Imaging Active Galactic Nuclei with 3mm-VLBI

by

Sheperd S. Doeleman

B.A. Physics, Reed College (1986)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1995

Author . . . .
Department of Physics
May 18, 1995

Certified by . . . .
Alan E. E. Rogers
Senior Research Scientist
Assistant Director - Haystack Observatory
Thesis Supervisor

Certified by . . . .
Bernard F. Burke
Professor - Physics Department
Thesis Supervisor

Accepted by . . . .
George F. Koster
Chairman, Departmental Committee on Graduate Students

Imaging Active Galactic Nuclei with 3mm-VLBI

by

Sheperd S. Doeleman

Submitted to the Department of Physics
on May 19, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

The technique of Very Long Baseline Interferometry (VLBI) allows data recorded simultaneously at distant telescopes to be combined to provide information on the smallest angular scales. To increase resolution we have continued an ongoing effort to higher frequencies. At a frequency of 86GHz, the coherence limitation of the atmosphere becomes severe and attention to accurate recovery of the complex visibility on each baseline is crucial. Traditional detection methods that do not take into account the coherence losses at high frequencies are inadequate to detect weak sources.

The first goal of this project was to develop mm-VLBI capabilities in areas of both hardware and software. As a starting point, Chapter 2 reviews basic relations of interferometry to aid in later discussion. Chapter 3 details the design, construction, and testing of a polarizing window for the Haystack antenna. Chapter 4 begins with a review of VLBI detection methods and then describes the new incoherent methods developed to lower detection thresholds for low coherence data. Application to mm-VLBI data is explained and methods of recovering phase information are discussed.

The second part of the thesis was the scheduling and observation of a global 3mm-VLBI experiment. Chapter 5 gives the particulars of antennas, sources, scheduling, calibration and data reduction. Chapter 6 focuses on the source 3C111 which was included in the observations only as a test source. It is a nearby radio galaxy and with its widely spaced radio lobes is atypical of the core dominated sources that make up most of the source list. We examine the possibility that 3mm observations may be able to set constraints on its orientation, investigate physical parameters of the core, and find an improved position using VLBI observables. Chapter 7 describes observations of the galactic center source SgrA*, thought to be an AGN in our own galaxy. A reanalysis of data taken in 1994 is performed and a new size limit for this compact object is found which is consistent with the scattering size for a point source in the ISM. Two different models of emission are discussed and one is shown to be inconsistent with observations.

Thesis Supervisor: Alan E. E. Rogers
Title: Senior Research Scientist
Assistant Director - Haystack Observatory

Thesis Supervisor: Bernard F. Burke
Title: Professor - Physics Department

My Family

Acknowledgments

Contents

List of Figures

9

10

11

List of Tables

Chapter 1

Introduction

For two decades, radio astronomers have used the technique of Very Long Baseline Interferometry (VLBI) to make images of the sky at radio wavelengths. This technique combines signals recorded on a network of widely spaced antennas to achieve exceptional resolution - up to 1000 times that of optical images. This resolution depends on both antenna separation (B) and observed wavelength as λ/B. The effort to increase resolution has, therefore, involved both increasing the separation of antennas and increasing the observing frequency. The baseline length is limited by the Earth's diameter but high frequency observing is limited primarily by instrumention and data analysis techniques. This thesis aims to apply new methods of analysis to mm-wavelength VLBI data which increase sensitivity and overcome some problems inherent in high frequency observations.

1.1   Motivation to mm-VLBI

- High Resolution

  The motivation to pursue mm-λ VLBI begins with the increased resolution which, on long baselines, can approach $\sim 50\mu as$ or one hundered millionth of a degree. A resolving power of this magnitude is currently the only method available to model and image the nuclei of highly active galaxies (AGN). Current theories of AGN postulate accretion of matter onto a very massive black hole to explain the high luminosity cores and jet structure seen in these sources (Blandford,1990). At the size scales probed by 3mm-VLBI (down to 0.02 light years for the nearest sources) the resolution will be approaching the base of the jet.

- Self-Absorption

  Higher frequency VLBI can also look deeper into those sources whose cores are optically thick at lower frequencies. This issue arises because in most AGN the emission mechanism appears to be synchrotron radiation resulting from charged particles interacting with a magnetic field. When this radiation is intense and the emission region compact, it is possible for the charged particles to re-absorb the radio photons - a process which masks the true source structure with an optically thick shell. This typically happens when the brightness temperature becomes comparable to the electron energy : $kT \sim \gamma m_e c^2$. At lower frequencies this problem is pronounced and is the reason for the spectrally inverted cores seen in many cm-VLBI maps. This effect also explains the behavior of the quasar 3C273 during a flare in 1990 that was monitored at many frequencies (Robson, et al., 1993). The light curves of this source show increased emission at all wavelengths shortward of 3.3mm but *none* at 8.1mm. At 3mm we begin to see emission during the flare from regions which are invisible at longer wavelengths due to self-absorption.

- Plasma Effects

  3mm-VLBI also reduces the effects of scattering by the interstellar medium in our galaxy. The scattering size of a point source near the galactic plane is proportional to λ^2. Observations of the galactic center SgrA* (Rogers, et al., 1994; Krichbaum, et al., 1994) are limited by this scattering as are data on all sources at low galactic latitudes such as CygA (Carilli, et al., 1991). The 3mm-VLBI study of SgrA* (see Chapter 7) specifically depends on the reduced scattering to set new size limits and rule out emission models.

  Faraday rotation also increases as λ^2 and the ISM in our galaxy and in host galaxies of AGN may hide fine polarization details at low frequencies. Interest in this aspect of 3mm-VLBI is increasing as more dual-polarization antenna become available. Polarization experiments at 3mm have just begun but will certainly probe a new regime of AGN.

1.2 Sensitivity

The long baselines which give VLBI its high resolving power require that information be recorded separately at each antenna and correlated later at a central location. This adds an extra step in the detection process in which the complex correlation is computed at many different time lags in a search for peak correlation. Detection depends on integrating the correlated signal in time. Traditionally, this has been done using a vector average within the so called coherence time : the time after which phase noise in the signal will reduce the average. Within the coherence time, the interferometric SNR can be written as

$$\text{SNR} = \sqrt{\frac{A_1 A_2 \epsilon_1 \epsilon_2}{T_{s1} T_{s2}}} \left(\frac{S_{source}}{2k} \right) \sqrt{2 \Delta \nu \tau_c} \qquad (1.1)$$

where T_s is the antenna system temperature, A is the collecting area, ϵ is aperture efficiency, $\Delta \nu$ is bandwidth, and τ_c is the coherence time. To get a feel for the capabilities of modern day 3mm-VLBI, it is illustrative to compare its typical SNRs with those of cm-VLBI when it was first being established. Table 1.1 gives averaged values for antennas and acquisition systems of these two periods. To put the SNR figures in perspective we note that a SNR

	cm(1975)	mm(1995)
Diam.	50m	25m
Eff.	50%	20%
T_{sys}	100K	300K
$\Delta \nu$	2MHz	112MHz
τ_c	250sec	20sec
SNR	60	4

Table 1.1: Comparison of array sensitivities between cm-λ twenty years ago and mm-λ today. SNR calculations are for a 1Jy source.

of 7 is needed for reliable detection. The exact numbers in Table 1.1 can be shifted somewhat but the general concensus is that detection was simply easier by a factor of ~ 10 at the start of cm-VLBI than it is twenty years later for mm-VLBI. This is due mostly to the brief coherence times which result from path length variations in the atmosphere and high frequency contamination of receiver oscillators. Even in moderate weather conditions coherence problems can reduce detection levels and let weaker sources go undetected.

To deal with this problem we have reformulated the search process in terms of inco-

16

herently averaged quantities which are immune to phase noise. With this method one can average beyond the coherence time although the SNR does not increase as fast as $\sqrt{t}$. If we incoherently average 20 second coherence intervals for the duration of a 6.5 minute scan then the detection threshold corresponding to the mm SNR in Table 1.1 is ~ 0.5Jy. This alone marks an important change from past mm VLBI experiments. In the past, all mm-VLBI experiments chose targets from virtually the same source list not because other objects were uninteresting but because only the brightest and most compact sources could be detected. With new averaging techniques now available it is hoped that a broader spectrum of sources may be investigated.

There is still a strict brightness temperature requirement for 3mm-VLBI detection due to the small beam size of an array. If we take 0.5Jy as our flux density threshold and place it within the beam of a two element interferometer then the minimum brightness temperature can be expressed as

$$T_{B,min} = \frac{2}{\pi k} B^2 S_{min} \tag{1.2}$$

which is $2 \times 10^{10} K$ when evaluated for the longest (10^4km) baselines.

1.3 Thesis Outline

The first goal of this project was to develop mm-VLBI capabilities in areas of both hardware and software. As a starting point, Chapter 2 reviews basic relations of interferometry to aid in later discussion. Chapter 3 details the design, construction, and testing of a polarizing window for the Haystack antenna. Haystack was a new addition to the 3mm-VLBI network and required a quarter wave plate that would allow it to record in LCP mode at 86GHz. Chapter 4 begins with a review of VLBI detection methods and then describes the new incoherent methods developed to lower detection thresholds for low coherence data. Application to mm-VLBI data is explained and methods of recovering phase information are discussed.

The second part of the thesis was the scheduling and observation of a global 3mm-VLBI experiment. Chapter 5 gives the particulars of antennas, sources, and scheduling concerns. It also describes the data calibration procedure and the data path through the methods of Chapter 4. The last two chapters are analyses of two specific sources observed at 3mm. Chapter 6 focuses on the source 3C111 which was included in the observations only as a test

source. It is a nearby radio galaxy and with its widely spaced radio lobes is atypical of the core dominated sources that make up most of the source list. We examine the possibility that 3mm observations may be able to set constraints on its orientation, investigate physical parameters of the core, and find an improved position using VLBI observables. Chapter 7 describes observations of the galactic center source SgrA*, thought to be an AGN in our own galaxy. A reanalysis of data taken in 1994 is performed and a new size limit for this compact object is found which is consistent with the scattering size for a point source in the ISM. Two different models of emission are discussed and one is shown to be inconsistent with observations.

Chapter 2

Interferometry

A full review of interferometry is beyond the scope of this thesis but introduction to select terms and ideas will be necessary for the chapters to follow. Accordingly, a brief outline is given here with attention paid to those areas particularly important for mm-wavelengths.

2.1 Preliminaries

A basic interferometer consists of two receiving elements (antennas) that track a source on the sky. The signals from each are multiplied and averaged in a device called a correlator. Buried in this correlation is information about the source structure and the art of interferometry is to collect, calibrate and transform this data into a map of sky brightness. The idealized picture of an experimental setup (Fig. 2-1) shows the two antennas, the wavefront from a distant source, the time delay between signal reception at both sites, and the connecting baseline.

We will examine the content of the box labeled "Correlator" later, for now assume that we have access to its output. The delay term τ_g can be written in terms of source and station coordinates :

$$\tau_g = X_\lambda \cos(\delta)\cos(H) - Y_\lambda \cos(\delta)\sin(H) + Z_\lambda \sin(\delta) \tag{2.1}$$

where (H, δ)=source hour angle (from Greenwich) and declination and $(X_\lambda, Y_\lambda, Z_\lambda)$ are the baseline coordinates in the directions $(H = 0, \delta = 0), (H = -6^h, \delta = 0), (\delta = 90°)$. As the Earth rotates, τ_g will change as will the projected baseline d_p. The quantity $\omega \tau_g$ where ω is

19

the observing frequency is the fringe phase and the magnitude of the correlation coefficient is the visibility magnitude.

The value of $\omega \tau_g$ is very sensitive to the specific geometry of an observation. Its rate of change, the fringe frequency :

$$\nu_f = \frac{\omega}{2\pi} \frac{d\tau_g}{dt} = -\frac{\omega_e}{2\pi} \left[X_\lambda \cos(\delta) \sin(H) + Y_\lambda \cos(\delta) \cos(H) \right] \qquad (2.2)$$

(ω_e is Earth's angular rotation) can cause the phase to rotate at frequencies higher than 10kHz. For the longest millimeter baselines of $2 \times 10^9 \lambda$ the fringe rate has a maximum of $\sim$ 25kHz. This rapid variation at the correlator ouput is slowed to a near zero rate by putting a compensating delay (τ_g' shown in Fig. 2-1) into the signal path from antenna 2. A given τ_g' removes fringe phase and rate for exactly one point on the sky - the phase center. For other points in the interferometer's field of view, this correction is not exact and leaves a net phase offset and rate. The phase response due to a point source is, therefore, always referenced to the zero phase at the phase center.

From Fig. 2-1 we can write the correlator output as an averaged product of two monochromatic radio signals. For a point source the normalized correlator output is the complex visibility

$$\mathcal{V}_n = \exp\left\{ -2\pi j (u\xi + v\eta) \right\} \qquad (2.3)$$

where the baseline vector has been projected onto the sky plane in the direction of the source with components (u, v) measured in λ and the (ξ, η) are displacements on the sky from the origin of the image plane. An expression for the correlator output for a complex source is the integral of the point source response over the brightness distribution on the sky

$$\mathcal{V}(u, v) = \int \int B(\xi, \eta) \, e^{-2\pi j (u\xi + v\eta)} \, d\xi \, d\eta \qquad (2.4)$$

This key relation ties the output of the correlator to the sky brightness distribution via a fourier transform relation. The visibility is then a fourier component of the sky brightness distribution specified by a particular baseline length and orientation. We note that since $B(\xi, \eta)$ is a real valued function, $\mathcal{V}_n(u, v)$ will be conjugate symmetric so that $\mathcal{V}_n(u, v) = \mathcal{V}_n^*(-u, -v)$ and only one half of the (u, v) plane need be sampled in order to invert the fourier expression and recover the image. If the Earth did not rotate, each baseline would

contribute only one point to the visibility data. The Earth's motion increases the visibility coverage by changing the projected baselines as they appear to someone on the source looking back at Earth. Each baseline traces out an ellipse in (u, v) space whose eccentricity and dimensions depend on antenna positions and source declination. In practice, coverage of the (u, v) plane is sparse due to the limited number of antennas and the fourier inversion process is not uniquely determined. Fig. 2-2 shows the coverage for the source 3C84 using 4 antennas.

To further complicate matters, in VLBI the phase of $\mathcal{V}(u, v)$ is not directly recoverable at the correlator. This is due to the fact that separate time standards run each station and have small clock and rate offsets that cause offsets and drift in the visibility phase. The delay τ_g is also affected by the unknown phase delay through the atmosphere to each antenna. Both these effects render the measured visibility phase essentially random. Without a proper phase for each visibility, any map made by direct fourier transform of Eq. 2.4 will have little relation to the true sky image. The solution to this problem (one that plagued VLBI for many years) lies in forming the so called closure phase (Rogers et al,1974;Jennison,1958).

Though the baseline phases are corrupted by the factors mentioned above, these effects can nearly always be distributed on a station basis. So, for example, each station has associated with it an atmospheric delay, a clock rate, and a clock delay. If these assignments are made then summing the measured phase around a closed triangle of baselines will cause these factors to cancel and leave us with the sum of true visibility phases. Appendix D shows the mechanics of this sum and the cancelling terms. The resulting phase is known as the closure phase and contains only information about the source structure. Use of the closure phase in creating an image has been a subject of some debate over the past two decades. On a singe triangle of baselines, the closure phase only allows determination of the sum of all three baseline phases. With three unknowns and only one relation containing them, one cannot uniquely determine all three phases. In general, the number of independent closure phases for an N station array, $(N - 1)(N - 2)/2$, is always less than the required $N(N - 1)/2$ baseline phases. The remaining $N - 1$ phases are either determined iteratively by first assuming a source model or are left to vary with the closure phases being used in a 2-D least squares fit.

2.2 Field of View

The monochromatic interferometer response does not tell the whole story. Visibility is not determined for a single time or at a single frequency but as an average over a frequency range and for a finite time. This averaging smears the final image and also limits the instrument's field of view.

2.2.1 Finite Bandwidth

Interferometer sensitivity is proportional to the square root of system bandwidth so it is to the observer's advantage to record signal over the widest range of frequencies possible. The visibility function, however, responds differently to each frequency. If we make the frequency dependence of the normalized visibility function explicit, we can write for a point source :

$$\mathcal{V}_n = \exp\left\{-2\pi j \frac{\nu}{c}(u'\xi + v'\eta)\right\} \qquad (2.5)$$

where (u', v') are baseline components in meters, (ξ, η) are source displacement coordinates on the sky. Integrating this response over a bandwidth centered on ν_o we find that the visibility will have the following bandwidth dependence:

$$\mathcal{V}_n \propto \text{sinc}\left(\pi\frac{\Delta\nu}{\nu_o}(u\xi + v\eta)\right) \qquad (2.6)$$

So, moving the source off the phase center of the array incurs a response loss. How far the source can move is dictated by the fractional bandwidth inside the sinc expression. We try to maintain :

$$\xi < \frac{\nu_o}{u\Delta\nu} \quad \text{and} \quad \eta < \frac{\nu_o}{v\Delta\nu} \qquad (2.7)$$

A limiting case for mm-VLBI would be $\Delta\nu = 112\text{MHz}$, $\nu = 86\text{GHz}$ and u,v $= 1\text{G}\lambda$. This case limits the field of view on the sky to be $\xi, \eta < 150\text{mas}$. This is known as the delay beam size and is not the primary limitation on mm-VLBI observing : the high brightness mm structure that mm-VLBI is sensitive to rarely extends more than 10mas from the core. The delay beam size does not change with observing frequency since ν_o and (u, v) scale together. As long as bandwidth remains constant, the delay beam stays the same.

2.2.2 Visibility Averaging

The strength of aperture synthesis lies in its ability to greatly increase sampling in the u,v plane by observing as the Earth's spin changes the direction and projected length of the baselines. As an observation progresses, one can monitor the evolution of each baseline by watching its time evolving track on the u,v plane. Since source structure can be represented as texture on the u,v plane, the visibility amplitude recorded on a baseline will change with time. This implies a natural limit to how long one may average the signal from an interferometer. If, for example, the source structure causes very rapid beating in the interferometer response, then too long an average in the correlator or during post-processing will wash out these variations and our sensitivity to structure. For a simple double source, the frequency of beating in the visibility amplitude is proportional to component separation on the sky : the farther apart the components, the faster the visibility will oscillate in time. Visibility averaging is, therefore, usually taken as a field of view limitation for an array. For mm-VLBI this effect dominates all others contributing to the field of view. In the case of very weak sources for which one would like to integrate as long as possible, the limitation can be very strict.

The normalized visibility for a point source displaced from the phase reference center by (ξ, η) is given in Eq. 2.3. The exact frequency of oscillations in visibility will vary based on the geometry of the particular situation. r.a. and declination of the source, baseline vectors, and time of observation all contribute to this rate. Writing u,v in terms of time and expanding the visibility about time t_o we can express visibility as :

$$\mathcal{V}_n = \mathcal{V}_{n_o} \exp\left[-2\pi j \tau \omega_e \left(\xi(\cos(H_o)X_\lambda - \sin(H_o)Y_\lambda)\right) + (\eta \sin(\delta)(\cos(H_o)Y_\lambda + \sin(H_o)X_\lambda))\right]$$

(2.8)

where $\mathcal{V}_{n_o}$ is the visibility at t_o, H_o is source hour angle at t_o, τ is the offset in time from t_o and (X_λ, Y_λ) are baseline components in units of wavelength. Integrating this form over a time window $(t_o - \tau/2, t_o + \tau/2)$ and arbitrarily setting $X_\lambda = 0$ we find a reduction factor of :

$$R_{av} = \text{sinc}\left\{\pi Y_\lambda \omega_e \tau[-\xi \sin(H_o) + \eta \sin(\delta) \cos(H_o)]\right\}$$

(2.9)

The maximum frequency of oscillations will occur when $\tan(H_o) = -\xi/\eta \sin(\delta)$ and we can

23

rewrite the reduction as :

$$R_{\mathrm{av,max}} = \mathrm{sinc}\left\{\pi Y_\lambda \omega_e \tau \sqrt{\xi^2 + \eta^2 \sin^2(\delta)}\right\} \tag{2.10}$$

The argument of the sinc function must be much less than one for the edges of the field not to fade so we write:

$$d < (\pi \omega_e \tau)^{-1} \tag{2.11}$$

where d is the source offset from the image plane origin measured in beamwidths. A typical averaging time in mm-VLBI is $\tau = 390$ seconds which means that the effective imaging field may be only ~ 30 beamwidths wide. This can limit the field of view to 15mas or less depending on the specific array. There are sources that we predict will have only a single component near the phase center such as SgrA*. For these, the field of view limit is not as important and one can safely average much longer.

2.3   Signal Recording

Due to the separate and isolated nature of each VLBI antenna, signals must be recorded onto some storage media for later playback at a central location. The path from source to correlator is long and tortuous so much of current receiving systems are devoted to isolating and correcting for sources of loss and phase corruption. Little can be done about the effects of intergalactic and interstellar media. However, added phase and signal strength loss incurred in our atmosphere can be removed to some degree by measurements made during the VLBI observing run. The quasi-reconstruction of phase through the use of closure was touched upon in the last section and the atmospheric loss corrections will be discussed in a later section on calibration. Suffice it to say that we are able to approximately recreate the signal as it appeared just before entering the Earth's atmosphere.

When it reaches the antenna, the dish collects the signal then directs it through a circular polarizing window (see next chapter) and into the receiver feed. From there, it encounters a number of mixing stages which shift the desired observing band down to some intermediate frequency (IF). Fig. 2-3 shows an example of such a local oscillator chain that converts sky frequency to the intermediate band at Haystack Observatory. This IF is then sampled by up to 14 base band converters (BBC) that make up the analog acquisition portion of the

24

Mark III VLBI system. Each of the BBCs can be set for up to 8MHz bandwidth making the maximum bandpass of the system 112MHz. The function of each BBC is to convert a different section of the IF band to baseband : 0 to 8MHz where in this case 8MHz is the preset BBC bandwidth. Since every BBC generates its own mixing frequency, phase offsets from one to another can show up during the down-conversion. To keep track of this a $1\mu s$ period pulse train is injected into the IF and shows up in the frequency spectrum as a spike at every multiple of 1MHz. By following the phase of this "phase-cal" signal, the variations between the BBC can be found and removed at the correlator. This step is necessary in any multi-band acquisition system to align the bands in phase for averaging.

The Mark III formatter samples these video freqency channels at the nyquist rate with 1-bit sampling. This feature ensures that the sign of the signal is kept but all amplitude information is lost. Deriving the true correlation from this clipped version makes use of the Van Vleck relation which states:

$$R_{12}(\tau) = \sqrt{T_{s1}T_{s2}} \sin\left(\frac{\pi}{2}\rho_{12}^c(\tau)\right) \simeq \sqrt{T_{s1}T_{s2}} \left(\frac{\pi}{2}\rho_{12}^c(\tau)\right) \qquad (2.12)$$

where R_{12} is the desired correlation coefficient, ρ_{12}^c is the normalized 1-bit clipped correlation, and the T_s are system temperatures. The T_s information cannot be recovered directly from the correlation and is determined separately during antenna calibration. High T_s and low antenna temperature of sources keep ρ_{12}^c very small (typically a few $\times 10^{-4}$) and sin can be replaced by its argument.

The formatter encodes station time along with the baseband signals and writes the ensemble to magnetic tape. Station time stamps allow the correlator to synchronize the tapes during playback at a later time.

