Trans.</u>, <u>IM-28</u>, #3, 184-187, 1979.

Conway, R.G., B.J. Burn and J.P. Valle, Measurement of
structure and polarization of 72 sources from the 4C
catalogue, <u>Astron. Astrophys., suppl. ser.</u>, <u>27</u>, #2,
155-170, 1977.

Costa, E., and M.C. Kelley, Calculations of Equatorial
Scintillations of VHF and gigahertz frequencies based on
a new model of the disturbed equatorial ionosphere, <u>Geop.
Res. letters</u>, <u>3</u>, #11, 677-680, 1976.

Cotton, W.D., Source structure corrections to the geodetic
very-long-baseline interferometry observables. <u>Radio
Interferometry Techniques for geodesy</u>, NASA Conference
Publication 2115, 193-197, 1979.

Davenport, W.B. and N.L. Root, <u>An Introduction to the Theory
of Random Signals and Noise</u>, McGraw Hill, New York, 1958.

Drake, A.W., <u>Fundamentals of applied probability theory</u>,
McGraw Hill, New York, 283, 1967.

Frenow, E.J. R.L. Leadabrand, R.C. Livingston, M.D. Cousins, C.L. Rino, B.C. Fair, and R.A. Long, Early Results from the DNA Wideband satellite experiment- Complex -Signal scintillation, Radio Science, 13, #1, 167-187, 1978.

Haves, P., Polarization parameters of 183 extragalactic radio sources, Mon. Nat.R. Astr. Soc., 173, 553-568, 1975.

Hildebrand, F.B., Advanced Calculus for Application, Prentice-Hall, Inc., Englewood Cliffs, New Jersey,733, 1976.

Hinteregger, H.F., The Mark III Wideband Digital Recorder in perspective, Radio Interferometry Techniques for Geodesy, NASA Conference Publication 2115, 305-316, 1980.

Hinterregger, H.F., The Density Upgrade: Mark IIIA (A Future Improvement of the Mark III VLBI System), Proceedings of Symposium No. 5: Geodetic Applications of Radio Inter- ferometry, NOAA Technical Report NOS 95 NGSS 24, U.S. Department of Commerce National Oceanic and Atmospheric Administration, 334, 1982.

Jackson, J.D., Classical Electrodynamics, John Wiley and Sons, New York, 1975.

Kane, J.A., Artic Measurements of Electron Collision Frequencies in the D-Region of the Ionosphere, J. Geophys. Res., 64, 133-139, 1959.

Kaula, W.M., The theory of satelite geodesy, Blaisdell Publishing Co., Waltham, Mass., pp.125, 1966.

Kong, J.A., <u>Theory of elctromagnetic waves</u>, John Wiley & Sons, New York, 339, 1975.

Krauss, J.D., <u>Radio Astronomy</u>, McGraw Hill, New York, 97-99, 1966.

Lieske, J.H., T. Lederle, W. Fricke, and B. Morando, Expressions for the precession quantities based upon the IAU (1976) System of astronomical constants, <u>Astron. Asrophys.</u>, <u>58</u>, 1-16, 1977.

Lorrain, P., and D.R. Corson, <u>Electromagnetic Fields and Waves</u>, W.H. Freeman and Company, San Francisco, Calif., 706, 1970.

Ma, C., <u>NASA Goddard Report X-592-73-259</u>, Goddard Space Flight Center, 1973.

Marcaide, J.M., VLBI studies of the extragalactic radio sources 1038+528A,B. Ph.D. Thesis, MIT, 1982.

Mason, S.J. and H.J. Zimmerman, <u>Electronic Circuits, Signals and Systems</u>, John Wiley and Sons, New York, 616, 1960.

Mark III System Documentation, Vol. 1 and Vol. 2, 1980, (For information contact A.E.E. Rogers, Haystack Observatory, Westford, MA, 01886).

Morrison, N., <u>Introduction to sequential smoothing and prediction</u>, McGraw-Hill Book Company, New York, 645, 1969.

Murden, J., E-Region temperature and con-neutral collision frequency from incoherent scatter radar mono-pube ACF measurements, <u>Journal of Atmospheric and Terrestrial Physics</u>, <u>43(1)</u>, 13-21, Jan. 1981.

Nicolet, M., Collision frequency of Electrons in the Terrestrial Atmosphere, _Physics of Fluids_, _2_, #2, 95-99, 1959.

Rino, C.L., A power law phase scrum model for ionospheric scintillation 1. Weak scatter, _Radio Science_, _14_, #6, 1135-1145, 1979(a).