2.4   Correlation

In theory we can follow the signal from antenna to correlator to see the dependence of phase on geometric and instrumental effects. The phase between signal from antenna 1 and 2 can be written:

$$\phi_{12} = 2\pi(\nu - \nu_{lo})(\tau_2 - \tau_1) + (\theta_2 - \theta_1) + 2\pi(\nu\Delta\tau_g + \nu_{lo}\tau_g') \qquad (2.13)$$

with ν=sky frequency, ν_{lo}=local oscillator frequency, τ_1 and τ_2 are clock errors at each station, θ_1 and θ_2 are phase offsets in frequency references, $\Delta\tau_g = \tau_g - \tau'_g$ is the difference between actual geometric delay and geometric delay applied at the correlator. With each antenna's characteristic response $A_i(\omega)$, we can write down the cross spectrum: $S_{12}(\omega) = S_1(\omega)S_2^*(\omega) = |A_1(\omega)||A_2(\omega)|\exp(j\phi_{12})$. The cross correlation is the fourier transform of the cross spectrum:

$$R_{12}(\tau) = \int_{-\infty}^{\infty} S_{12}(\omega)e^{j\omega\tau}\frac{d\omega}{2\pi} \tag{2.14}$$

which, for rectangular passbands at each antenna is

$$R_{12}(\tau) = 2\Delta\nu\cos(2\pi\nu_{lo}\tau_g + \theta_{21} + \pi\Delta\nu(\tau + \Delta\tau_g + \tau_2 - \tau_1))\operatorname{sinc}(\pi\Delta\nu(\tau + \Delta\tau_g + \tau_2 - \tau_1)) \tag{2.15}$$

The remaining τ_g in the cos term changes with time and gives rise to the fringe frequency discussed earlier. This effect is taken out by phase shifting one signal path by $2\pi\nu_{lo}\tau'_g$ and the correlation becomes

$$R_{12}(\tau) = \Delta\nu\exp[j(2\pi\nu_{lo}\Delta\tau_g + \theta_{21} + \pi\Delta\nu(\tau + \Delta\tau_g + \tau_2 - \tau_1))]\operatorname{sinc}(\pi\Delta\nu(\tau + \Delta\tau_g + \tau_2 - \tau_1)) \tag{2.16}$$

and the visibility is found by maximizing $R_{12}(\tau)$ in a search over lag space. In the frequency domain one writes the cross spectrum:

$$S_{12}(\nu') = |A_1(\nu')||A_2(\nu')|\exp j[2\pi\nu'(\tau_2 - \tau_1 + \Delta\tau_g) + 2\pi\nu_{lo}\Delta\tau_g + \theta_{21}] \tag{2.17}$$

where $\nu' = \nu - \nu_{lo}$. The delay that is measured from the cross correlation is the group delay $(1/2\pi)(d\phi/d\nu')$ across the bandpass. This tells us that the MarkIII system, which comprises many frequency channels, will need one more parameter to recover visibility. Knowledge of fringe rate allows us to average the complex correlation over time, and with measurement of phase across a single channel we can average in frequency across the band. To average over the entire system bandwidth (up to 28 channels) the group delay over all ~ 112MHz is needed. This is done by maximizing a multiband delay function which is the sum of correlation functions from each individual channel. Therefore, to make sense of the correlations in time and across all frequency bands, the rate, single band delay (SBD) and multi-band delay (MBD) must be found. The process of searching for a correlation

26

maximum in this 3-D space is called the fringe search and is discussed in detail in the chapter on detection.

To source

$$\tau_g = \frac{d \cos\theta}{c}$$

$$d_p = d \sin\theta$$

τ_g

d_p

θ

$V_1(t)$

d

Correlator

τ_g'

Delay

$V_2(t + \tau_g)$

$< V_1(t) \; V_2(t) >$
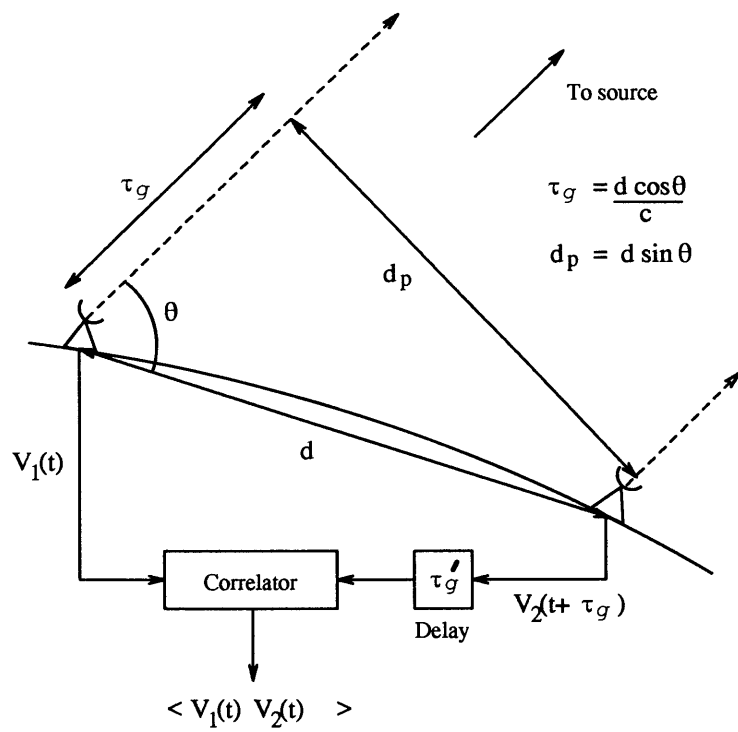
Figure 2-1: Schematic view of simplified interferometer.
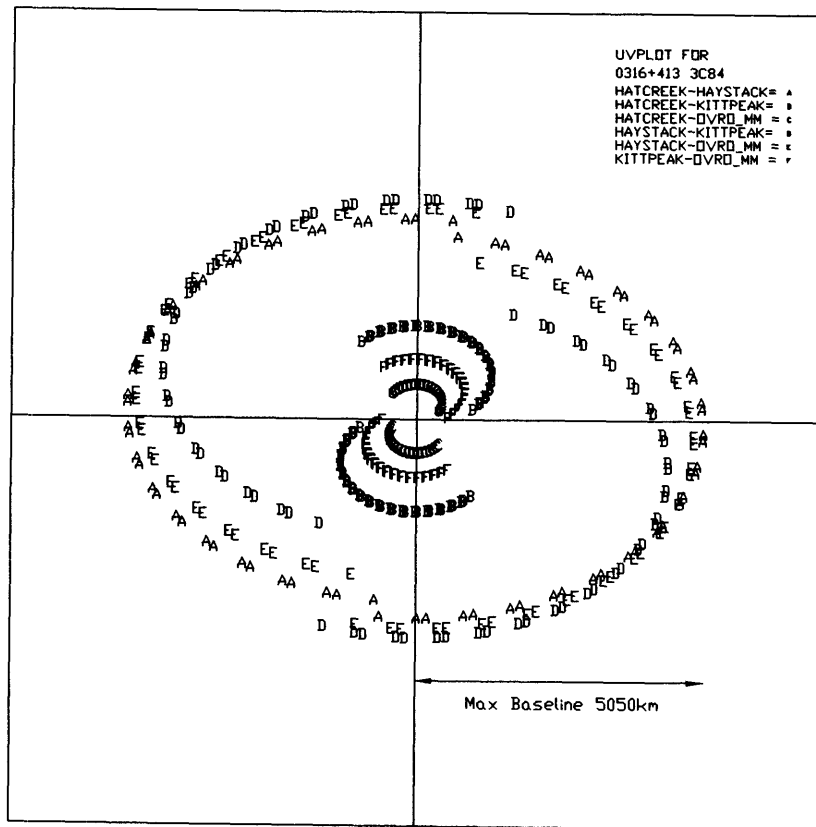
Figure 2-2: u,v coverage for a 4-station observation of 3C84 over 12 hours.
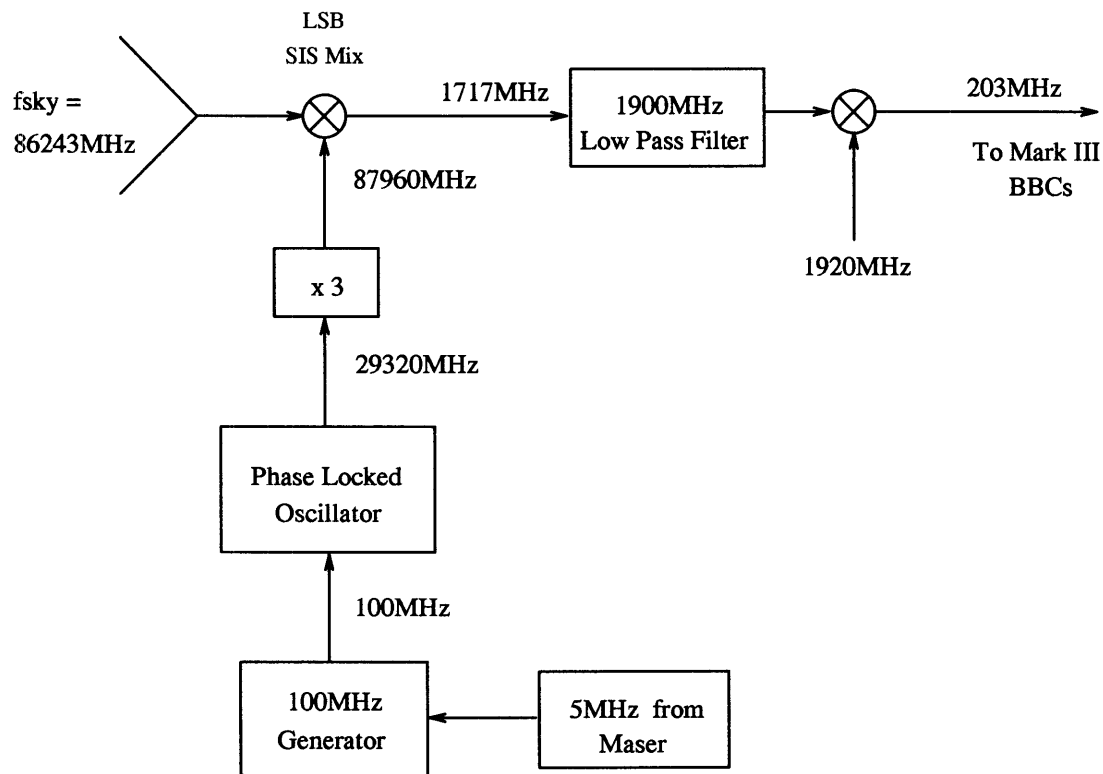
Figure 2-3: Schematic of Haystack Local Oscillator chain for 86GHz VLBI. 5MHz is derived from the station's hydrogen maser and is used to drive two phase locked oscillators, one at 100MHz and another at 29320MHz. After a final multiplication by 3, the first local oscillator is combined with signal from the sky in an SIS mixer. A 1900MHz low pass filter provides image rejection for a 2nd mixer with an LO of 1920MHz. The resulting frequency band centered at 203MHz is sampled by the MarkIII BBCs.

Chapter 3

Polarizer for Haystack Observatory

Since the Haystack antenna was a new addition to the 3mm network, it required, among other things, a quarter wave plate for 86GHz operation. This plate, also called a polarizer, converts linearly polarized radiation to circular polarization and visa-versa. The need for this becomes clear considering that each antenna observes the same source, but have their feeds oriented at different angles. So the orientation of the source with respect to each antenna's receiver feed will vary in a determinable way.

For an alt-az mounted telescope, the parallactic angle is given by

$$\sin(\psi_p) = \frac{\cos(L) \ \sin(H)}{\cos(E)} \tag{3.1}$$

with L, H and E being station latitude, source hour angle and station elevation respectively. This angle is shown in Fig. 3-1 for the Haystack Antenna tracking the source 3C84 which is at a declination of 41.5 degrees. For the same time range, this angle is also shown for the Onsala telescope in Sweden. Since both have linear feeds, each of these telescopes is sensitive to a different polarization from the source at any given time. For an unpolarized source, the resulting correlation depends on the angle between the feeds and will go to zero as that angle approached 90 degrees (cross-polarization).

The general fix for this situation in VLBI has been to mount quarter-wave windows in front of the receiver. This has the effect of converting a linear feed to one sensitive to LCP or RCP radiation. With all antennae observing either RCP or LCP, no possibility for cross
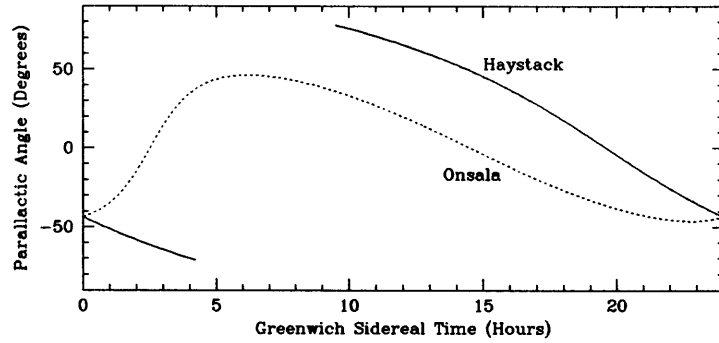
Figure 3-1: Plot of parallactic angle for Haystack and Onsala antennas tracking the source 3C84. The plot for Haystack is not shown for times when the source is not visible at that site. For Onsala, the source is circumpolar.

polarization exists. Though this technique solves the polarization question, it introduces another source of loss in the signal path and another point of possible reflection that can add noise to the system. Schematically, we show the polarizer in its position before the feed in Fig. 3-2.
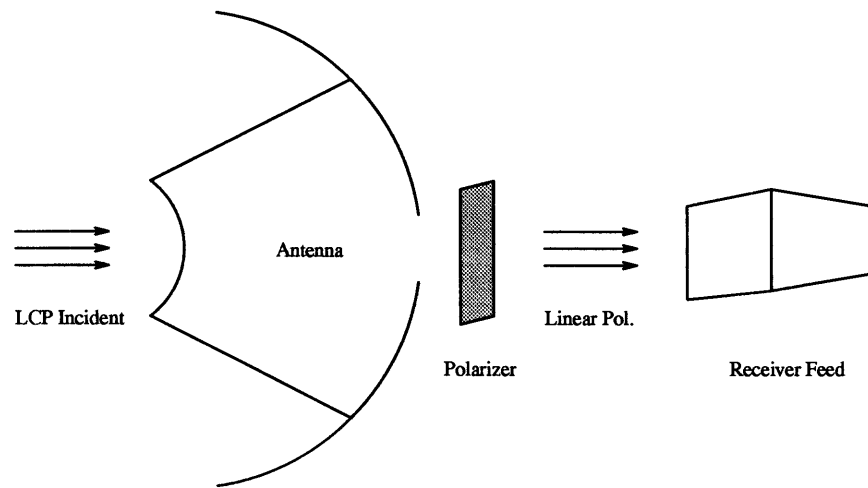
Figure 3-2: The polarizer is positioned just before the feed and may introduce reflections, loss, or add system noise if not designed properly.

Should the polarizer exhibit undue loss, it will contribute to the overall noise of the

system as follows (τ = attenuation) :

$$T_{noise} = T_{ambient}(1 - e^{-\tau})$$ (3.2)

and it will reduce the signal coming from the subreflector :

$$T'_{signal} = T_{signal}e^{-\tau}$$ (3.3)

The polarizer then is required to keep polarizing efficiency high and additive loss low. We examined two different approaches to this issue both of which will be described here. The final design which took the form of a window of stacked waveguides will be shown in more detail. Later tests of the polarizer showed scattering loss to be a factor but at only a 5% level.

3.1 Grooved Dielectric Design

Any polarizer design must incorporate a medium that presents different propagation vectors to incident light based on its polarization. One way of doing this is to cut grooves in a slab of dielectric material so that light with an E vector aligned with the grooves sees a different effective dielectric constant than light with E field oriented perpendicular to the grooves. By adjusting the effective dielectric constants for each direction, the linear components of CP light can be brought into phase and added while propagating through the polarizer. If we groove the dielectric as shown in Fig. 3-3 (the central ungrooved portion does not effect the differential phase shifts) we can calculate the perpendicular and parallel dielectric coefficients to be:

$$\epsilon_\perp = \frac{\epsilon(a + b)}{b\epsilon + a}$$ (3.4)

$$\epsilon_\parallel = \left(\frac{a}{a + b}\right)(\epsilon - 1) + 1$$ (3.5)

where ϵ is the dielectric constant of the ungrooved medium (usually telflon: $\epsilon \simeq 2.7$). These relations can be easily derived by enforcing boundary conditions at the dielectric surface and averaging the E-field over the volume of the grooved medium. They are also given in Born & Wolf (p.705) and in (Kirschbaum and Chen, 1957).

33

To ensure a 90° phase difference between $\perp$ and $\parallel$ orientations we require that :

$$\frac{\lambda}{4} = 2L(\sqrt{\epsilon_\perp} + \sqrt{\epsilon_\parallel}). \tag{3.6}$$

Given that a model obeys Eq. 3.6, we can then compute the expected first order reflections from each of the 4 interfaces of the dielectric. Fig. 3-4 shows the normalized reflected intensity from a sample polarizer tuned to a value near 86GHz.

The Quabbin antenna in Massachusetts has used a dielectric polarizer at mm-wavelengths for many years. They report losses of over 10% loss in signal with the polarizer installed. This is probably due to a number of factors including tangent losses in the teflon and uneven machining of the high aspect ratio grooves. In addition, the extremely narrow grooves ($\sim$ 0.5mm) are prone to capture condensation and dirt. Haystack, therefore, opted to design and build a plate using a different mechanism to achieve a window with different propagation constants along each axis.

3.2 Parallel Plate Polarizer

3.2.1 Simple Treatment

An alternative method of creating a birefringent material is to stack parallel plate waveguides together to form a window with two different indicies of refraction. The key to a simple treatment of this polarizer is assuming that we can view the metal vane medium as a homogenous dielectric. We ignore any problems that arise due to fringing fields near the vane edges and idealize the situation as the section of stacked parallel plate waveguides shown in Fig. 3-5.

Adopting this approach, we can write the electric fields in the polarizer with the following z dependence :

$$E_\parallel \propto e^{ik_\parallel z} \quad \text{with} \quad k_\parallel = \frac{2\pi}{\lambda}\sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} \tag{3.7}$$

$$E_\perp \propto e^{ik_\perp z} \quad \text{with} \quad k_\perp = \frac{2\pi}{\lambda} \tag{3.8}$$

In other words, the component of incident plane wave polarized in the x direction will pass through the polarizer unaffected. It already satisfies all boundary conditions imposed by the metal plates. The y component, however, will see a waveguide with cutoff wavelength

$\lambda_c = 2b$ and effective index of refraction :

$$n_{\parallel} = \sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} \qquad (3.9)$$

The net effect will be to change the relative phase of these orthogonal components which is exactly what we need to build our quarter-wave plate.

A Fabry-Perot type analysis of transmission through the polarizer yields two relations which, when combined, specify the required polarizer depth and vane spacing. Viewing the polarizer as a dielectric slab, we can derive (or find in Born & Wolf p.325) the intensity of the reflected wave:

$$\text{Reflection Intensity} = \frac{4\rho^2 \sin^2\left(\frac{\delta}{2}\right)}{(1 - \rho^2)^2 + 4\rho^2 \sin^2\left(\frac{\delta}{2}\right)} \qquad (3.10)$$

$$\text{where} \qquad \delta = \frac{4\pi d n_{\parallel}}{\lambda} \qquad \text{and} \qquad \rho = \left(\frac{1 - n_{\parallel}}{1 + n_{\parallel}}\right)$$

So, to completely cancel all reflective intensity, we require that $\delta = 2m\pi$ where m is an integer; or, equivalently :

$$d = \frac{m\lambda}{2n_{\parallel}}. \qquad (3.11)$$

From the same analysis, we can find the phase difference between the two polarizations of E field and require that it be 90 degrees (for circular polarization) to get the second relation :

$$\frac{\pi}{2} = \frac{2\pi d}{\lambda}(1 - n_{\parallel}) - \arctan\left(\frac{\rho^2 \sin(2\delta)}{1 - \rho^2 \cos(2\delta)}\right) \qquad (3.12)$$

Ultimately, if these two equations are satisfied, our polarizer will advance the phase of $E_{\parallel}$ by 90 degrees and transmit all the energy of the incident wave. When we combine Eqs. 3.11 and 3.12, we find that there is a family of polarizers for a given wavelength :

$$b_m = \frac{\lambda/2}{\sqrt{1 - \left(\frac{2m}{2m+1}\right)^2}} \qquad (3.13)$$

$$d_m = \frac{(m\lambda/2)}{\sqrt{1 - \left(\frac{\lambda}{2b_m}\right)}} \qquad (3.14)$$

This prescription guarantees us a reflectionless quarter-wave plate at the wavelength λ for each integral value of m. But although it appears that any value of m will do, we have already made an implicit assumption that severely limits the range of m. The general expression for the parallel propagation constant is :

$$k_{\parallel,n} = \frac{2\pi}{\lambda}\sqrt{1 - \left(\frac{n\lambda}{2b}\right)^2} \tag{3.15}$$

which means that our value of $k_\parallel$ in Eq. 3.7 corresponds to the lowest order TE mode in the waveguide. Including higher order terms makes the analysis much more complex so we restrict ourselves to considering only the TE_{10} mode ($n = 1$). To this end we design the waveguide such that all modes with $n > 1$ have pure imaginary propagation constants and are exponentially damped. Specifically we want :

$$\sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} \quad \text{to be real}$$

$$\text{and} \quad \sqrt{1 - \left(\frac{n\lambda}{2b}\right)^2} \quad \text{to be imaginary if } n > 1.$$

This simplification introduces the restriction $1/2 < b/\lambda < 1$ which forces a limit on how high m can go. The table below shows the acceptable m values for $\lambda = 86\text{GHz}$ and the resulting polarizer dimensions. Theoretical results of the three possible designs are also shown in Fig. 3-6.

m	b_m (mm)	d_m (mm)
1	2.34	2.62
2	2.91	4.36
3	3.39	6.11

3.2.2 Complex Treatment

A more exact look at polarizer performance means addressing the non homogeneity of the metal vane media. Carlson and Heins (1947) (CH) were the first to rigorously solve a related problem : the reflection and phase shift of a plane wave incident on an infinite half plane of spaced metal plates. Lengyel (1951) extended their results to the case of a slab of metal vane media and established a credible agreement with experiment. It should be noted that the final analytical results are expressed in terms of infinite sums, and their evaluation in

1951 was not easy. The increased speed and efficiency of modern computers allowed us to evaluate the necessary sums and search a large parameter space to find an optimal design.

In this section we outline the method of CH, give Lengyel's results and numerically solve for the E-M fields present in the CH infinite half-plane problem. Calculation of the fields will enable us to estimate additional losses due to the increased magnetic fields near the vane edges.

CH formulate the problem as a contour integral over the vane surfaces using Green's theorem. When they impose all boundary conditions, they are left with an integral equation which they have to solve in order to get the surface currents on each vane. Their next step is to fourier transform the integral (which leads to an easier equation) and solve for the current. The final result for the E_y field is in the form of a complex contour integral which is evaluated by looking at the singularities of the integrand. Physically, each of the singularities corresponds to a TE mode in the waveguide. So, as one might expect, the solution for $E_y(x, z)$ will be a sum of waveguide modes weighted by coefficients determined by C+H's contour integral. For the case of $f = 86\text{GHz}$, $a = 3.2\text{mm}$, $d = 5.25\text{mm}$ the E-M fields just inside the metal vanes are shown in Fig. 3-7 .

Lengyel(1951) assumes a lossless interface to show how the phase of a wave transmitted through the polarizer is related to the phase shift in the CH paper. He shows that for a polarizer surrounded by air,

$$\rho' + \rho'' - 2\tau' = \pm\pi \tag{3.16}$$

where

$$\rho' \quad = \quad \text{phase of reflection from front surface of polarizer}$$

$$\rho'' \quad = \quad \text{phase of reflection from back surface of polarizer}$$

$$\tau' \quad = \quad \text{phase of transmission across air - metal vane interface}$$

He uses this relation and the standard reflection/transmission relations for a slab with air on both sides to write :

$$|\text{Reflection Amp.}| = \frac{2\rho|\sin\Psi|}{\sqrt{(1-\rho^2)^2 + 4\rho^2\sin^2\Psi}} \tag{3.17}$$

37

$$\text{phase shift} = \rho' + \Psi' + \pi - \frac{2\pi d n_{\parallel}}{\lambda} \tag{3.18}$$

where

$$\begin{aligned}
\Psi &= \rho'' + \frac{2\pi d}{\lambda} \\
\Psi' &= \arctan\left\{\left(\frac{1+\rho^2}{1-\rho^2}\right)\tan\Psi\right\}
\end{aligned}$$

and the expressions for ρ' and ρ'' are

$$\rho'(x) = 2\left\{x(\ln 2 - 1) - \sum_{n=2}^{\infty}(-1)^n\left[\arcsin\left(\frac{x}{n}\right) - \frac{x}{n}\right]\right\} \tag{3.19}$$

$$\rho''(y) = -2\left\{y(\ln 2 - 1) - \sum_{n=2}^{\infty}(-1)^n\left[\arcsin\left(\frac{y}{\sqrt{n^2-1}}\right) - \frac{y}{n}\right]\right\} + \pi \tag{3.20}$$

where

$$x = \frac{2b}{\lambda} \quad \text{and} \quad y = \frac{2b n_{\parallel}}{\lambda} \tag{3.21}$$

Fig. 3-9 shows graphs comparing Lengyel's predictions with those of the simple approach.

3.2.3 Experiment

In order to asses how well the theory predicted polarizer behavior, we set up a test fixture to measure transmission. Our main goals were to take measurements with the incident E field vector perpendicular, parallel, and also at 45 degrees to the vanes. One expects the first two orientations to pass nearly all incident signal and the last to cut the signal intensity by a factor of 2. This is most easily seen by considering that at 86 GHz the $1/4$-λ plate will convert the incident linear polarization to CP. When this transmitted wave strikes the linear receiving feed, only one half of the original intensity will be recovered. The setup is shown in Fig. 3-8.

The aperture of each horn was reduced using echo-sorb plugs to ensure that all path lengths from oscillator to bolometer differed by no more than $\lambda/3$. This was done to reduce the problem of phase front curvature. Another problem involving reflections along the signal path was not so easily dealt with. It turned out that moving the horns together or apart caused the bolometer output to oscillate. This was interpreted as fluctuations in signal

frequency combining with reflections to cause destructive interference at the detector. An effort was made using a frequency sweep mode of the BWO to average over these unwanted oscillations but results were not as good as expected.

Fig. 3-10 shows transmission for the three orientations. We also graph, for comparison, the expected dependence of 45 degree transmission on phase shift over this frequency range. Each point is an average of a few readings and, based on the variance of these readings, the points shown are good to within 10%.

3.2.4 Losses

As it turns out, the largest contribution to the loss will probably come from the phase error introduced by variations in vane spacing. The prototype polarizer was designed for a spacing of 3.2mm and, when measured, showed an average spacing ($\bar{b}$) of 3.28mm with a standard deviation (σ) of 0.17mm. The polarizer frame was then re-tensioned in a vise to reduce any buckling. When it was re-assembled these numbers changed to $\bar{b} = 3.2$mm and $\sigma = 0.12$mm. To get an idea of how this variation would effect loss, we expressed loss as a phase integral making the vane spacing follow a normal distribution with a specified average and variance. As an approximation, the simple phase model was used. The expression for transmission is :

$$
\text{Transmission} = \left| \frac{\int \exp\left(-x^2/2\sigma^2\right) \exp \frac{2\pi i d}{\lambda} \left(\sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} - \sqrt{1 - \left(\frac{\lambda}{2(b-x)}\right)^2} \right) dx}{\int \exp\left(-x^2/2\sigma^2\right) dx} \right|^2
$$

(3.22)

For a $\sigma = 0.15$mm we find a loss of 2.6% and an increase of σ to 0.3mm pushes the loss to 5.7%. A new polarizer design using spring tensioned vanes was built and reduced this source of loss considerably.

As with any waveguide component, we have to consider ohmic losses due to the finite conductivity of the guiding material. The loss in parallel plate waveguide for the dominant mode is given by :

$$
\text{loss} = e^{-2\beta z}
$$

(3.23)

$$\beta = \frac{1}{b}\sqrt{\frac{c}{2\sigma b}}\frac{(\lambda/2b)^{\frac{3}{2}}}{\sqrt{1-\left(\frac{\lambda}{2b}\right)^2}} \tag{3.24}$$

with σ the conductivity. This amounts to a loss of approximately 0.15% in the polarizer. But we also have to note that the metal vanes distort the normal waveguide fields near the vane edges causing additional loss. In general, the time averaged power absorbed per unit area is roughly :

$$\frac{dP_{\text{loss}}}{da} = \frac{\mu\omega\delta}{4}\left|H_{\parallel}\right|^2 \tag{3.25}$$

where $H_{\parallel}$ is the tangential magnetic field at the vane surface and δ is the skin depth. Using worst case values of the surface H field yeilds an additional loss of 0.05%. If we double the theoretical loss to account for surface roughness we get a total ohmic loss of $\sim$0.5%. From these considerations, ohmic losses should account for no more that 1-2% of the overall loss.

Thickness of the vanes does play a role in transmission but Lengyel's 1951 paper suggests that the effects are not noticeable until the vanes are quite thick. The analytical approach doesn't easily accomodate vane thickness as a parameter but Rogers & Doeleman (1992) have numerically solved for polarizer transmission and include this parameter in the calculations. The result is that our vane thickness of 0.003 inches is a minor loss issue.

3.2.5 Mechanical Design

Initial tests showed that very thin brass shim could be strung within a frame to form the waveguide window. Using this idea we bolted together a frame of thick brass stock and cut slits in two sides to guide the brass shim. Each slit was located with required spacing for the waveguide and made wide enough to accomodate the shim with sufficient epoxy to secure it. This model was tested both in a laboratory setup and in operation on the Haystack antenna with details of both reported in the sections below.

Lab tests showed the polarizer to work but antenna tests revealed that it was scattering too much of the incident radiation. We determined that uneven spacing of the metal ribbon was causing non-uniform waveguides and reducing the efficiency of the window. The final design sought to address this concern by spring tensioning the metal ribbon forming the walls of each waveguide. Loops of ribbon were stretched around steel dowels of the waveguide width and tensioned with a spring internal to the loop. This construction alleviates many

of the problems encountered in the first design. It allows the ribbon and frame to deform in temperature gradients and still maintain the shape of the waveguide window. Since the frame itself is constructed out of one aluminium block, there is much less chance of frame buckling due to the tension in the ribbon. Also, slack caused by loosening and creep over time should be taken up by the springs. The spring tensioned design was tested on the antenna and performed much better than the prototype. Much of the design work and constructin of the polarizer baseplate was performed by David Breslau of the MIT Center of Space Research.

3.2.6 Tests on the Antenna

In addition to laboratory tests, the behavior of the polarizer was also characterized on the antenna at Haystack. The metal vane spacing was, at that time, cause for some concern as measurements made under a microscope showed the mean spacing to be $x = 3.15$mm with a variation of $\sigma = 0.12$mm. Variations in spacing, in addition to causing phase losses, can scatter signal from the subreflector before reaching the feed. By measuring system and antenna temperature with and without the polarizer inserted we characterized the overall loss.

The first test used the prototype polarizer to observe the SiO maser emmission in Orion. The results are shown below :

	T_{sys} (K)	T_{SiO} (K)	SNR
Polarizer In	275.7±2.5	12.5±2.5	0.045
Polarizer Out	249.4±6.6	13.2±0.8	0.053

The polarizer, therefore, caused a loss in SNR of 17±5%. If all the loss was ohmic then the load should have recalibrated the feed-polarizer combination and the line strengths in the two cases should not differ. That this did not happen indicates scatter by the polarizer within the angle subtended by the load. This reduces the signal coming from the radio source and is not taken into account by the calibration system. Assuming an effective opacity for the polarizer and using nominal values for reciever temperature and atmospheric temperature the estimate for scattering loss is 13% and remaining losses are 4%.

When the final version of the polarizer was tested in this way, a noticeable improvement was evident. This time, the test source was Venus and the same type of measurements were made :

	T_{sys} (K)	T_{Venus} (K)	SNR
Polarizer In	230.1±3.0	2.24±0.05	0.0097
Polarizer Out	213.3±3.0	2.26±0.05	0.0106

The SNR loss of 9% is an improvement by nearly a factor of two from the prototype design. This increase clearly demonstrates the deleterious effects of uneven vane spacing. In this case the ohmic losses were estimated to be near 3% with 6% left as small angle scattering. These results place the metal vane polarizer with at least as low a loss as standard dielectric models. The spring tensioning should enable the vanes to maintain their present spacing even during temperature fluctuations.

Figure 3-3: The polarizer as shown extends into the page with a being the groove width and b the spacing between the grooves.
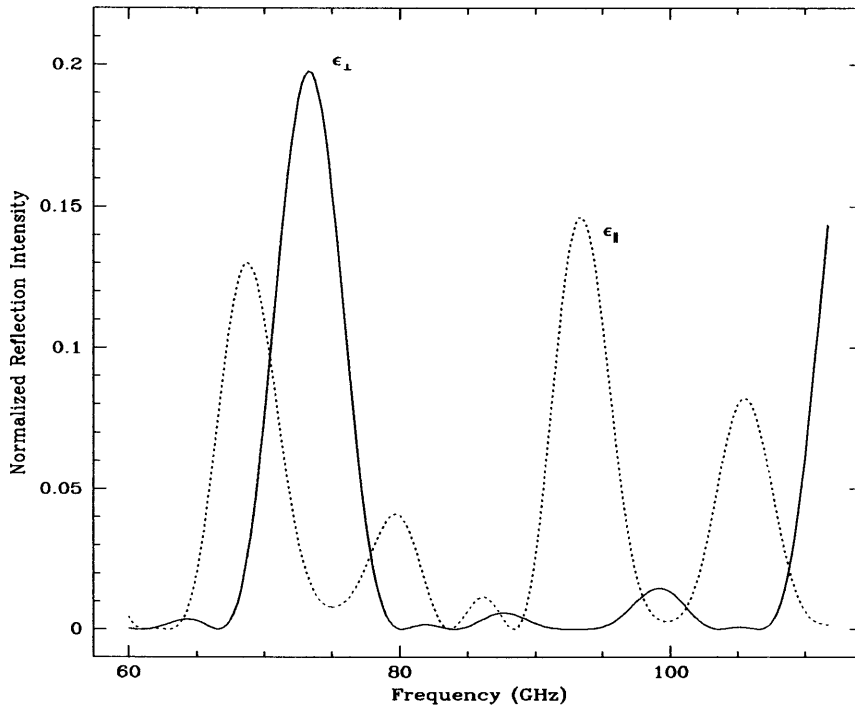
Figure 3-4: First order reflections off grooved dielectric polarizer made of teflon. The dimensions of the grooves are $a = 0.5$mm $b = 0.4$mm. The length of each comb section is 3.3mm and the thickness of the middle solid section is 2.0mm.

b   is the vane spacing

d   is the polarizer depth

$E_{\parallel}$ is the component of incident field parallel to the vanes

$E_{\perp}$ is the component perpendicular to the vanes
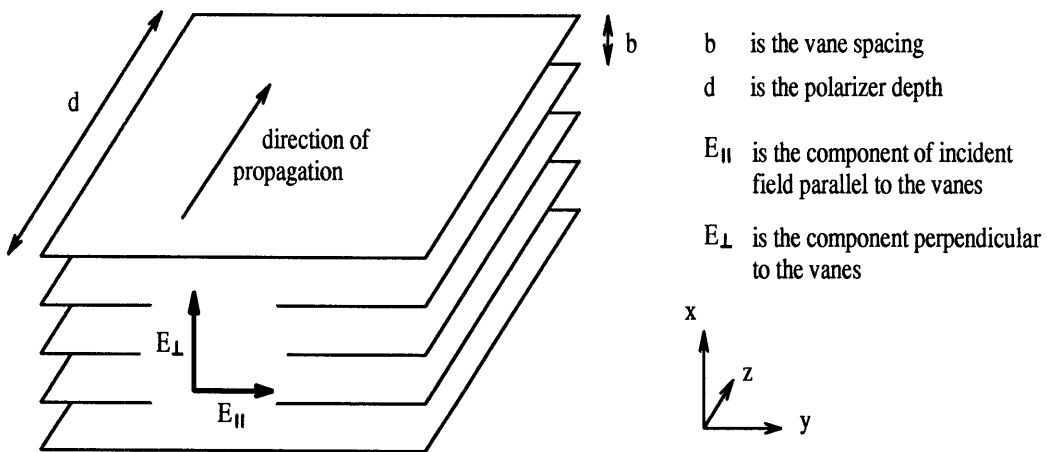
Figure 3-5: Schematic view of polarizer
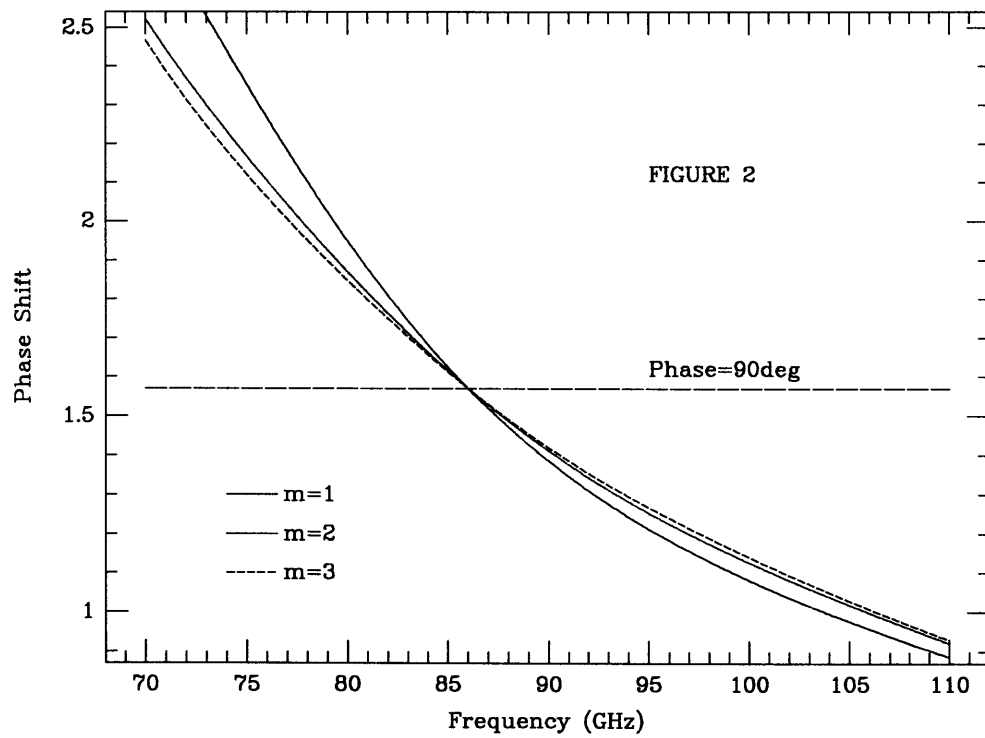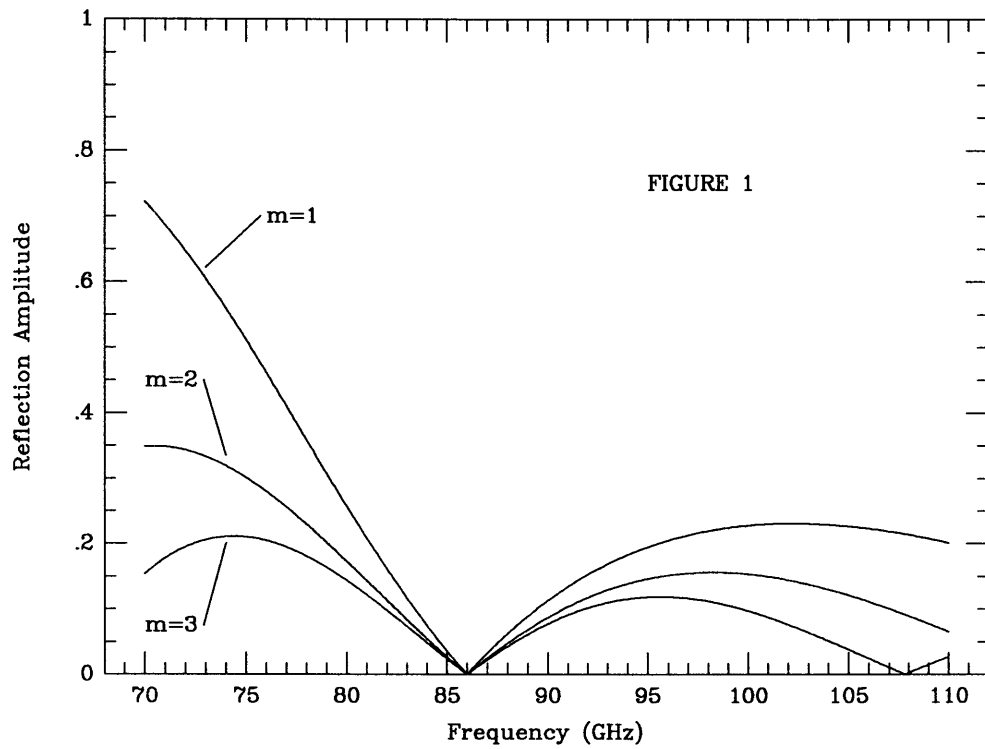
Figure 3-6: Reflection and phase shift from simple model of polarizer.
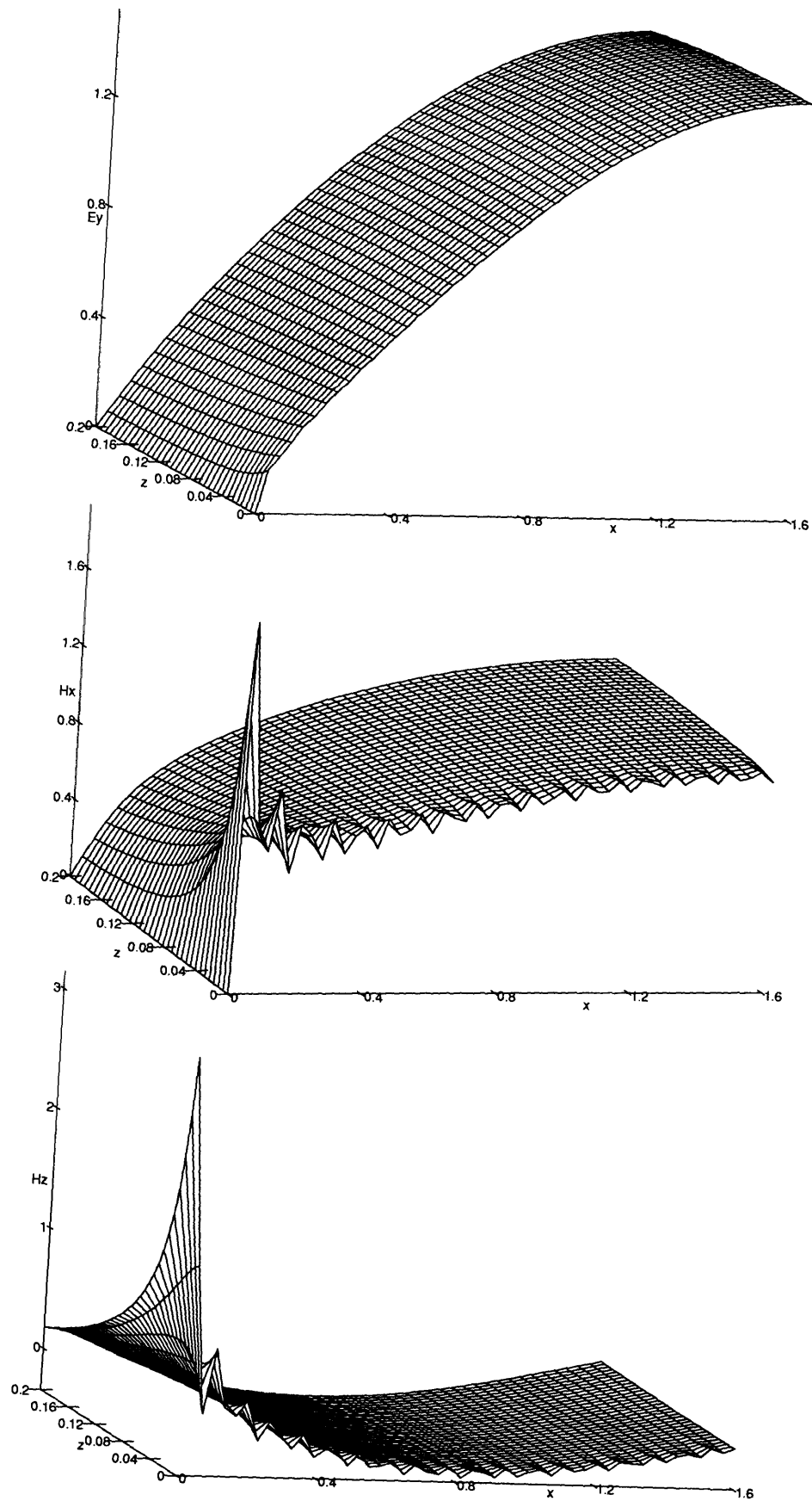
Figure 3-7: Ey, Hx, Hz Fields just inside the metal vanes. x ranges from vane surface to midway between vanes. Spatial dimensions in mm.