Rino, C.L., A power law phase scrum model for ionospheric scintillation 2. Strong scatter, _Radio Science_, _14_, #6, 1147-1155, 1979(b).

Robertson, D.S., Geodetic and Astrometric measurements with very-long-baseline interferometry. Ph.D. Thesis, MIT, 1975.

Robertson, D.S., and W.E. Carter, Operation of the National Geodetic Survey Polaris Network, Proceedings of Symposium No. 5: Geodetic Applications of Radio Interferometry, _NOAA Technical Report NOS 95 NGS 24_, U.S. Department of Commerce, 334, 1982.

Rogers, A.E.E., Very-long-baseline interferometry with large effective bandwidth for phase-delay measurement, _Radio Science_, _5_, 1239-1248, 1970.

Rogers, A.E.E., Broadband Passive 90° RC Hybrid with Low Component for Use in the Video Range of Frequencies, _Proc. IEEE_, _59_, 1617-1618, 1971.

Rogers, A.E.E., Phase and Group delay Calibration of a Very Long Baseline interferometer by East Coast VLBI Group, Radio Interferometry Techniques for Geodesy, _NASA Conference Publication 2115_, 255-261, 1980.

Rogers, A.E.E., Performance tests of the Mark III and S/X receivers, <u>Memorandum Northeast Radio Observatory Corporation</u>, 29 October, 1981.

Rogers, A.E.E., R.J. Cappallo, H.F. Hinteregger, J.R. Levine, E.F. Nesman, J.C. Weber, A.R. Whitney, T.A. Clark, C. Ma, J. Ryan, B.E.Corey, C.C. Counselman, T.A. Herring, I.I. Shapiro, C.A. Knight, D.B. Schaffer, N.R. Vandenberg, R. Lacasse, R. Mauzy, B. Rayhrer, B. Schupler, and J.C. Pigg, Very-long Baseline Radio Interferometry: The Mark III System for Geodesy, Astrometry, and Aperture Synthesis, <u>Science</u>, <u>219</u>, 51-54, 1983.

Schlegel, K., H. Kohls and K. Rinmert, Temperatures and Collision frequency in the polar E-Region, measured with the incohert scatter techniques, <u>J. Geophys. Res.</u>, <u>85</u>, 710-714, 1980.

Shapiro, I.I., Possible experiments with long-baseline interferometers, <u>NEREM Record</u>, 10, 70-71, 1968.

Shapiro, I.I., and C.A. Knight, in <u>Earthquake Displacement and the rotation of the Earth</u>, Reidel, 285, 1970.

Shapiro, I.I., J.J. Wittels, C.C. Counselman, D.S. Robinson, A.R. Whitney, H.F.Hinteregger, C.A. Knight, A.E.E. Rogers, T.A. Clark, L.K. Hutton and A.E. Niell, Submilli-arsecond Astrometry Via VLBI. I. Relative position of the radio sources 3C345 and NRAO 512, <u>Astron. J.</u>, <u>84</u>, 1459-1469, 1979.

Simard-Normandin, M., P.P. Kronberg and J. Neidhofer, Linear
    Polarization observations of extragalactic radio sources
    at   2 cm and at 17-19 cm., Astron. Astrophys. Suppl.
    Ser., 43, #1, 19-22, 1981.

Simard-Normandin, M., P.P. Kronberg, and S. Button, Integrated
    linear polarization of extragalactic radio sources at
    10.5 GHz (2.86 cm.) II, Astron, Astrophys. Suppl. Ser.,
    48, 137-138, 1982.

Stacey, F.D., Physics of the Earth, John Wiley and Sons, Inc.,
    New York, 414, 1977.

Wahr, J.M., The forced nutations of an elliptical, rotating,
    elastic, and oceanless Earth, Geophys. J. R. Astron.
    Soc., 64, 705-727, 1981.

Whitney, A.R., Precision geodesy and astrometry via very long
    baseline interferometry, Ph.D. thesis, Mass. Inst. of
    Tech., Cambridge, Mass., 1974.

Whitney, A.R., A.E.E. Rogers, H.F. Hinteregger, C.A. Knight,
    J.I. Levine and S. Lippincott, A Very-long-baseline
    interferometer system for geodetic applications, Radio
    Science, 11, #5, 421-431, 1976.

Wittels, J.J., Positions and Kinematics of Quasars and related
    radio objects inferred from VLBI observations, Ph.D.
    thesis, Mass. Inst.of Tech., Cambridge, Mass.,1975.

Wozencraft, J.M. and I.M. Jacobs, Principles of Communication
    Engineering, John Wiley and Sons, Inc., New York, 1965.

END

FILMED

4-85

DTIC