46
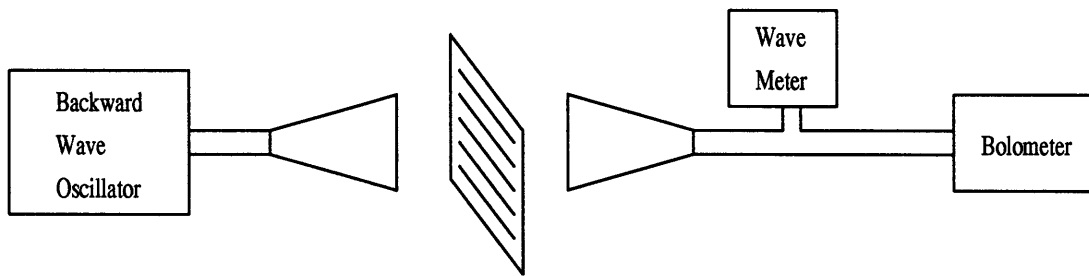
Figure 3-8: Experiment setup

Reflection from Polarizer Surface
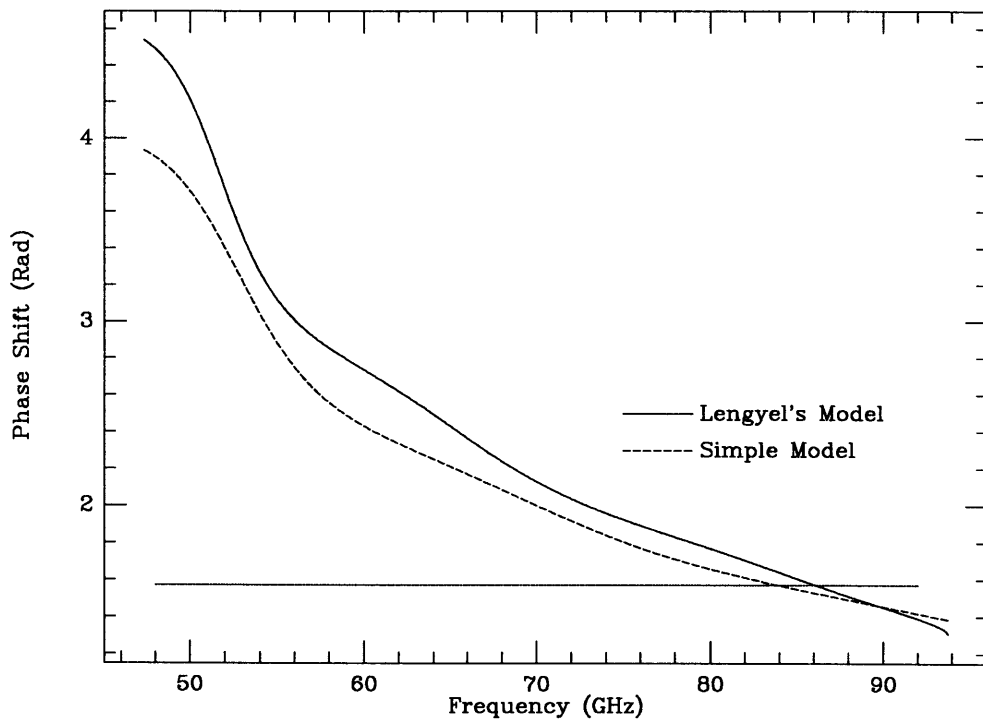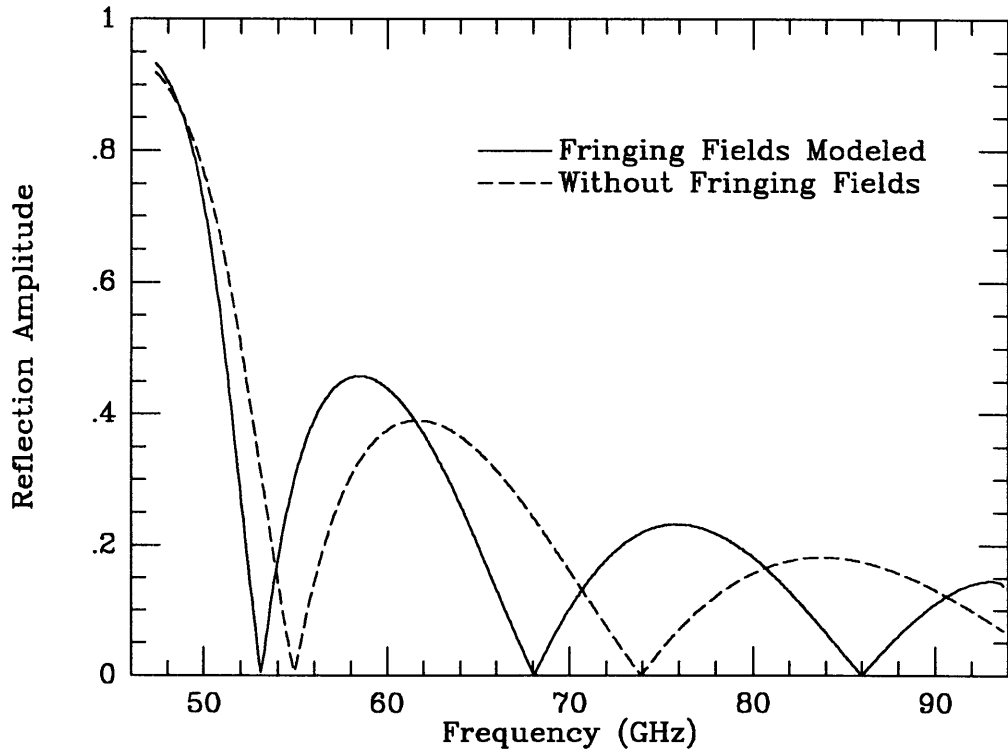
Figure 3-9: Comparison of Lengyel's reflection and phase shift with those of the simple treatment.
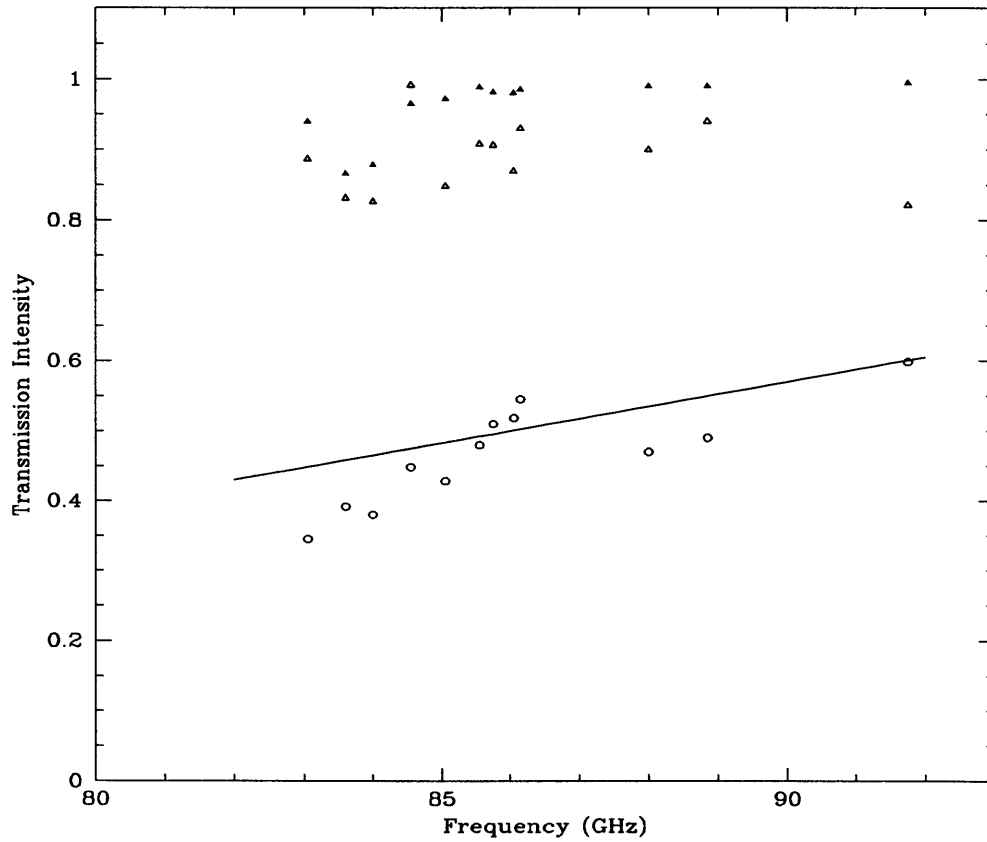
Figure 3-10: Transmission results for three polarizer orientations. The solid curve is the expected 45 degree transmission.

Chapter 4

Interferometric Detection

Noise in the correlator that affects determination of the complex visibility originates in the high system temperatures at the antennas. At mm-λ a nominal T_{sys} might be $\sim$300K while an average source signal would be only $\sim$0.5K. With this ratio in signal to noise it is remarkable that the signal can be recovered at all. Advantage is taken of the high data rates possible (256 Mb/s) to increase the number of samples and, hence, the SNR. With the bandwidth of current acquisition systems up to 112MHz, the number of samples ($2\Delta\nu\tau$) in an averaging interval of 20 seconds is $\sim 4.5 \times 10^9$. If the inputs to the correlator are two streams of 1-bit clipped data from a Mark III formatter, then the signal to noise of the resulting correlation can be expressed as :

$$\text{SNR} = L \sqrt{\frac{T_{a1}T_{a2}}{T_{s1}T_{s2}}} \sqrt{2\Delta\nu\tau} \tag{4.1}$$

where T_a= antenna temp., T_s= system temp., $\Delta\nu$= bandwidth, τ= averaging time, and L=loss due to digital signal processing. We can also write, in the case where the noise is much greater than signal, the correlation amplitude and noise as :

$$V = \sqrt{\frac{T_{a1}T_{a2}}{T_{s1}T_{s2}}} \quad and \quad \sigma = \frac{1}{L\sqrt{2\Delta\nu\tau}} \tag{4.2}$$

These expressions hold only if there are no losses incurred due to loss of coherence. Thus, it is assumed that the averaging time τ is not so large that the complex correlation cannot be averaged over that time. The loss factor L includes all digital losses due to the necessary quantization of signals from the antennas. $L \simeq 0.5$ for nyquist sampling of 1-bit data

processed by the MarkIII VLBI system.

To recover phase of the complex visibility requires that within the correlator two separate correlations take place. One takes as input the two antenna signals with one shifted so that they align in the time domain. This yields the real component of the correlation. The second uses a quadrature network to shift the phase of one input to the real correlator by $\pi/2$ which results in the imaginary component of the correlation. Taken together, the amplitude and phase of the complex visibility can be determined. The vector quantity that exits the correlator consists of visibility amplitude and phase. To increase the SNR it is coherently averaged for the duration of the observation :

$$\vec{Z} = \frac{1}{\tau} \int_0^\tau \vec{Z}_c(t) \, dt \tag{4.3}$$

where $\vec{Z}$ is the averaged visibility and $\vec{Z}_c(t)$ is the visibility output at each moment from the correlator. The degree to which this average properly estimates the true visibility depends on the phase stability of $\vec{Z}_c(t)$. Phase noise added to the antenna signals by atmospheric turbulence or noisy electronics will cause Eq. 4.3 to underestimate the average. If τ is small compared to any phase noise variation then the average can be used directly. In all VLBI applications one should always check that the integration time is at least comparable to the coherence time : $\tau \leq \tau_c$.

Estimating the visibility in the presence of noise is complicated by the two dimensional nature of this correlation. Since the product of signals in the real correlation is uncorrelated with the product in the imaginary correlator, there are two noise components associated with the visibility vector. This is shown in Fig. 4-1 with $\vec{V}$ being the true visibility, $\vec{Z}$ the measured visibility and $\vec{\epsilon}$ the noise vector.

The noise is made up of a real and imaginary component such that $\vec{\epsilon} = (\epsilon_x, \epsilon_y)$ with $\sigma = \langle \epsilon_x \rangle = \langle \epsilon_y \rangle = (2\Delta\nu\tau)^{-1/2}$. With this picture in mind we can now discuss and formalize methods of fringe detection.

4.1   Traditional Detection

The probability densities of $|\vec{Z}|$ and $|\vec{\epsilon}|$ are well determined and are Rice and Rayleigh distributions respectively. Detection algorithms normally sense the measured amplitude and check to see it it rises above some threshold SNR. The probability distribution of Z is
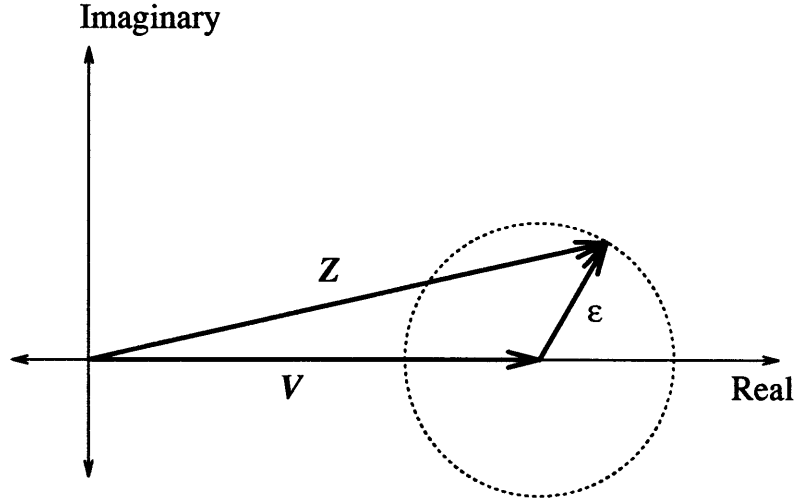
Figure 4-1: Depiction of noise in the correlator where $\vec{V}$ has been taken, arbitrarily, to be real and $\vec{Z}$ is the result after addition of the noise $\vec{\epsilon}$.

(Thompson, Moran & Swenson (1986), p.260) (hereafter TMS):

$$p(Z) = \frac{Z}{\sigma^2} \exp\left(-\frac{Z^2 + V^2}{2\sigma^2}\right) I_o\left(\frac{ZV}{\sigma^2}\right) \quad Z > 0 \tag{4.4}$$

with I_o= modified Bessel function of order zero. Now the question that must be asked is : what SNR is required of Z that we may correctly identify it as a detection? It has already been pointed out (Section 2.4) that a fringe search means evaluating Z at many values of delay and rate. Therefore, since noise is a stochastic process, the wider the search windows, the greater the chance a noise spike will rise high enough above the noise floor and be misidentified as a detection. TMS (p.266) give the probability that one or more samples of Z with no signal will exceed the threshold and be falsely identified as a signal detection. They find this probability or error to be :

$$\mathrm{PE} = 1 - \left[1 - e^{-\mathrm{SNR}^2/2}\right]^n \tag{4.5}$$

so that in a search over 350,000 pairs of (rate,delay) a PE of 10^{-4} requires a SNR of 6.6.

Once a detection is made the best estimate of amplitude is

$$V = \sqrt{Z^2 - 2\sigma^2} \tag{4.6}$$

since $\langle Z^2 \rangle = V^2 + 2\sigma^2$. In other words, the measured visibility amplitude has to be adjusted

for the fact that it is always positive and still registers a non-zero value if there is no signal. This noise bias correction lowers the amplitude of a 100 SNR detection (nominal cm-λ SNR) by only 0.02% and is ignored except in the low SNR case. For our argument in the next section the salient points here are that this scheme of detection uses a time average of the vector $\vec{Z}$ to increase SNR and hence to detect the source. If this time average is limited in duration due to phase noise then other methods not susceptible to phase variations can be employed which allow significantly longer averaging intervals. A new detection method of this type is discussed in Section 4.2.

4.1.1 Narrowing the Search

The dependence of PE on the number of search points presents the possibility of lowering detection SNR by restricting the scope of the search. A typical fringe search would be done in three dimensions over single-band delay, multi-band delay and rate. In SBD, there are 16 delay channels to search in the Haystack correlator. The search range for the rate is given by the reciprocal of the correlator accumulation period (AP) and the number of rate search points is twice the total number of APs. The search in MBD depends on the particular setup of frequencies observed but for 3mm-VLBI, one generally searches over twice the number of frequency channels. This type of search should exactly nyquist sample the 3-D search space; nominal values are given here :

$$\text{MBD points} = 28 \text{ ("Mode A")}$$
$$\text{SBD points} = 16$$
$$\text{Rate points} = 780 \text{ (6.5 min. scan, AP=1 sec.)}$$
$$\text{n} = 350{,}000 \text{ search points}$$

If n can be reduced by using a-priori information then a lower SNR will correspond to the same PE given in Eq. 4.5.

Alef and Porcas (1986) describe a method which uses known detections to estimate SBD, MBD and rate on baselines for which no fringes are found. They assume that all rate and delay errors can be attributed to station based factors so that in an array of N stations there are N-1 delays and N-1 rates to be found. They make the further approximation that delay and rate add to zero around closed loops of baselines (Appendix D). Baseline based closure errors that would not close around a loop do exist but introduce small effects and are

53

more important in imaging than in detection. A discussion of non-closing effects is given in (Thompson & D'Addario, 1982; Pearson & Readhead, 1984). In the Alef and Porcas algorithm, a least squares analysis then uses the rates and delays from firm detections to assign new rates and delays for the non-detection baselines. New fringe searches center on these new values and decrease the search range. Other information such as adjacent detections in time can also be used to interpolate delay and rate.

To show how this works we take a situation in which three stations observe a source and two make detections but the third does not. By forcing rate and delay to close around the triangle of baselines, the search range on the weak baseline is limited to a smaller area where detection is more likely. If this reduction means that n decreases from 350,000 to 600 points, then the SNR needed for detection falls to 5.6 from an original value of 6.6. Now the weaker signal stands a better chance of detection.

4.1.2 Early Incoherent Averaging

Another accepted method of boosting sensitivity, especially in low SNR data, is to average the visibility beyond the coherence time. Section 4.2 discusses this idea in detail but as an historical and motivational platform, the basics as understood in the mid-1970's are important and outlined here. In most cm-λ observations, SNR is high enough and the coherence time is long enough that this extended averaging is either unecessary or too slow to deliver the promised sensitivity. For mm-λ, the benefits are more clear.

We start with the unbiased estimate of amplitude given in Eq. 4.6 that applies to one time interval of VLBI data. Many such intervals (all of length equal to τ_c) are combined to maintain the unbiased nature of the amplitude estimate :

$$|V|^2 = \left(\frac{1}{N} \sum_i Z_i^2 \right) - 2\sigma^2 \tag{4.7}$$

with N = the number of segments so that a total interval of $N\tau_c$ is considered in the average. TSM show that the SNR of this estimate is :

$$\mathrm{SNR} = \Delta\nu\tau_c \frac{T_{a1}T_{a2}}{T_{s1}T_{s2}} \sqrt{\frac{N}{1 + \left(\frac{2T_{a1}T_{a2}\Delta\nu\tau_c}{T_{s1}T_{s2}} \right)}} \tag{4.8}$$

Bad weather has the cumulative effect of lowering both τ_c and T_a while increasing T_s so

54

that at high frequencies it is not unusual to enter a regime where $T_a \ll T_s/\sqrt{2\Delta\nu\tau_c}$ which simplifies the SNR to

$$\mathrm{SNR} \simeq \tau_c\Delta\nu\left(\frac{T_{a1}T_{a2}}{T_{s1}T_{s2}}\right)\sqrt{N} \qquad (4.9)$$

Choosing an SNR that implies likely detection, we can write:

$$S_{min} = \frac{2kN^{-\frac{1}{4}}}{A_1A_2}\sqrt{\mathrm{SNR}_{min}\frac{T_{s1}T_{s2}}{\Delta\nu\tau_c}} \qquad (4.10)$$

with S_{min} being the flux density threshold. TSM do not discuss how the SNR_{min} might be chosen and this is covered in the next section. Historically, the $N^{-\frac{1}{4}}$ factor was seen as too slow an improvement to warrant the large amounts of observing time needed. However, some past experiments have attempted to use this technique. Stellar binaries, for example, are often observed at cm wavelengths with VLBI but are also quite weak with flux densities in the mJy range. Early efforts to detect these sources tried to incoherently average the visibility to extend the integration time. No sources were ever convincingly detected in this manner possibly due to their variability and faintness, and this avenue was left largely unexplored as new methods such as phase referencing proved reliable and easy to implement. In phase referencing the coherent integration time can approach hours by using a nearby bright source as a phase calibrator. For mm-VLBI, the phase variations are on such a short time scale that phase referencing is not an option and we must give incoherent averaging a closer look.

4.2   Incoherent Detection

Though incoherent averaging was recognized as early as 1968 (Clark) as a potential detection method, it lacked a rigorous determination of SNR and clear detection thresholds. Without such a framework we cannot reliably protect against the possibility of latching onto a false detection due to noise. Since an incoherent search evaluates a different quantity than the coherent case, a new formalism has to be created that can accurately asses the validity of a detection. In this section we state clearly the goals and methods of an incoherent search. The effects of noise on the search are included and a prescription for the probability of error in the search is explored. The reference for much of the work in this chapter is Rogers, Doeleman & Moran (1995).

55

The incoherently averaged visibility as given in (Moran,1975) is Eq. 4.7 in Section ref-sec:early. His formulation implies that all segments to be averaged are the same length and weights them all equally. It turns out that there are weighting schemes that optimize SNR and so the weights, for now, will be explicitly left in the estimate of visibility :

$$A^2 = \frac{1}{\sum w_i} \sum w_i (Z_i^2 - 2\sigma_i^2). \tag{4.11}$$

Here, A^2 is the square of the visibility estimate, w_i are the weights for each segment, Z_i is the measured amplitude and σ_i is the noise in the ith segment. As long as $\langle w_i \rangle = w_i$ we see that the estimate will be unbiased since

$$\langle A^2 \rangle = \frac{1}{\sum w_i} \sum w_i |V|^2 = |V|^2. \tag{4.12}$$

The signal to noise of this estimate is then found by computing the expectation divided by the variance : $\mathrm{SNR}_A = \langle A^2 \rangle / \sqrt{\langle A^4 \rangle - \langle A^2 \rangle^2}$. Appendix B shows some of the interim steps required to write

$$\langle A^4 \rangle - \langle A^2 \rangle^2 = \frac{1}{\left(\sum w_i\right)^2} \left[\sum w_i^2 (4V^2 \sigma_i^2 + 4\sigma_i^4) \right]. \tag{4.13}$$

With no approximations, the form of the SNR is then

$$\mathrm{SNR}_A = \frac{|V|^2 \sum w_i}{2\sqrt{\sum w_i^2 \left(\sigma_i^2 |V|^2 + \sigma_i^4\right)}} \tag{4.14}$$

With this expression for the SNR we can return to the question of what the weights should be. If no assumptions are made about equal length segments then the optimum weighting will depend on the SNR of individual segments. In the low SNR case ($|V| \ll \sigma$) the weights should be segment length squared : $w_i = t_i^2$. For high SNR, $w_i = t_i$ is appropriate. Appendix B contains a discussion of this optimization.

Taking stock, it is seen that if the individual segments are themselves coherent averages over intervals small compared to the coherence time, then $\langle A^2 \rangle = |V|^2$ will accurately reflect the true visibility. The philosophy behind this estimate is to always segment the data and work with averages of these segments. This sidesteps the phase noise problem by never allowing coherence loss to occur within a segment where it could propagate into the average

and contaminate the visibility estimate. The problem of coherence loss is thus solved but the difficulty of determining the required SNR_A for fringe detection remains.

4.2.1 Probability of Error

One way to quantify the probability of error in a fringe search is to ask how high a peak would we expect in the absence of signal. Then, if we find a SNR_A peak at a point of delay and rate that well exceeds this expected noise peak, we can say that the probability of error is low. For each independent point in the search a chance exists that a noise spike will masquerade as signal so the form of PE_A must be

$$PE_A = 1 - \left[1 - \int_{A_c^2}^{\infty} p(A^2)\, dA^2\right]^n \qquad (4.15)$$

In this equation, the term in brackets is the probability that the peak in visibility amplitude due to noise lies below A_c^2. This is raised to the nth power to take into account each search point. Thus, PE_A is the probability that at least one noise peak will be above the threshold.

We first evaluate PE_A by examining the limiting case of many segments. For a large number of segments, the probability density of A^2 ($p(A^2)$) will approach a gaussian form by the central limit theorem. The probability of error in the search is then

$$PE_A = 1 - \left[1 - \frac{1}{\sqrt{2\pi}} \int_{SNR_A}^{\infty} e^{-x^2/2} dx\right]^n \qquad (4.16)$$

and can be approximated for large n and small PE_A by

$$PE_A \simeq \frac{n\, e^{-SNR_A^2/2}}{\sqrt{2\pi}\, SNR_A} \qquad (4.17)$$

Where the asymptotic expansion for the error function erf(x) has been used. We frequently, though, need to evaluate PE_A for a moderate number of segments and the formula above will not apply.

Since, in the general case, the gaussian approximation made in Eq. 4.16 fails, other means for calculating PE_A are needed. One method is to find an analytic expression for $p(A^2)$ and use Eq. 4.15. Fortunately, the exact form of $p(A^2)$ is not too difficult to write

(Appendix A):

$$p(A^2) = 2 \left(\frac{N}{2\sigma^2} \right)^N \frac{(A^2 + 2\sigma^2)^{N-1}}{(N-1)!} \exp(-\frac{N(A^2 + 2\sigma^2)}{2\sigma^2}) \tag{4.18}$$

So given a search size n and a measured maximum A_o^2, the prescription for PE_A is to integrate the probability destribution and use Eq. 4.15. Fig(A.1) shows the form of $p(A^2)$ for a few values of N along with the gaussian approximation for N=50.

A second method of finding PE_A uses simulations to statistically estimate results of a search. A single simulation represents a search for a maximum over n samples of A^2 *with no signal*. We then divide the resulting distribution of noise maxima into two groups : those above A_o^2 and those below. The fraction of those maxima above A_o^2 is the probability of error.

Fig. 4-2 combines these two methods and shows how the PE_A grows as the search size does. The curves hold for an incoherent average of 20 segments. If the number of segments were to increase, the curves would all shift to the left so that a smaller SNR_A would give the same probability of detection. The opposite is true for averages of fewer segments. Intuitively we can see that this should be the case because as more segments are averaged, the effect of a noise spike in one of the segments will be diminished. And, if only a few segments are combined, then anomalously high noise in one can more easily raise the average value. But the satisfying aspect of Fig. 4-2 lies in our ability to examine a map of SNR_A and say with assurance that a peak is *probably not* noise. A typical search over 300 points of delay and rate with a 20 segment average will need to achieve an SNR of at least 3.8 before one can confidently say that the probability of error is less that 10^{-6}. It must be noted that this PE describes a situation in which one mistakenly detects a source in the presence of pure noise. If one attempts to detect a source but finds none then it is possible to calculate rigorous upper limits on the source strength. No effort has been made here to do so. Such a calculation will have to be completed for any future 3mm-VLBI survey which may result in many non-detections.

4.2.2 Comparison with Coherent Detection

A new detection method cannot be proposed without directly comparing it to previous and verified work. Using the results above, the flux thresholds for both the incoherent and co-
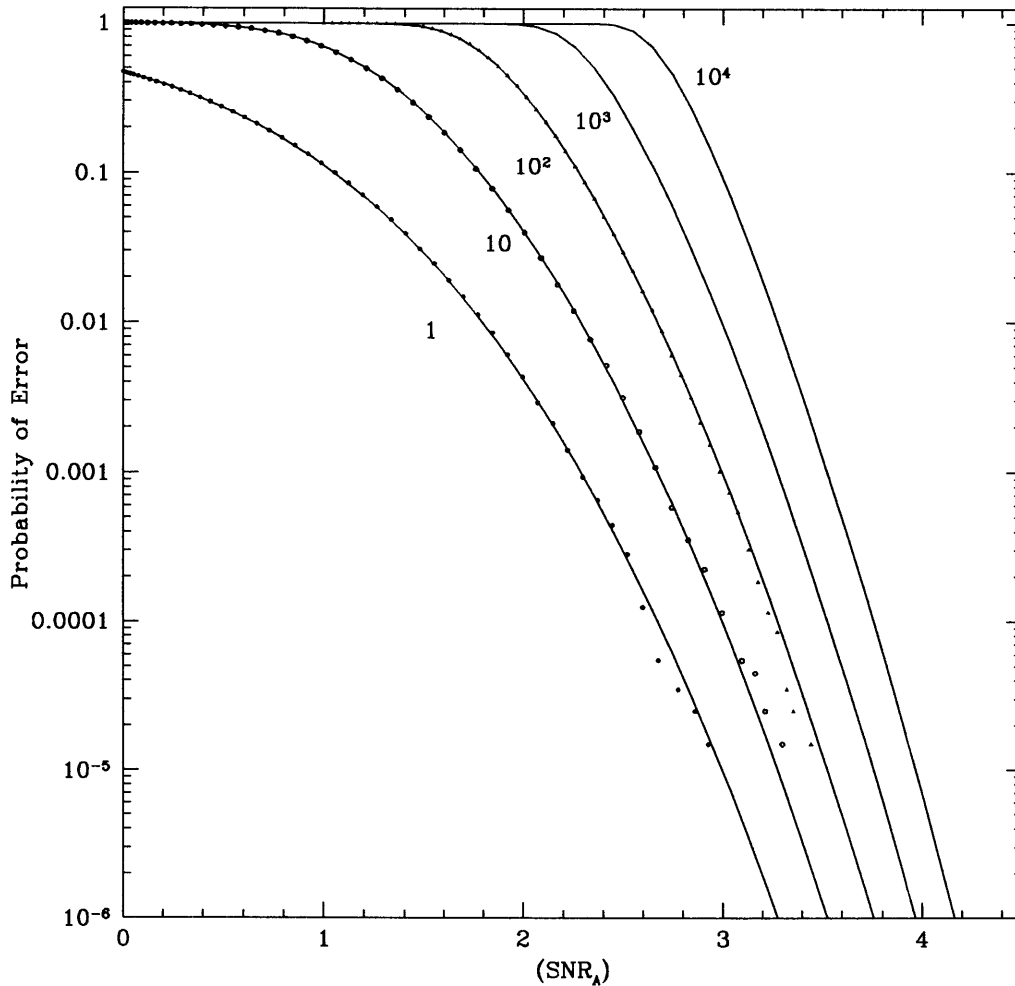
Figure 4-2: Probability of Error for an incoherent search. Each solid curve represents a search of a 20 segment incoherent average over the number of elements shown. The points show the same probability found by simulating many searches.

herent searches can be compared under different sets of assumptions. For these comparisons we will assume that the segments are of equal length and that effects of weighting can be ignored.

A value of incoherent SNR (SNR_A) corresponds to an equivalent signal for a single segment which is found by inverting Eq. 4.14. We do this first for the case of low SNR and a large number of segments as a limiting case, then use the more general (but messier) approach of simulation and numerical integration. In the low SNR case we write

$$\frac{|V|_A}{\sigma} = \sqrt{2\,SNR_A}\ N^{-\frac{1}{4}} \tag{4.19}$$

where we take SNR_A and $|V|_A$ to be threshold values. The thresholds for coherent detection

in one segment can be expressed as

$$\frac{|V|_T}{\sigma} = \sqrt{\mathrm{SNR}_T^2 - 2} \qquad (4.20)$$

with SNR_T and $|V|_T$ also threshold values. The subtraction of 2 from SNR_T^2 is required to correct the coherent average for noise. The ratio of threshold flux densities in the incoherent and coherent methods is then:

$$\frac{S_{incoh}}{S_{coh}} = N^{-1/4} \sqrt{\frac{2\,\mathrm{SNR}_A}{\mathrm{SNR}_T^2 - 2}} \qquad (4.21)$$

To directly compare the thresholds, care must be taken that the values of SNR_A and SNR_T each correspond to the *same* probability of error in *identical* searches. For example : choosing values of $\mathrm{SNR}_A = 6.6$ and $\mathrm{SNR}_T = 7$ results in the same PE=0.01% for both methods in a search of 10^6 points. This puts the ratio of flux thresholds at

$$\frac{S_{incoh}}{S_{coh}} = (0.53)N^{-1/4}. \qquad (4.22)$$

The numerical factor in the last equation very much depends on the search size and the PE one requires for detection. And the overall advantage of incoherent averaging further depends on factors such as weather, observing frequency, and the number of segments being averaged. In fact, the careful reader will have noticed that the proffered example necessitates an extremely large number of segments (~ 3000). This is due to the fact that we have assumed small signal (to use Eq. 4.19) but have also demanded low PE (to use Eq. 4.17). This example was not meant as a typical situation one might encounter but as a verification of the $N^{-1/4}$ dependence seen in Section 4.1.2 for large N.

When a more realistic average is computed, and smaller values of N apply, the above analysis will not hold. First of all, the full form of PE would have to be used, not the gaussian approximation; secondly, the assumption that $|V|/\sigma \ll 1$ may not be valid. In this case, integration of $p(A^2)$ is one way to circumvent these problems. To compare the detection methods, a PE and search size are fixed and the expression for PE evaluated to find the value of A^2. An equivalent simulation analysis can be performed by creating a distribution of search maxima. The threshold value of A^2 will divide the distribution so that a fraction PE of maxima are above the threshold. Fig. 4-3 shows the threshold

reduction using incoherent averaging as a function of segment number with $SNR_A = 3.8$ and $SNR_T = 4.3$ which both correspond to a PE of 0.01 in a search of 100 points. The asymptotic limit described by Eq. 4.21 is also shown and agrees well with both the simulation and numerical integration for large values of N.

The advantage gained by incoherent averaging has, then, two regimes which are qualitatively distinct. At large N, the asymptotic fourth root of N dependence takes over but for small and mid-range values of N a transition region is entered with a power law slope of approximately 1/3. This shows incoherent averaging to be more powerful than previously recognized and an especially potent tool to deal with data of poor coherence. In general, we can now split interferometric observations into three catagories based on the coherence time (τ_c) allowed by weather and other sources of phase noise and the total duration of the observation (τ_s). If the total observing time remains within the coherence time then one should always coherently average to gain maximal sensitivity. If the observation exceeds the coherence time and a moderate number of segments are combined incoherently, then the threshold lowers more slowly. And, if the observation is long enough so that many segments must be averaged, the improvement is worse. Sensitivity (S_{th}) dependence on time in all three domains are specified in Table 4.1.

$\tau_s < \tau_c$	$S_{th} \propto \tau_s^{-1/2}$
$\tau_s > \tau_c$ small N	$S_{th} \sim \tau_s^{-1/3} \tau_c^{-1/6}$
$\tau_s > \tau_c$ large N	$S_{th} \propto \tau_s^{-1/4} \tau_c^{-1/4}$

Table 4.1: Three different regimes of signal integration.

As a final note, the point should be stressed that more than one algorithm exists for the segmentation of data. The general practice has been to average coherent non-overlapping segments but many more segments can be formed by shifting the boundaries of a fixed segment length and resegmenting. As an example, two sets of segments can be derived from a common dataset by offsetting the segment boundaries so that the boundaries of one set are midway between the boundaries of the other set. In this case we multiply the number of segments by two with some overlapping. Incoherently averaging the combined sets of segments results in a small reduction in detection threshold of $(3/4)^{1/4}$ or $\sim 7\%$. The

optimum segmentation scheme is a true running mean which offsets each segment by the smallest time unit possible - the accumulation period of the correlator. This increases the number of segments dramatically and lowers the detection threshold by $(2/3)^{1/4}$ or about 10%. Appendix C details the calculation of threshold reduction for offset segments.

4.3 Application to VLBI Data -

The success or failure of an incoherent search on VLBI data depends almost entirely on how one segments the data. In most VLBI applications, the raw data format will be "scans" that correspond to the number of bits that fit onto one pass of a magnetic tape. In the MarkIII system, 6.5 minutes is a typical scan length - a duration which usually far exceeds the coherence time at mm wavelengths. This situation - a fixed amount of data and the need to segment and average - naturally leads to an optimal segmentation length that maximizes SNR_A. Before mass procesing of a large experiment begins, a few sample scans are always examined to determine the best coherent averaging interval.

As we segment one of these test scans, we expect to see a specific evolution in SNR_A based on segment length. When the scan is split into only a few segments, then each one is still fairly long and suffers losses due to phase noise so the SNR_A will remain low. As the segment length approaches the coherence time, SNR_A will reach a peak. But since we have assumed a fixed scan length, each segment must grow smaller if this process is continued. If this happens, SNR_A will start to decrease even though we keep generating more segments. This tendency can be shown by rewriting the SNR formula in Eq. 4.14 with explicit dependence of $|V|/\sigma$ on N :

$$ SNR_A \propto \left(\frac{1}{N} \right) \left(N^{1/2} \right) = \frac{1}{\sqrt{N}}. \qquad (4.23) $$

The same arguments also predict how the measured amplitude will respond to segmentation. At first, the average will underestimate the true amplitude and the estimate will increase the more we segment. Once the coherence time is passed, the amplitude will sit on a plateau but become more noisey as SNR_A drops due to the arguments above. Examples taken from an April 1993 experiment illustrate this point in Fig. 4-4. The top two graphs show SNR_A and amplitude on a low SNR scan that fits the description given above. The SNR_A peaks at an optimum τ_{seg} and the amplitude increases to some definite limiting value at small

τ_{seg}. The bottom two graphs show a pathological case where the coherence time is less than two seconds. The SNR_A continues to increase with smaller τ_{seg} and never turns over. The amplitude estimate never reaches a final value even at a $\tau_{seg} = 2$ seconds. In this case we cannot calibrate the visibility since we underestimate the true visibility by an unknown amount.

Resolution in rate is affected by the number of segments and can change the number of points needed in a search. Recall that segmenting a scan and averaging is equivalent to integrating the fringe rate spectrum and that the resolution in fringe frequency will be approximately $1/\tau_{seg}$. This has the effect of smearing out SNR_A peaks in rate as the number of segments grows. So, if the segment length was 20 seconds with one second APs from the correlator, then the search range would be $\pm 0.5Hz$ with a search grid spacing of 25 mHz in rate. This reduces the number of possible rate points to 40 from an original 780 for a full 6.5 minute scan undergoing a coherent search.

Fig. 4-5 shows four maps of SNR_A for a poor coherence 3mm-VLBI scan. Each contour plot reflects a different segmentation length and a different probability of error that the peak is a true detection. At 60 seconds, the PE for the peak at (rate=-0.4ps/s, delay=-25ns) is marginal due to the large number of points to search and the SNR_A loss due to phase noise. As the segment length decreases, the peak increases in height and begins to smear in the rate direction. The much reduced PE for $\tau_{seg} = 12$ seconds is due to a smaller search grid and the increased SNR_A. The coherence time of this scan is between 3-10 seconds.

4.3.1 Segmented Closure Phase and Bispectrum

Much effort has been spent to protect the amplitude estimate and detection limits from interferometer phase variations. But, at some point, the visibility phase itself must be recovered - imaging cannot proceed without it. The point has been made that the phase on an individual baseline cannot be recovered and VLBI relies on the closure phase. This sum of phases around a triangle of baselines is subject to the same coherence-type problems as the amplitude. The best estimate of closure phase will, therefore, come from an average of closure phases determined on a segment by segment basis. The exact method of determining this average has been a subject of debate and (Rogers,Doeleman,Moran, 1995) give an argument that supports averaging the bispectrum or triple product of the visibilities. In this section, the SNR of the bispectrum is given and shown to be superior to that of an

average of closure phase. A method to reassign baseline phases based on closure phases is shown to be necessary and is also described.

We consider the small signal case where even small advantages in SNR are important. To average the closure phase with no attention paid to the individual SNR of each segment we write :

$$\vec{C} = \frac{1}{N} \sum_k e^{i(\theta_c)_k} \tag{4.24}$$

where $(\theta_c)_k = (\theta_1 + \theta_2 + \theta_3)_k$ is the closure phase on the kth segment. For a point source we will want to compute the SNR as

$$\text{SNR}_C = \frac{\langle \text{Re}C \rangle}{\sqrt{\langle (\text{Re}C)^2 \rangle}} \tag{4.25}$$

since the visibility for an unresolved source should be completely real. RDM show that

$$\langle \text{Re}C \rangle = \sqrt{\left(\frac{\pi}{8}\right)} s_1 s_2 s_3 \quad \text{and} \quad \langle (\text{Re}C)^2 \rangle = \frac{1}{2N} \tag{4.26}$$

where $s_i = |V_i|/\sigma_i$. The SNR_C is then

$$\text{SNR}_C = \sqrt{\left(\frac{\pi}{8}\right)} \sqrt{2N} \, s_1 s_2 s_3. \tag{4.27}$$

A similar analysis can be applied to the bispectrum :

$$\vec{B} = \frac{1}{N} \sum_k \left[|Z_1||Z_2||Z_3|e^{i\theta_c} \right]_k \tag{4.28}$$

and the SNR_B, defined as

$$\text{SNR}_B = \frac{\langle \text{Re}B \rangle}{\sqrt{\langle (\text{Re}B)^2 \rangle}} \tag{4.29}$$

is found to be

$$\text{SNR}_B = \frac{\sqrt{N} s_1 s_2 s_3}{2}. \tag{4.30}$$

Comparison of these two SNR for the closure phase shows

$$\frac{\text{SNR}_B}{\text{SNR}_C} \simeq 1.4 \tag{4.31}$$

so that for low SNR data on each baseline, the averaged bispectrum shows up to a 40%

improvement in SNR over a straight phasor average. Or, put another way, the flux density threshold for a good determination of closure phase is ~10% less than if the phasor average is used. For all data reduction in this work the bispectrum will be used exclusively. In addition to this formulaic approach, there is a more heuristic argument for use of the bispectrum. Consider the representation of visibility in Fig. 4-1. On the average, we can see that the larger the SNR (Z/σ) the nearer the argument of Z will be to the true phase. When the bispectral average is formed we are just giving slightly more weight to those segments whose phases are nearer the true value. This bispectral average also lowers the weight on anomalously low SNR segments which drop out for instrumental or atmospheric reasons.

In analogy with the amplitude, error in phase estimation depends on the degree of scan segmentation. For if the segments are too large then by the very definition of coherence time the phase variations will prohibit accurate determination of closure phase. When the segments themselves start to approach phase stability the bispectrum will begin to deliver believable results. Fig. 4-6 demonstrates this principle. The bispectrum derived closure phase on the Haystack-Onsala-Quabbin triangle starts off with large errors and random values when the segment length is too high. When all three scans are segmented to a length of ~10 seconds, the closure phase converges to a value just below zero with smaller uncertainty. The rms phase error starts a descent and the SNR_B a rise right around this point as well.

The use of an average to estimate closure phase introduces an added complication. Most imaging software packages make use of the fact that in any VLBI array of more than three stations, there are redundant triangles. They are redundant in the sense that their closure phase can be arithmatically determined by adding linear combinations of other triangles. So for four stations ABCD, the closure phase on BCD is equaivalent to the sum of closure phases : ABC - ABD + ACD. But this is true only for an average of one segment. Indeed most VLBI applications just segment data and do no further averaging which leaves this equivalence intact. When averaged though, the noise on each triangle is independent and no longer exactly cancels in the linear combination. Therefore, it is not permissible to keep only the "independent" triangles in an array. All triangles contain different information and provide different estimates of the same closure phases. Fig. 4-7 show the correlation between the phase on BCD and phase of the sum above for different numbers of segments

in the average.

The probability density of the closure phase can be approximated by the same functional form as that of the single baseline phase (Rogers et al,1984). The relation between the closure phase SNR and that for each individual baseline is given in Eq. 4.30 for low SNR. If a linear method of image reconstruction is employed, such as least squares fitting of a model in the u,v plane, then errors in closure phase can be related to effects in the resulting map. The method of self-calibration (Wilkinson,1988), however, is inherently non-linear and it is not clear how closure phase errors directly affect the final image.

Threshold Reduction

○ Simulation Results

— Results of Numerical Integration

····· Curve Showing $N^{-0.25}$

·—· Asymptote For Large N (eq. #4.21)

Initial Slope is = −0.36

Number of Segments Averaged
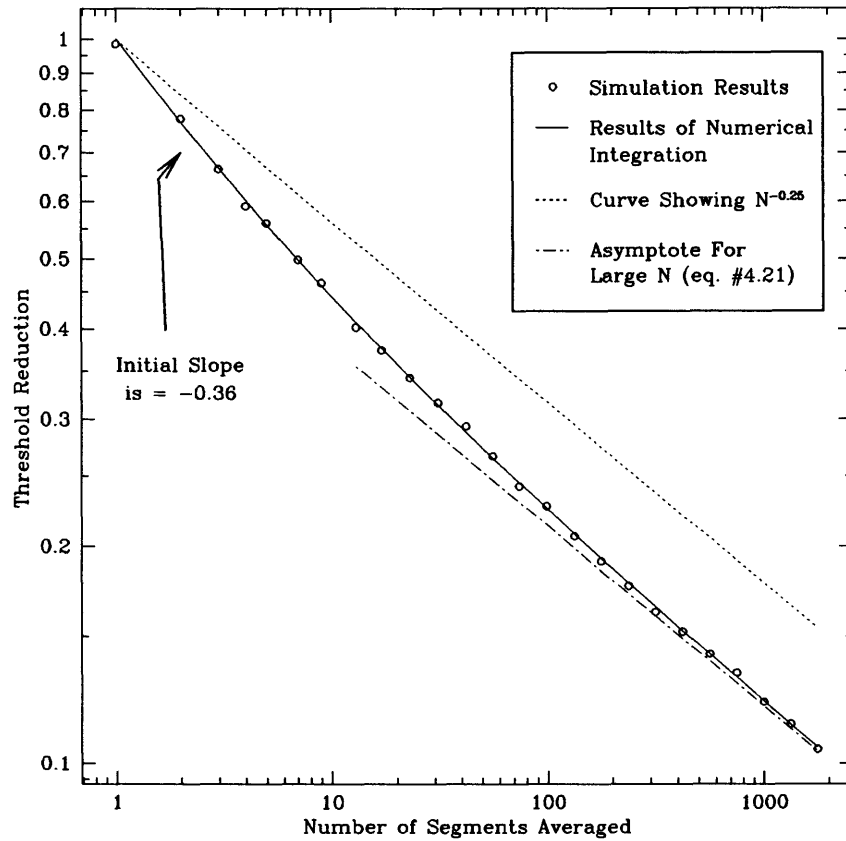
Figure 4-3: Reduction in the flux detection threshold obtained by searching over an incoherent average of many segments. The open circles show results of a simulation in which a maximum in segmented and averaged amplitude was found in a search of 100 points in (delay delay-rate) space. 2000 of these searches were made and the threshold was defined as the dividing line between the top 1% and lower 99%. The solid line is calculated by numerical intergration of the probability density for A^2.
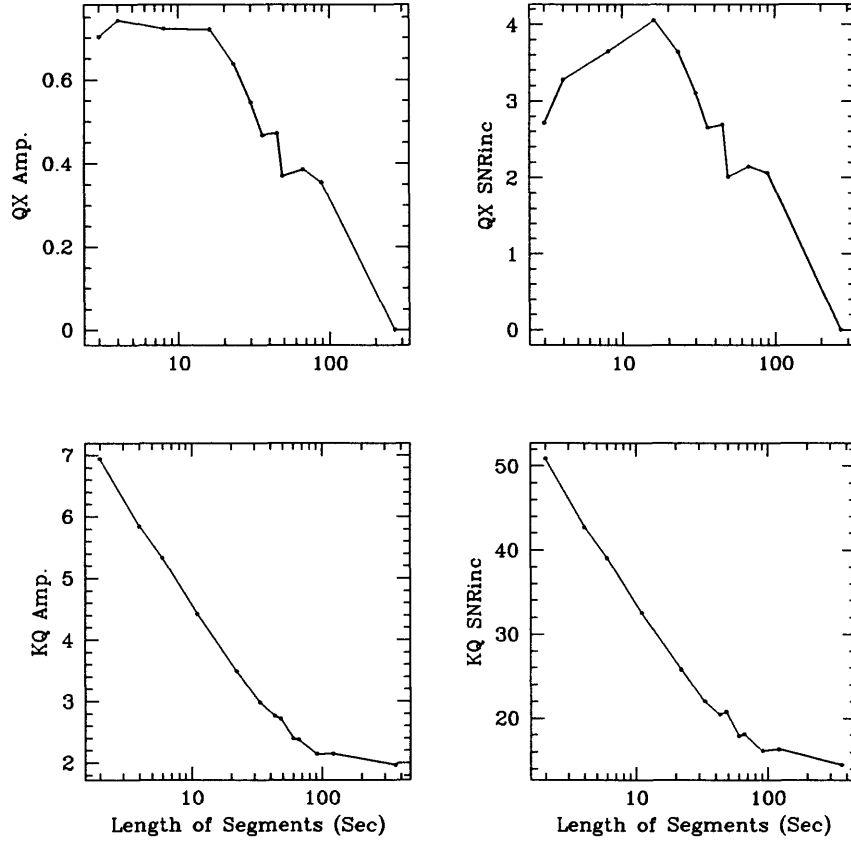
Figure 4-4: Incoherently averaged amplitudes and SNR as functions of segmentation length. The top two panels show a low SNR scan of 3C111 on the Quabbin-PicoVeleta (QX) baseline whose SNR_{inc} peaks near $\tau_{coh} = 12$sec. The amplitude increases as segment length decreases to τ_{coh} and reaches a limit for smaller segments. Lower panels show the same quantities for a Haystack-Quabbin (KQ) scan on 3C273. Here the SNR is higher but the coherence is much worse : even at a segment length of 2 sec. there is no peak in SNR_{inc} and we have not recovered all the amplitude.
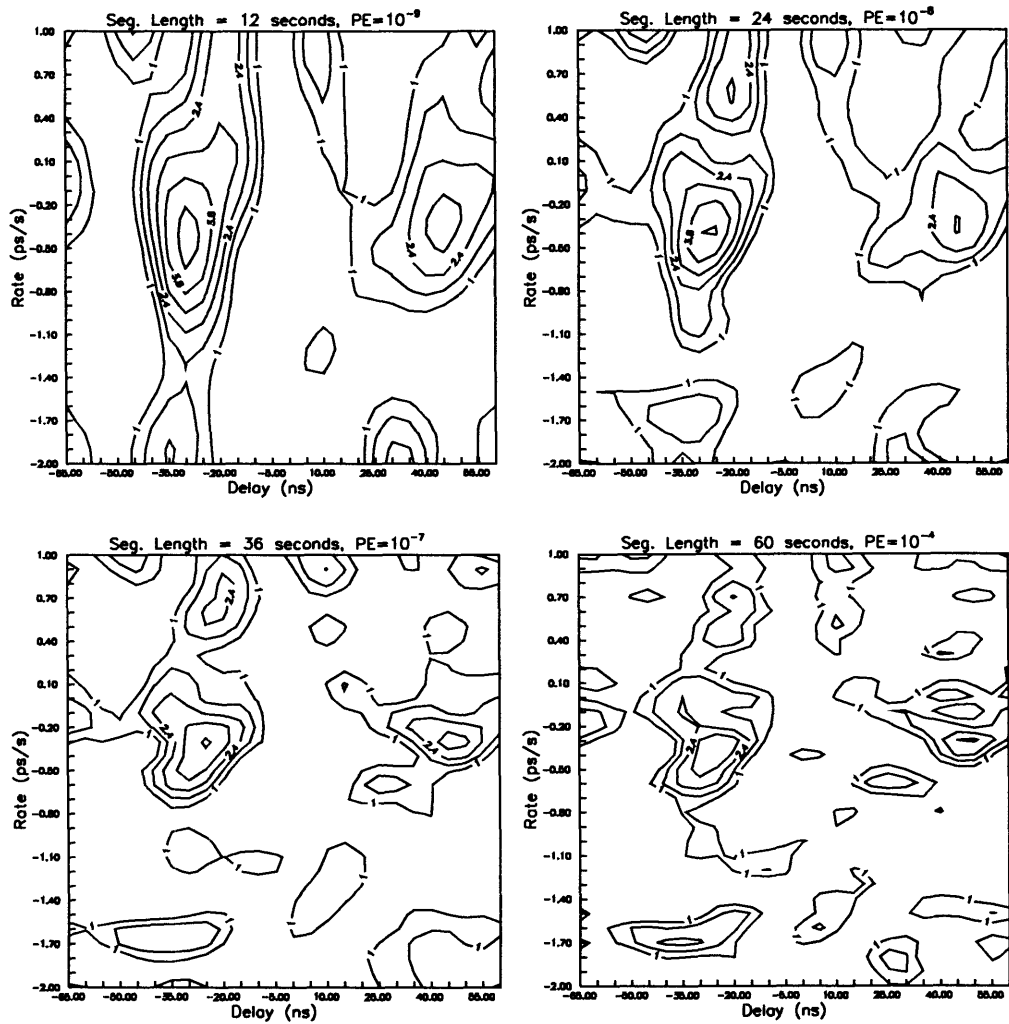
Figure 4-5: Four searches on the same data but with different segmentation lengths. Each contour map shows SNR_A as a function of delay and rate with the segment length and formal probability of error at the top. At 60 seconds the detection is marginal due to coherence losses but at 12 seconds the detection is quite strong.
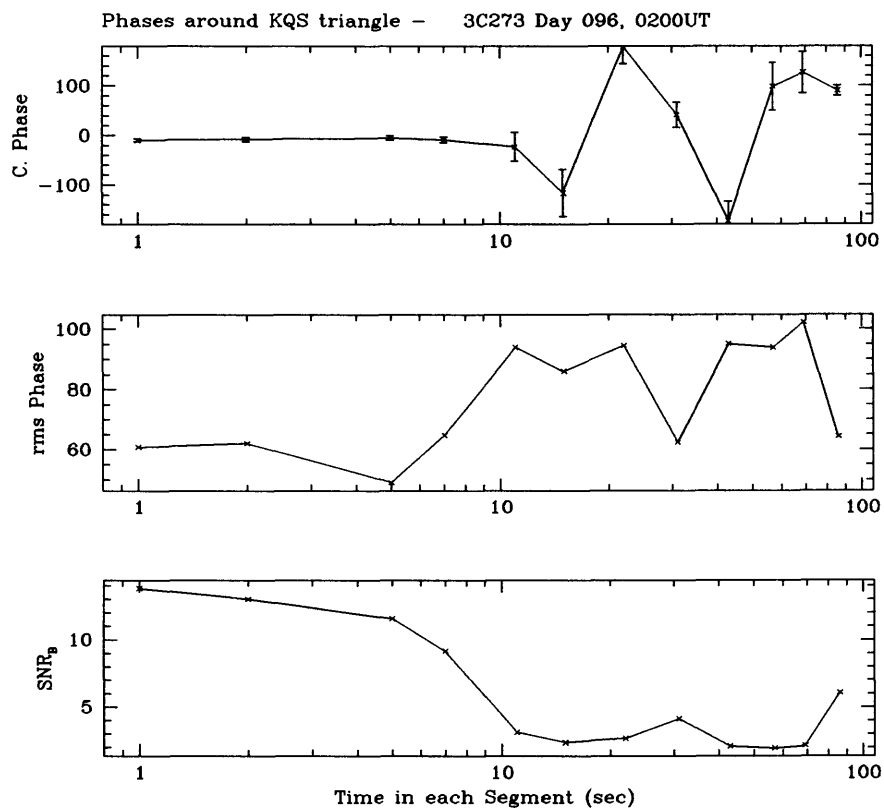
Figure 4-6: Top panel shows closure phase on the Haystack-Quabbin-Onsala triangle a scan on 3C273. Errors are 1-σ, determined empirically by comparing noise and signal contributions from each segment in the bispectral sum (Rogers, et. al., 1994b). Last two panels are the rms phase of all segments in bispectral sum and the SNR_B.
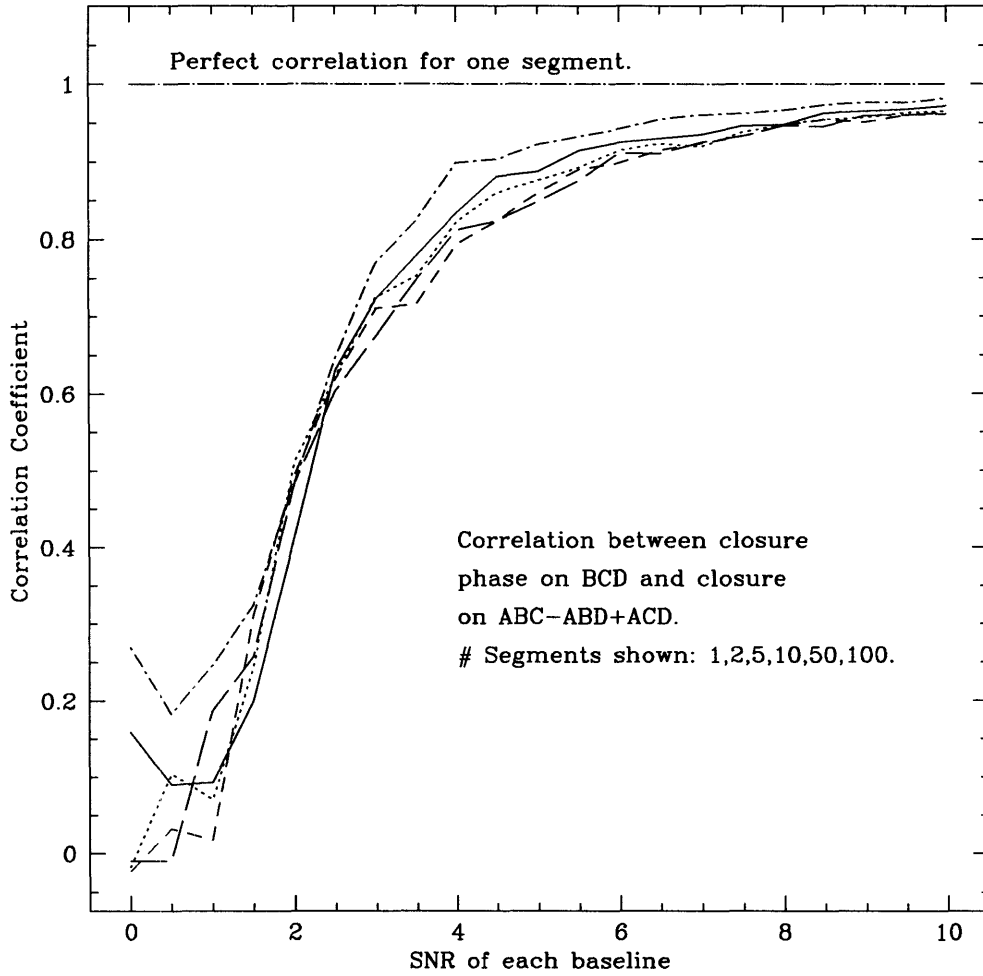
Figure 4-7: Correlation between closure phase on one triangle and the linear combination of closure phases on the other three. For one segment these quantities are the same. When an average is taken, the noise on each triangle is independent and the two estimates of the same closure phase decorrelate at low SNR.

71

Chapter 5

Observations

5.1 Antennas

As previously mentioned, 3mm-VLBI has not always had the freedom to pursue more than a handful of bright compact sources. This limitation, due to sensitivity, is relaxing somewhat but remains the biggest hurdle in the field today. The new detection methods of Ch. 4 lower detection thresholds but to make further progress, system temperatures and dish efficiency must improve. The current 3mm-VLBI array is characterized in Table 5.1 which gives each antenna's SEFD (System Equivalent Flux Density) and other features.

Station	Symbol	Diameter(m)	Efficiency	T_{sys}	SEFD (Jy)
Haystack	H	37	0.07	200	7500
Bonn	B	inner 60	0.13	400	2600
Pico Veleta	P	30	0.55	200	1500
SEST	D	15	0.5	300	9000
Onsala	S	20	0.48	250	4600
Quabbin	Q	14	0.5	400	14400
KittPeak	K	12	0.55	200	9000
OVRO	O	3x10	0.55	300	6400
Nobeyama	N	45	0.4	400	1800

Table 5.1: Antennas used in 1993 April experiment.

If we include the incoherent averaging improvement to sensitivity, then the sensitive Pico-Bonn baseline should be able to detect a source of ~ 0.1Jy. This limit assumes a coherence time of 20 seconds, a 6.5 minute total scan length, and the antenna attributes in Table 5.1.

Source	Alias	90GHz Flux (Balonek) 3/Jan/93	10/Nov/92	Type	z	Scale (h^{-1}pc) $\theta \sim 100\mu$as
0316 +413	3C84	6.6	7.0	G	0.0172	0.24
0851 +202	OJ 287	2.5	3.3	B	0.306	0.28
1226 +023	3C273B	20.7	24.7	Q	0.158	0.18
1253 -055	3C279	15.0	13.0	Q	0.538	0.37
1641 +399	3C345	7.1	7.5	Q	0.595	0.38
2200 +420	BL Lac	2.0	2.1	B	0.0695	0.09
0333 +321	NRAO 140	1.5	1.6			
0415 +379	3C111	4.4	4.4	G	0.049	0.066
0420 -014	—	3.5	4.8			
0458 -020	—	1.3	1.7			
0735 +178	DA 237	1.5	1.6			
0748 +126	—	2.0	2.6	Q	0.889	0.42
1055 +018	OL 093	3.6	—	Q	0.888	0.42
1228 +126	3C274	7.0	6.8	G	0.004	0.006
1335 -124	—	4.9	4.8	Q	0.541	0.37
1510 -089	—	1.4	—			
1730 -130	NRAO 530	5.6	4.3	Q	0.902	0.42
1828 +487	3C380	1.1	0.9			
2145 +067	DA 562	4.5	4.7	Q	0.990	0.43
2223 -052	3C446	2.5	2.9	B	1.404	0.43
2251 +158	3C454.3	10.1	9.0	Q	0.859	0.42

Table 5.2: Sources scheduled for observation 1993 April. B - Bl Lac, Q - Quasar, G - Galaxy. The scale column shows the linear size at source redshift assuming a 100μas beam with q_o=0.5 and H_o=100h kms^{-1} Mpc^{-1}.

5.2 Sources

For the April 1993 global 3mm-VLBI experiment, Table 5.2 shows the sources that were chosen for observation. All of these sources had either been observed at 3mm before or had some characteristic that would suggest detection at higher frequencies. Such characteristics include previous detection with 7mm-VLBI, short timescale variations in the mm, and compact structure at low frequencies with an inverted spectrum.

5.3 Scheduling

The actual scheduling of a five day VLBI experiment is non-trivial. The widely spaced antennas that allow high resolution frequently cannot observe the same source due to mutual visibility considerations. Each antenna also places its own slew speed and sun avoidance

angle constraints on the schedule. Since there are two main antenna groupings (Europe and U.S.), it is natural to subnet antennas so that at any given time, different antennas may be observing different objects. This option keeps fairly isolated antennas such as Nobeyama from remaining idle as a single source is tracked from Europe to the U.S. A schematic of the final schedule is shown in Figs. 5-1 and 5-2 to illustrate the scope of the experiment and the subnetting possible.

5.4 Correlation

The recorded tapes were shipped to the Haystack correlator where station clocks and clock drift rates were determined. Most of the correlation took place on an older model MarkIII system which supported only 4 playback drives. The small number of drives quickly increases the time required to correlate all baselines, especially during periods that include all antennas. The number of correlation passes with n drives and m stations is $m(m-1)/2 \mod n(n-1)/2$.

Preliminary inspection of the data indicated that several antennas had experienced instrumental problems. It is the nature of interferometry that the data on tape is meaningless until correlated with another station. This means that the most useful diagnostics of an acqusition system usually occure long after the experiment is over and must be remembered for the next observations. The Haystack antenna experienced gain fluctuations that had to do with an incorrectly set thermal control system. This affected all of the data but was partically corrected during the calibration. the Quabbin site was plagued by phase noise that was traced to a faulty buffer amplifier on the hydrogen maser. This problem was corrected for the last 6 hours of the experiment but reduced the coherence time for all Quabbin baselines to less than 2 seconds for much of the observing run. Pico Veleta's local oscillator system was also contaminated with phase noise and only the last 12 hours of data was salvaged. No fringes were found to Bonn due to weather. OVRO's local oscillator was affected by power line noise which spread some spectral power out to ±120Hz side lobes. Onsala and Kitt Peak both performed very well. Both Nobeyama and SEST experienced difficulties and produced no fringes.

During the last day of observations five of the antennae were working well enough to collect good quality data. This thesis examines one source observed during this time :
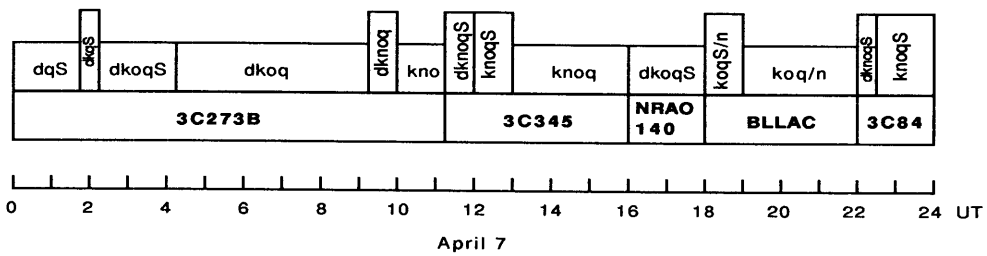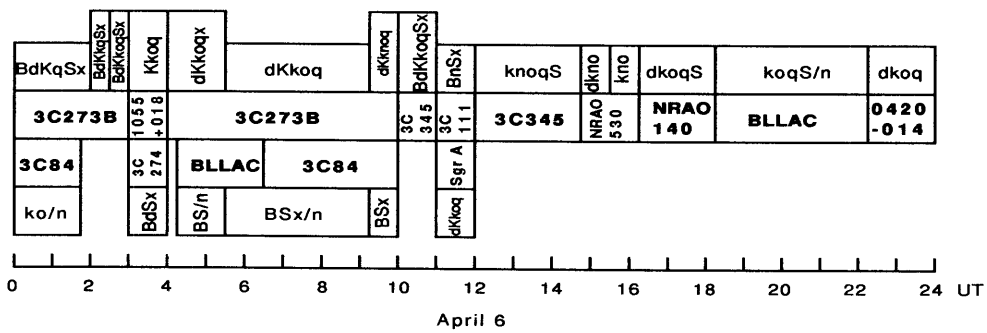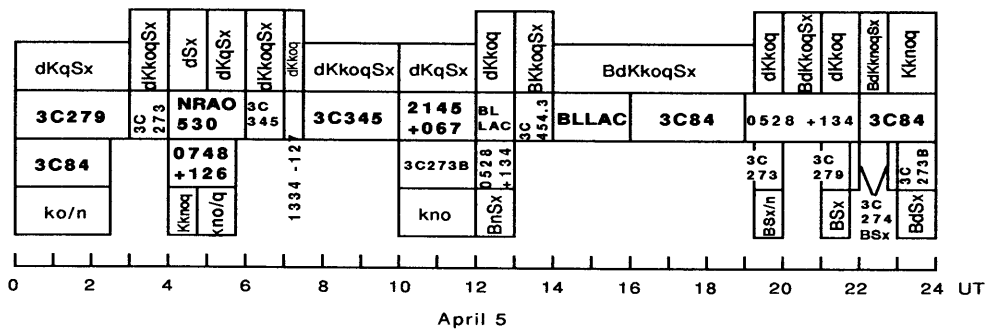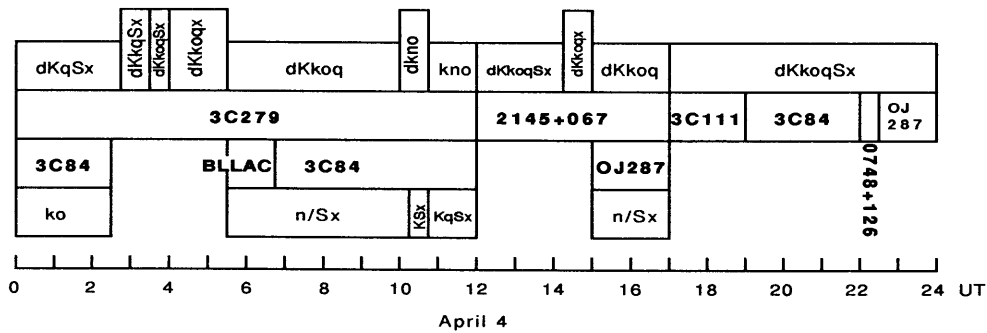
Figure 5-1: First 4 days of April 1993 schedule.
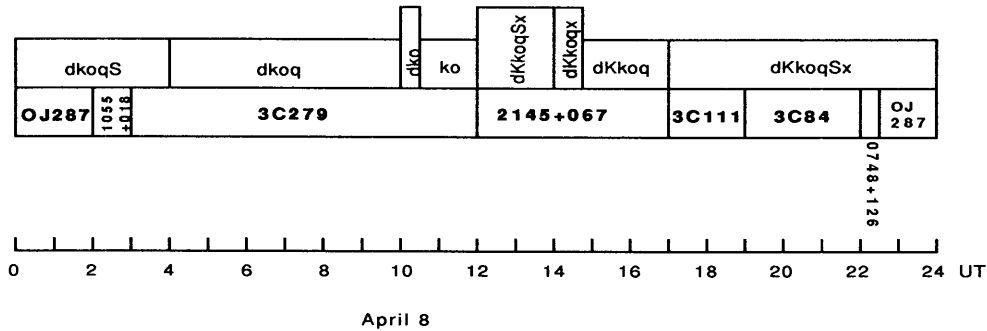
Figure 5-2: Last three days of April 1993 schedule.

3C111. The other sources have all been correlated as well but will require much more effort to calibrate due to the problems outlined above.

As the data was correlated, coherence tests were performed to determine the extent of amplitude reduction due to phase variations. A plot of integrated amplitude vs. coherence time is shown in Fig. 5-3. The amplitudes were obtained by segmenting each baseline and incoherently averaging the segments as in Eq. 4.6 of Ch. 4.

5.5 Calibration

After the signals from each antenna have been correlated there remains one more step before the visibility amplitude is in usable form. Since the correlator cannot know the sensitivity of each antenna or how local weather may have affected the signal, calibration information must be applied to convert correlation coefficient to flux density. If the antenna temperature rise due to the source (T_a) is much less than the system temperature (T_s) then

$$S_{corr} = \rho \sqrt{\frac{T_{s1}T_{s2}}{T_{a1}T_{a2}}} S_{tot} \tag{5.1}$$

76
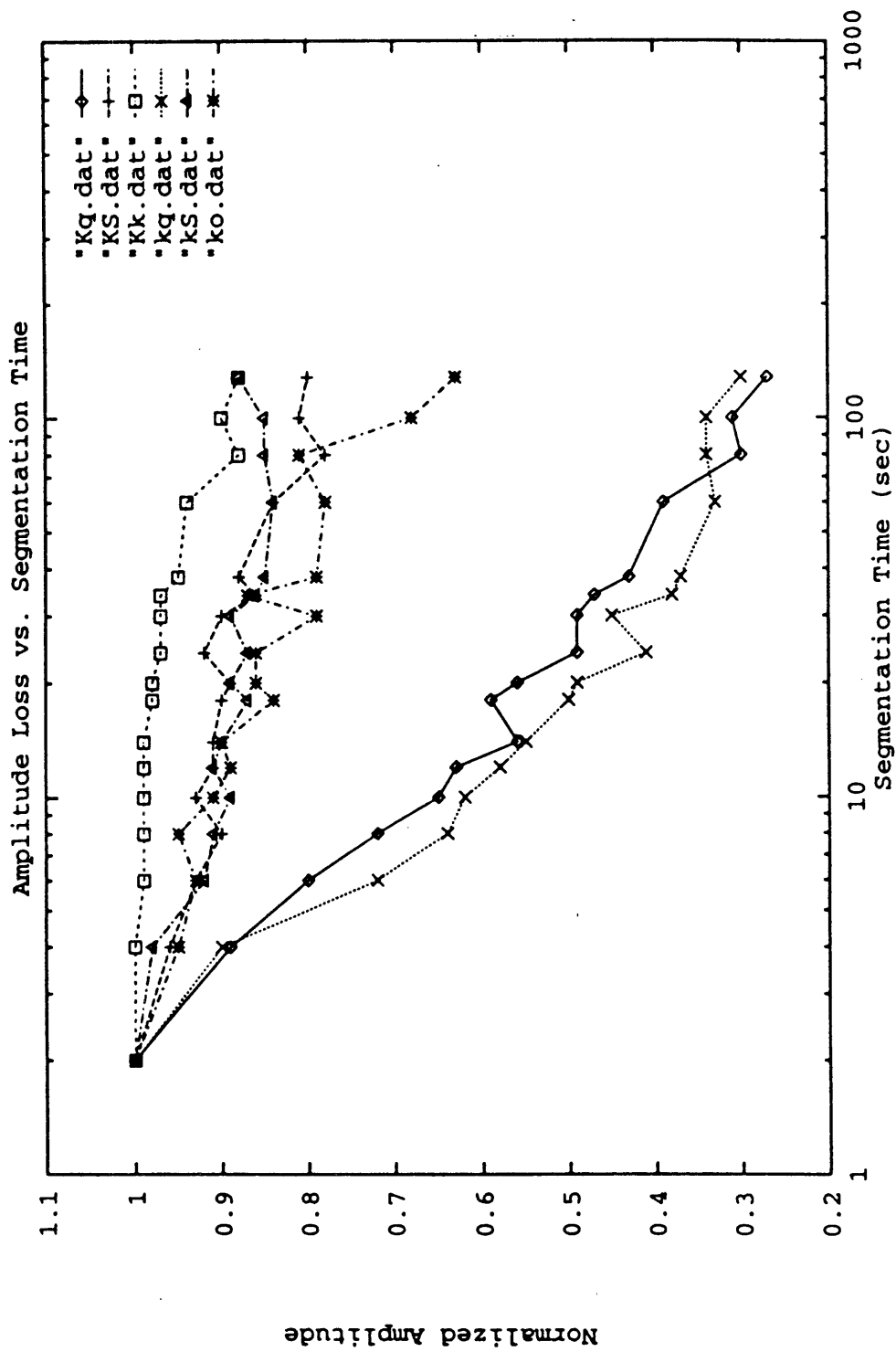
Figure 5-3: Averaged visibility amplitude for various baselines as a function of segmentation time.

with S_{tot} being the total or single dish flux of the source. S_{corr} is also written

$$S_{corr} = \rho\sqrt{\frac{T_{s1}T_{s2}}{K_1 K_2}} \tag{5.2}$$

where K_i is the antenna gain (K/Jy) at each site. The first step in calibration is then to gather a-priori values of system temperature and gain values from each antenna. Gain is typically a function of position on the sky due to telescope structural deformations as the azimuth and elevation are changed. Just before (or even during) the VLBI observations one should track the relative gain on a point source (SiO maser) as a function of elevation. This informs the observer as to the quality of telescope pointing and also yields a relative gain curve. To derive absolute gain the antenna temperature of a source of known brightness is measured and combined with the elevation dependence to give the absolute gain curve. Planets serve as good absolute calibrators having well established effective blackbody temperatures and whose antenna temperatures can be corrected for beam size effects. A current gain curve for the Haystack antenna is shown in Fig. 5-4 which shows the dramatic change in gain as elevation angle reaches its extremes.

To measure the system temperature of a VLBI system the appropriate question to ask is what is meant by the "system" and to what point should such a measure of noise be referenced. The currently favored techinique for mm system calibration is the absorbing vane method which measures the noise contribution from the telescope receiver and the atmosphere simultaneously (Penzias&Burrus,1973). This has the added benefit of directly finding a system temperature referenced above the atmosphere. So the vane-calibration system temperature is the noise temperature that would have to be added to a perfect (noiseless) antenna that was located above the Earth's atmosphere. With T_{sys}, T_{rx}, T_{atm} as system, receiver and atmospheric temperatures respectively and τ the atmospheric attenuation, we can write the desired system temperature :

$$T_{sys} = e^{\tau}\left[T_{rx} + T_{atm}(1 - e^{-\tau})\right] \tag{5.3}$$

To calculate this noise an absorbing vane of carbon foam at the ambient temperature is placed in front of the receiver feed and then removed. These two power measurements will be given by P_{vane} and P_{sky} where P_{sky} is measured with the antenna pointed at empty sky

(off source):

$$P_{vane} = g(T_{rx} + T_{vane}) \tag{5.4}$$

$$P_{sky} = g(T_{rx} + T_{atm}(1 - e^{-\tau})) \tag{5.5}$$

where g is a constant of proportionality and T_{vane} is the vane temperature. We can then form the ratio

$$\frac{P_{sky}}{P_{vane} - P_{sky}} = \frac{T_{rx} + T_{atm}(1 - e^{-\tau})}{T_{vane} - T_{atm}(1 - e^{-\tau})}. \tag{5.6}$$

Now the approximation is made that the vane temperature and the atmosphere temperature are approximately equal which allows the system temperature to be written

$$T_{sys} = T_{vane}\left(\frac{P_{sky}}{P_{vane} - P_{sky}}\right). \tag{5.7}$$

If the temperature of an observed source is desired then an additional power measurement P_{on} made while pointing at the source is required. The resulting source temperature is then

$$T_{source} = T_{vane}\frac{P_{on} - P_{sky}}{P_{vane} - P_{sky}} \tag{5.8}$$

in which the same approximation of $T_{vane} \simeq T_{atm}$ has been made. Note that using the vane-calibration removes the effects of the atmosphere so that independent measurements of atmospheric attenuation are unecessary. But it is still useful to ask how good the underlying approximation is and how the result T_{source} depends on elevation. To this end, one may introduce a difference temperature $\Delta = T_{vane} - T_{atm}$ and evaluate T_{source} including this term. Keeping Δ we can write the new source temperature (T'_{source}) in terms of the source temperature assuming $\Delta = 0$ (Eq. 5.7):

$$T'_{source} = T_{source}\left(1 + \frac{\Delta}{T_{vane}}(e^{\tau} - 1)\right). \tag{5.9}$$

We conclude that if $\Delta > 0$, Eq. 5.7 will underestimate the true source temperature. Fig. 5-5 shows the effect of including this temperature offset. More sophisticated analyses incorporate meteorological data to determine the actual atmospheric temperature profile and its effect on T_{source}.

5.6 Data Reduction

With only ~ 10 mm antennas active in mm-VLBI one would think that the data flow from an experiment would be moderate. Certainly there are nowhere near as many visibilities as for a VLA or 17-station VLBI observation. But the short coherence times at mm wavelengths require each VLBI observation to be segmented, sometimes down to the two second level (see Fig. 5-3). The number of data records grows as the reciprocal of the coherence time. If, in addition, one lowers detection thresholds by generating duplicate sets of offset segments (Section 4.2.2) the number of records doubles. A data pathway was created to reduce these large mm-VLBI data sets which incorporates the ideas in Chapter 4. A block diagram of the scheme is shown in Fig. 5-6.

The first step is correlation in which station clock information is used to help search for detections (coherently) on the brightest calibrator sources. At the time of this writing, the best of these "fringe-finders" are the quasars 3C273 and 3C279, both with flux densities ~ 20Jy at 86GHz. Detections on bright sources refine the station clock and rate offsets so that correlation can proceed on weaker sources. In the first pass of fringe searching a coherent algorithm is used which will usually detect stronger sources, even in moderately bad weather. For scans resulting in non-detections the search windows can be narrowed by enforcing closure of delay around closed loops of baselines (Section 4.1.1). If two sides of a triangle of baselines detect the source, then an estimate of delay on the third undetected baseline can be found (Appendix D;Alef&Porcas,1986). These delay estimates are fed back into the coherent search procedure which may detect the weak baseline in the reduced search space. All scans are then segmented in time which serves three purposes : 1) it allows one to determine the coherece time by segmenting at various timescales; 2) for as-yet-undetected sources, an incoherent search can be performed using the segmented data; 3) amplitude and bispectrum averaging will give estimates of visibility with effects of coherence loss removed. For convenience, the results of incoherent searches are used to set zero-width windows for a final coherent search which is equivalent to evaluating the cross-power spectrum at a fixed delay and rate.

Once averaged amplitudes and closure phases have been determined, estimates of individual baseline phases are found. This step is not necessary in the general sense but most imaging software requires this format. To do this a least squares fit for all baseline phases is

performed using the argument of the averaged bispectrum and its SNR as measured around all triangles of baselines in the observation.
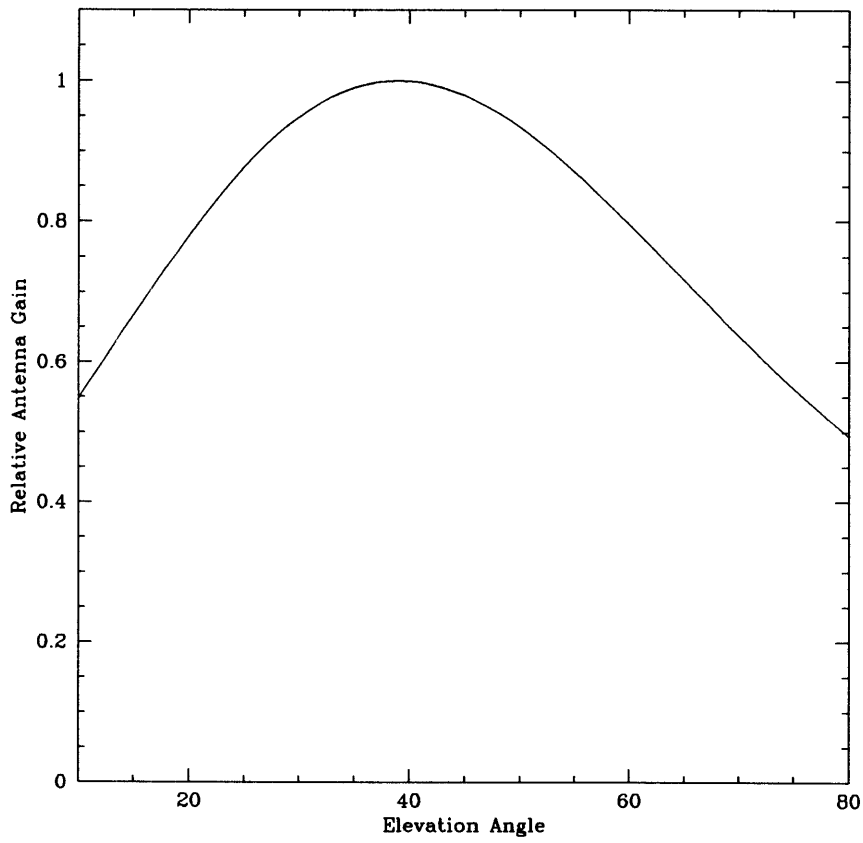
Figure 5-4: Relative Gain Curve for the Haystack antenna at 86GHz.
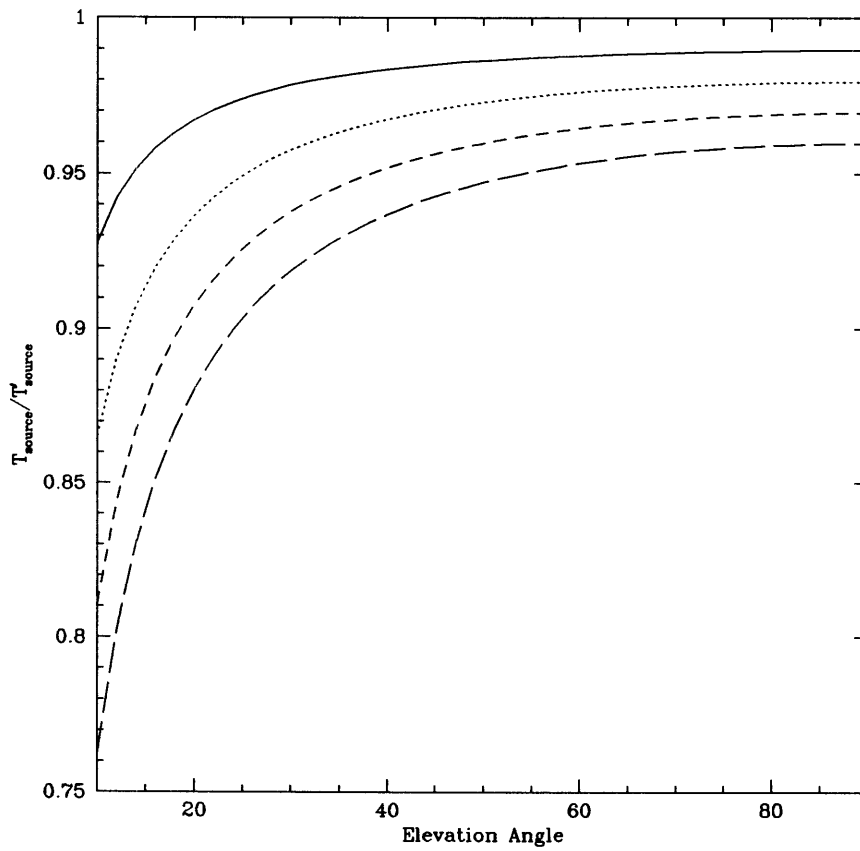
Figure 5-5: Amount by which source temperature is underestimated if atmospheric temperature and calibration vane temperature are unequal. In this plot $\tau = 0.1/\sin(elevation)$ and the curves are plotted for fractional temperature differences (Δ/T_{vane}) of 0.1,0.2,0.3,0.4.
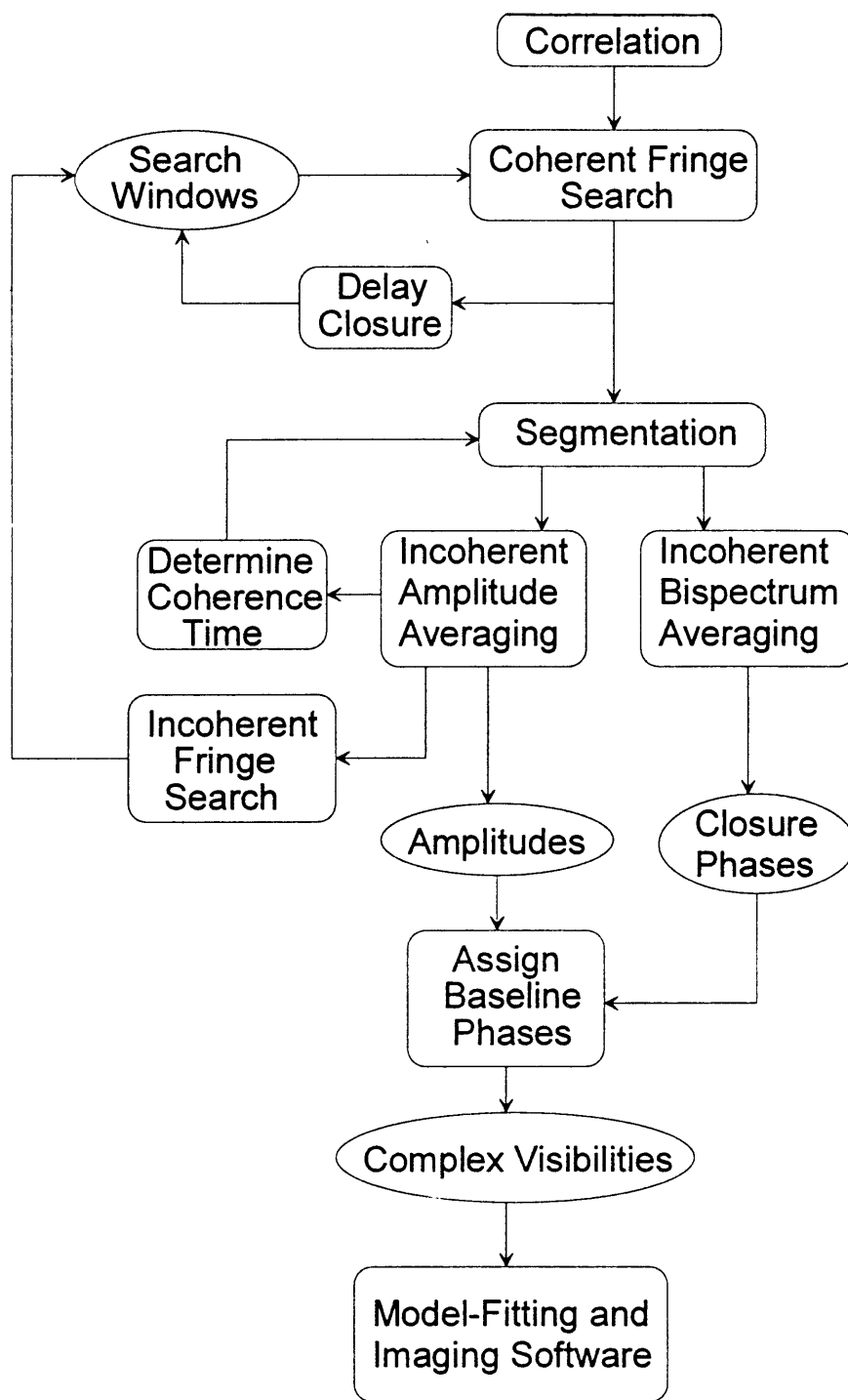
Figure 5-6: Data reduction pathway for mm-VLBI observations. Rounded rectangles are operations and ovals are data entities.

Chapter 6

The N-galaxy 3C111

As part of the April 1993 3mm-VLBI campaign we made a 1.5 hour observation of the N-galaxy 3C111. 3C111 was selected as a 3mm-VLBI target primarily for its high 3mm flux (4.4Jy in 1993) and because of its compact core seen at lower wavelengths. Little is known optically about this source due to its low galactic latitude ($b = -9°$) and its position behind the Perseus molecular complex in our own galaxy. Broadband optical spectra show 3C111 to be highly reddened (Sargent,1977) with an intrinsic spectrum similar to that of a Seyfert I. Its spectrum includes both broad emission lines ($H\alpha, H\beta$) and narrow lines ($[O\,III]$) which suggest that our view of the core is not blocked by a dusty molecular torus (Antonucci & Miller, 1985).

In the radio, much more work has been done and 3C111 stands out as a classical edge brightened FRII source with two large lobes to the SW and NE of a bright compact core. Attention was first drawn to 3C111 when its 90GHz flux was seen to jump in 1975. Wills (1975) used the observed radio spectrum to model the emission as two steep spectrum sources (the lobes) and a compact central component that was optically thick. This general scheme was borne out by later VLBI measurements (Pauliny-Toth et al,1976). VLA images at 5 and 15GHz combined with VLBI maps made at 1.7,5 and 10.7GHz show a well defined jet leading from the core to the NE lobe but no corresponding jet to the SW. The VLBI jet extends from pc to kpc scales with the same position angle as the VLA jet which reaches the NE hotspot some 100 arcsec from the core. The jet appears to be collimated on all scales by some mechanism other than thermal pressure and magnetic confinement has been suggested (Linfield,1987). The broad lines in this source's spectrum support the view that

3C111 is not oriented with its jet in the sky plane and may be oriented near to our line of sight.

FRII sources such as 3C111 with their extended structure can be counterintuitive when one considers their orientation on the sky. Since we see a 3-dimensional universe projected onto a 2-D map, it is tempting to think of these extended sources as lying "flat" on the sky with their jets making an angle of $\sim 90°$ to our line of sight. Abundant evidence, however, points towards relativistic beaming effects being important in some sources.

The current picture of 3C111 is one in which relativistic beaming is assumed to play a role. Evidence for this comes in a variety of forms but all points to a relativistic jet originating in the core oriented at an angle close to our line of sight. As early as 1975, large intensity variations were noted implying brightness temperatures that exceeded the self-compton limit (Hine&Scheuer,1980). This limit marks the point at which cooling by x-rays prevents prevents a further increase in synchrotron emission. The high brightness temperature can be reconciled with this limit by postulating a lorentz factor of a bulk flow in the jet ($\gamma = 1/\sqrt{1-\beta^2}$) of $\gamma \sim 2$. More recently, there are indications from VLBI monitoring at 5GHz that the jet structure is evolving superluminaly with an apparent speed $\beta_{app} \simeq 3h^{-1}$ where $H_o = 100h$km/s/Mpc (Preuss et al,1990). This measurement is preliminary and needs to be repeated but suggests a lorentz factor of $\gamma \sim 3$. Other possible evidence for beaming (though model dependent) concerns 3C111's lack of a counter jet to the SW lobe. If one assumes that two identical but oppositely directed jets emerge from the core, then the observed jet may be doppler boosted towards the observer while the counter jet is boosted away and falls below detection limits. The jet-counter jet ratio of 15:1 at 1.7 GHz (Linfield,1987), requires the jet to be oriented towards us with $\gamma \sim 2.5$.

One aspect of the beaming scenario which has not been analysed for 3C111 is a comparison of observed to predicted x-ray emission. This involves calculating the x-ray flux density due to the synchrotron self compton (SSC) process. In this chapter we begin by reviewing the orientation question and its analysis including the expected SSC effects. After an explanation of the 3mm-VLBI observations, we give results which include a new position for 3C111, a modeled brightness distribution of the source, physical parameters of the core, and the predicted SSC flux. Implications of these results are discussed.

6.1 Orientation

Before VLBI was able to deliver reliable images of the sub-arcsec structure of radio loud AGN, astrophysicists had to content themselves with analysing radio spectra. These can be divided into two main types : steep spectrum sources with $\alpha > 0.7$ (flux density $S_\nu \propto \nu^{-\alpha}$) and flat or inverted spectrum sources with $\alpha \simeq -0.3 \rightarrow 0.3$. Simple synchrotron theory predicts an optically thin power law emission spectrum $S_\nu \propto \nu^{(1-p)/2}$ where the relativistic electron distribution is $N(E) = KE^{-p}$. For steep spectrum sources a single emission region can often explain the observed spectra. For flat spectrum sources, if we keep the synchrotron assumption, something must intervene to keep the spectra "flat". The traditional solution deconstructs the sources into multiple synchrotron components each with a different lower cutoff frequency. In this way, as one component's spectrum begins to drop off in frequency, a new component emerges and so on. The resulting spectrum is essentially flat with bumps and wiggles marking the superimposed peaks of all components.

The typical spectrum of a single component has two distinct regimes that depend on the optical depth of the region to its own synchrotron radiation. When the energy of relativistic electrons becomes comparable to the brightness temperature of the radiation then electrons can begin to reabsorb emitted photons. This process causes the spectral slope below a certain cutoff frequency (ν_m) to be $+5/2$ with an eventual break to the optically thin $(1 - p)/2$ slope at higher frequencies. When VLBI images of compact sources were first made, many such components were found and portions of the spectra sucessfully dissected and ascribed to discrete blobs (Marscher & Broderick,1981).

The zoo of observed AGN comprises many different species that span ~ 11 orders of magnitude in luminosity. Even so, there are good observational reasons for unifying some groups of AGN together. Attempts to explain certain differences in terms of orientation effects have been partially successful (Antonucci,1993). At the center of these "unified" models is the hypothesis that most AGN are intrinsicaly similar. Their differences are correlated with the orientation of their jets to our line of sight. Those sources with jets directed towards us should show enhanced emission characteristics. One aspect of this argument is the effect of relativistic beaming on observed flux levels and proper motion of VLBI mapped components. One class of AGN for which these effects are particularly pronounced are the so called superluminal sources. This name comes from the fact that
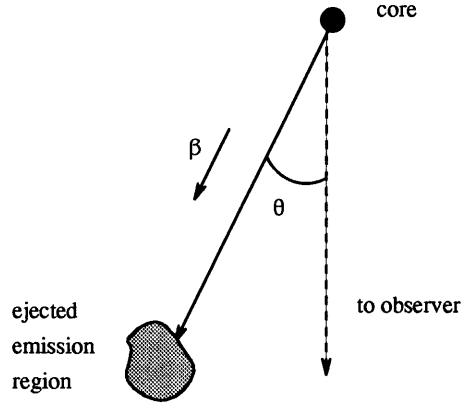
Figure 6-1: A component is ejected from the core moving at speed β at an angle θ from our line of sight.

under certain conditions the apparent speed of components ejected from an AGN core can be in excess of c. The geometry of the situation is shown in Fig. 6-1.

The apparent velocity can be shown to be:

$$\beta_{app} = \frac{\beta \sin(\theta)}{1 - \beta \cos(\theta)}.$$

(6.1)

The emission from an ejected blob will also be doppler boosted by an amount:

$$\delta = \frac{\sqrt{1 - \beta^2}}{1 - \beta \cos(\theta)}$$

(6.2)

and depends on direction and speed as well. Measurement of either one of these quantities leaves an ambiguity in β and θ: they cannot be determined simultaneously. If, however, independent measurements of both β_{app} and δ are available, β and θ can both be localized and the orientation of the jet constrained. Fig. 6-2 shows all of (β, θ) space with lines of constant doppler shift and constant apparent velocity. As it turns out, two independent approaches to estimating δ and β_{app} exist and this method has been used to constrain the source geometry of 3C345 (Cohen and Unwin,1984).

Measurements of both β_{app} and δ are problematic and prone to error. The apparent speed can only come from careful long term VLBI monitoring which follows the evolution of single components that can be unambiguously identified from one epoch to the next. A shift in a component's position is interpreted as bulk movement of a shock down the
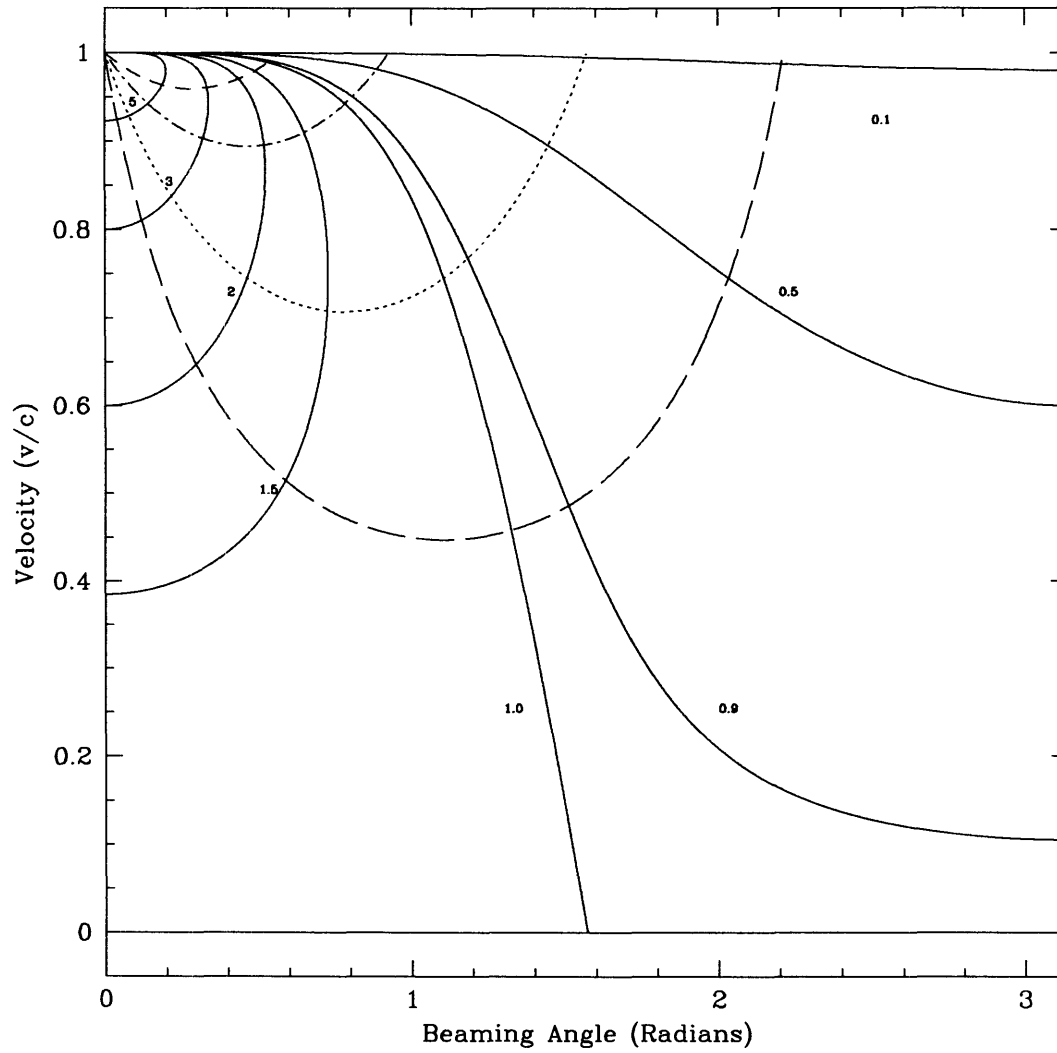
Figure 6-2: The (β,θ) plane with lines of constant doppler boost shown as solid curves with labeled values. The dashed curves are lines of constant apparent speed with values of 0.5c, 1.0c, 2.0c and 3.4c.

jet and the local apparent speed can be measured. Any extrapolation of speeds measured in the jet down to the core must be viewed with suspicion as such speeds are most likely due to propagating shocks and will depend on the local medium (Marscher,1992). Each jet component will typically have its own apparent speed and local doppler boosting factor. A measurent of β_{app} can, however, place an upper limit on beaming angle for the entire jet provided we assume it is relatively straight:

$$\theta_{max} = 2\cot^{-1}(\beta_{app}).$$ (6.3)

Independent determination of δ is harder and can be approached by two methods. The first and least accurate is to monitor the variation in total radio flux density with single dish measurements. Such programs of observation are routinely carried out and long term monitoring projects have been underway in Finland (Terasranta,1992) and the U.S. (Hughes,Aller&Aller,1992) for some time. From these measurements variation magnitude and timescale can be used to estimate the size and brightness temperature of emission regions. The brightness temperature for a flux density variation of ΔS in time τ_{obs} is found by Terasranta and Valtaoja(1994):

$$T_{b,obs} = 5.87 \times 10^{21} \frac{\lambda^2 \Delta S}{\tau_{obs}^2} \left[1 - \frac{1}{\sqrt{1+z}}\right]^2.$$ (6.4)

Should this estimate exceed the compton brightness temperature limit then a lower limit for δ can be set since the observed flux density scales as δ^3. One should note that an angular size derived in this way is dependent on distance and will introduce this dependence into any value of doppler factor found using this method. Values of brightness temperature from variability studies have been calculated for 3C111 by Hine and Scheuer (1980) who find instances in which it exceeds 10^{12}K by a factor of 4.

A second approach uses VLBI observables to estimate the SSC x-ray flux expected from high brightness radio components. When such regions become sufficiently compact, x-rays are generated as relativistic electrons compton scatter synchrotron photons. The doppler boosting factor in a jet artificially inflates the observed brightness temperature and can lead one to expect a higher x-ray flux than is actually seen. This occurs since SSC x-ray production is very sensitive to temperature and does not start until T_{bright} reaches $\sim 5 \times 10^{11}$ (Ghisillini, 1992). The SSC flux density from a homogenous sphere was first

calculated by Gould (1979) and Marscher(1983) included doppler boosting and redshift effects. Marscher's calculation gives :

$$S_\nu^C(E_{keV}) \simeq d(\alpha) \ln(\nu_2/\nu_m) \theta^{-2(2\alpha+3)} \nu_m^{-(3\alpha+5)} S_m^{2(\alpha+2)} E_{keV}^{-\alpha} \left[\frac{(1+z)}{\delta}\right]^{2(\alpha+2)} \mu Jy \qquad (6.5)$$

Here, α is the optically thin spectral index, $d(\alpha)$ is a numerical factor given by Marscher, ν_2 is the upper cutoff for synchrotron radiation, ν_m and S_m are the self-absorption turnover frequency and the flux density there, θ is the component size measured with VLBI or a less accurate method. In some cases the predicted SSC flux density far exceeds the observed x-rays and a $\delta > 1$ is invoked to bring the two values in accord. Since sources other than SSC can also generate x-rays, a lower limit on δ is really what this method yields.

We have attempted to combine the 3mm-VLBI data with other observations in an effort to determine whether or not geometric constraints can be placed on the source 3C111. Definitive measurements of β_{app} were first given in (Preuss et al,1990) as part of that groups monitoring of several nearby FRII galaxies. The latest measurements come from VLBI at 6cm wavelength which show components in the jet to be moving with an apparent speed of $\sim 3h^{-1}$. This is the only proper motion available on this source and we consider it subject to the caveats raised above.

6.2 Observations

The observations presented here ran from 17:30 to 19:00 UT on April 9, 1993 with u,v coverage shown in Fig. 6-3. A total of five antennas were used : KittPeak, Haystack, Onsala, Quabbin, PicoVeleta. All recorded a bandwidth of 56MHz using the MarkIII VLBI system (Rogers et al,1983). Fringes were detected using the coherent search technique on 25 scans; after constrained searching using incoherent averages, the source was detected in 37 scans. Coherence times ranged from 10-40 seconds and each baseline scan of 6.5 minutes was segmented and averaged to calculate visibility amplitudes and closure phases. On the longest baselines, the array fringe spacing of 90μas corresponds to a linear distance scale of $0.06h^{-1}$pc at $z_{3C111} = 0.049$ with $q_o = 0.5$.

3C111 was included in the April 1993 schedule as a test source in that it had not been observed at 3mm before. The 1.5 hour "snap-shot" observation is, of course, not optimal for detailed imaging but u,v coverage was sufficient for modeling. All visibilities were calibrated

91

using system temperatures and gain curves.

3C111 (Apr 1993, 86.241 GHz)

K P Q X S
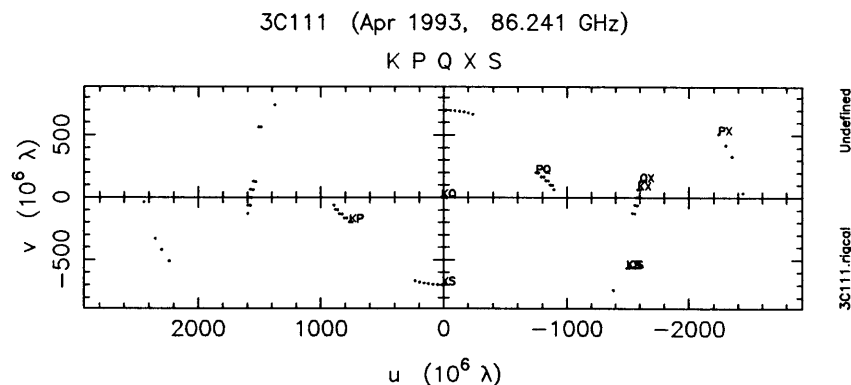
Figure 6-3: u,v coverage for 1.5 hour observation of 3C111.

6.3 Results

6.3.1 Refining the Position of 3C111

Many of the historical targets of 3mm-VLBI stand out as very compact, bright sources at longer wavelengths. For this reason, most have been used as astrometry references to accurately determine baseline lengths and their evolution with time. The use of these sources stems from the fact that astrometric observables depend not only on interferometer geometry but also on source structure. The difficulty in deconvolving effects due to source structure is removed by considering only the most point-like sources and assuming that measured quantities reflect only interferometer geometry.

3C111, due to its FRII classification and strong VLBI scale jet is not included among these reference sources and its position has not been steadily refined over years of geodetic work. During the course of correlating the recent 1993 April experiment, it became clear that the position we had assumed for 3C111 was incorrect as evidenced by high residual fringe rates in the correlator. The fringes, in other words, were not where we expected to find them. Using the observed fringe rates we performed a least squares analysis to find an improved position.

Astrometry makes use of measured fringe rate and interferometer delay to solve for source positions and baseline vectors. The expression for interferometer fringe rate is given in Eq. 2.2 of Ch. 2. If we assume that the baseline vector is well known, then the residual

fringe rate - that observed compared to that expected - can be written as:

$$\Delta\nu_f \;=\; \nu_{in} + \omega_e \left[Y_\lambda \sin\delta\, \Delta\delta + X_\lambda \cos\delta\, \Delta\alpha \right] \cos H \qquad (6.6)$$

$$+\, \omega_e \left[X_\lambda \sin\delta\, \Delta\delta - Y_\lambda \cos\delta\, \Delta\alpha \right] \sin H$$

where ν_{in} is an instrumental term and other quantities are as defined for ν_f earlier. If many sources are observed at a range of r.a. and declination then the instrumental term can be determined at each station and corrected for by introducing a clock drift rate in the correlator. This was done for the 1993 correlation and a position error was suspected as the reason for the anomalously high $\Delta\nu_f$ values that remained on baselines observing 3C111.

We took ν_{in} to be near zero and for each baseline formed the linear constraint given by Eq. 6.6. Each constraint can be graphed as a straight line on the plane of the sky and all lines should converge near one location marking the solution position. Fig. 6-4 shows this $(\Delta\alpha, \Delta\delta)$ plane and the constraint lines determined from six baselines. The point of convergence is well below the origin indicating a clear offset in sky position. The best solution was found by least squares and is marked on the figure.

The formal errors on this offset position are extremely small if one takes into account only the theoretical correlator errors in fringe rate estimation. Atmospheric effects that broaden the fringe frequency spectrum must also be included in the errors, especially at high frequencies. Treuhaft and Lanyi (1987) have calculated the expected deviations in fringe rate assuming a model atmosphere which includes the effects of turbulence. They find that for short integration times ($\sim$ 100sec), the standard deviation of fringe rate can be 0.2ps/s - by far the dominant contribution to random error. There may also be systematic errors contributing to the error which arise from incorrect clock rate values or baseline errors. To get a combined estimated of all errors we examined the typical delay rate errors on scans for which the source position was well known. These errors vary from 0.1 to 0.3 ps/s and we have conservatively used 0.4ps/s in the least squares fit. This yields a final value for the offset of $\Delta\alpha = -0.76 \pm 0.09$as and $\Delta\delta = 0.05 \pm 0.05$as which makes the new position for 3C111 :

$$\alpha(J2000) \;=\; 04^h\, 18^m\, 21^s.28 \;\pm\; 0.01$$

$$\delta(J2000) \;=\; 38^\circ\, 01'\, 35''.76 \;\pm\; 0.05$$

Least Squares Solution for 3C111 Offset

r.a. offset in Arcsec

Sx ———
kq – – –
kx ·······
qx ·········
Kx –·–·
qS –··–··
Solution ◊

Solution :
r.a. offset = -0.76 as
dec. offset = 0.05 as

Dec offset in Arcsec
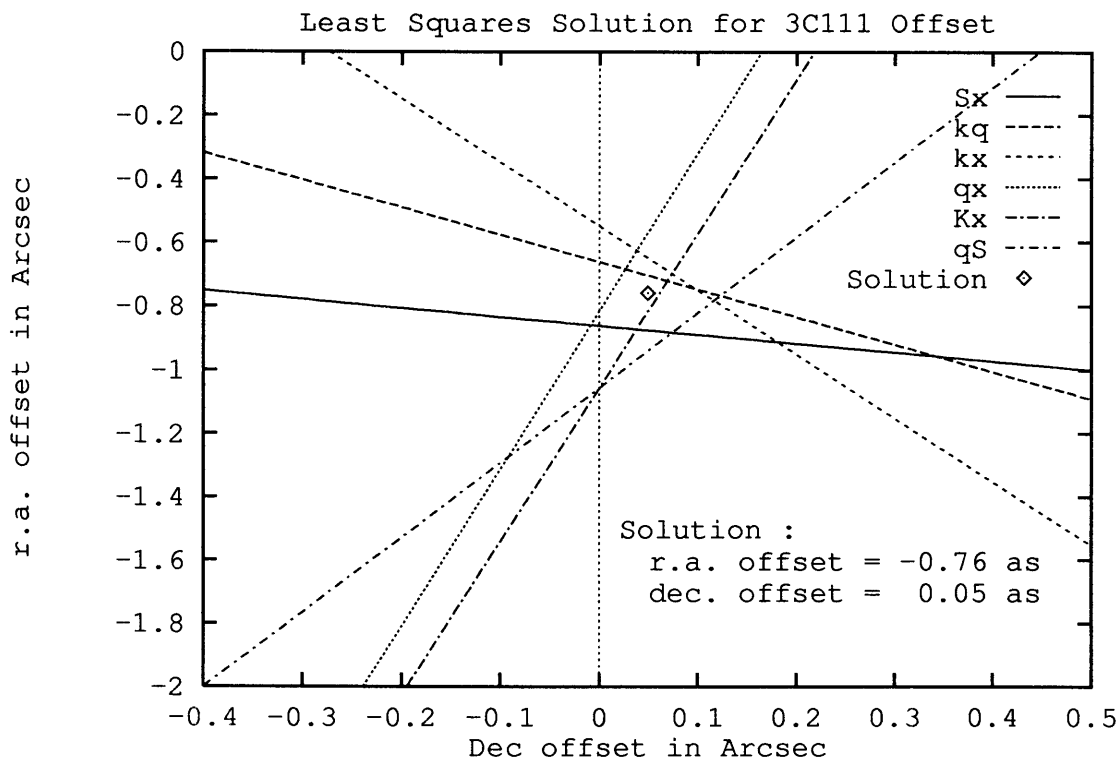
Figure 6-4: Offset solution for 3C111.

6.3.2 Modeling the Brightness Distribution

The closure phases (Fig. 6-5) do not deviate significantly from zero and this led to consideration of symmetric models. The simplest of these would be a single circular gaussian brightness distribution but a fit to such a model was poor ($\chi^2_\nu = 2.05$). Examination of the visibility amplitudes showed that the correlated flux observed on the longest baselines required that there be a very compact core in addition to the larger component that was used to fit data on the shorter baselines. A new model consisting of a circular gaussian component and a central unresolved source fit quite well with a minimum reduced chi-squared (χ^2_ν) of 0.95. There was no compelling reason to leave the simplicity of this three parameter model and introduce as a fourth parameter the size of the central component. A straight foward F-statistic test (Bevington,1969) to determine the need for such a fourth parameter yielded an F=2.3 for 34 degrees of freedom. F exceeds this value with a probability of $\sim$ 10% and thus indicates no definitive need to add a size parameter for the second brightness component. This contrasts with a similar test done for the validity of adding the second unresolved component to the original single gaussian model. In this case, the

94
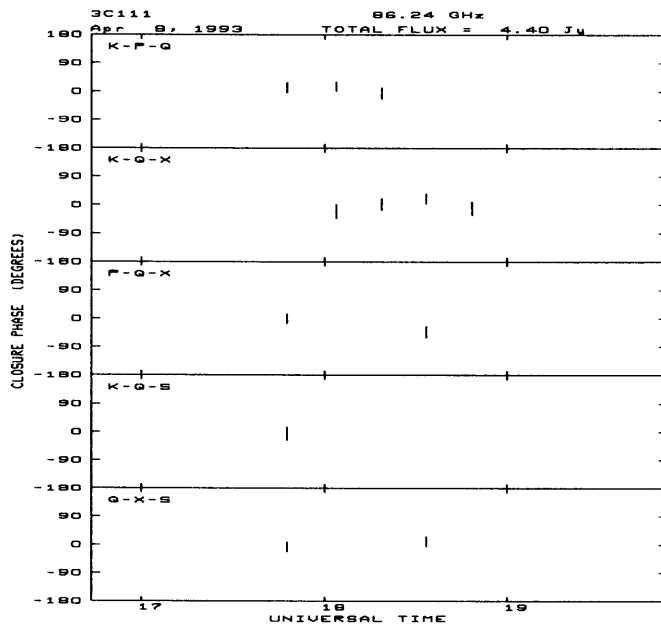
Figure 6-5: Closure phases for triangles in 3C111 observation.

F-statistic is 19.4 and the probability of F exceeding this value is vanishingly small which clearly suggested that a second component was necessary. Table 6.1 shows the values of parameters for the best fit of all three models. A plot of correlated flux vs. baseline length is shown in Fig. 6-6 along with the best fitting of all three models.

Model	χ^2	χ^2_ν	S_1 (Jy)	θ_1 (μas)	S_2 (Jy)	θ_2 (μas)
1 Gaussian	71.75	2.05	2.6±0.23	87.9±6.0	--	--
1 Gaussian + 1 Point Src.	31.9	0.94	2.77±0.45	121.0±19.0	0.36±0.11	--
2 Gaussians	29.7	0.9	$2.67^{+0.1}_{-0.7}$	$132.0^{+20.6}_{-11.0}$	$0.58^{+0.7}_{-0.22}$	36.1^{+30}

Table 6.1: Best fit parameters of brightness models for 3C111.

All quoted errors in Table 6.1 are $2\sigma = 95\%$ confidence contours for the appropriate number of free parameters in each model. The exception to this rule is that all errors cited for the double gaussian model are the acceptable ranges for each parameter as θ_2 is varied from 0μas to its upper 2σ limit of 66μas. The 4 parameter model is quite insensitive to the θ_2 parameter as might be expected from the good fit of the point source model. Another way of stating this is that the confidence contour value when $\theta_2 = 0\mu$as is only 30%.
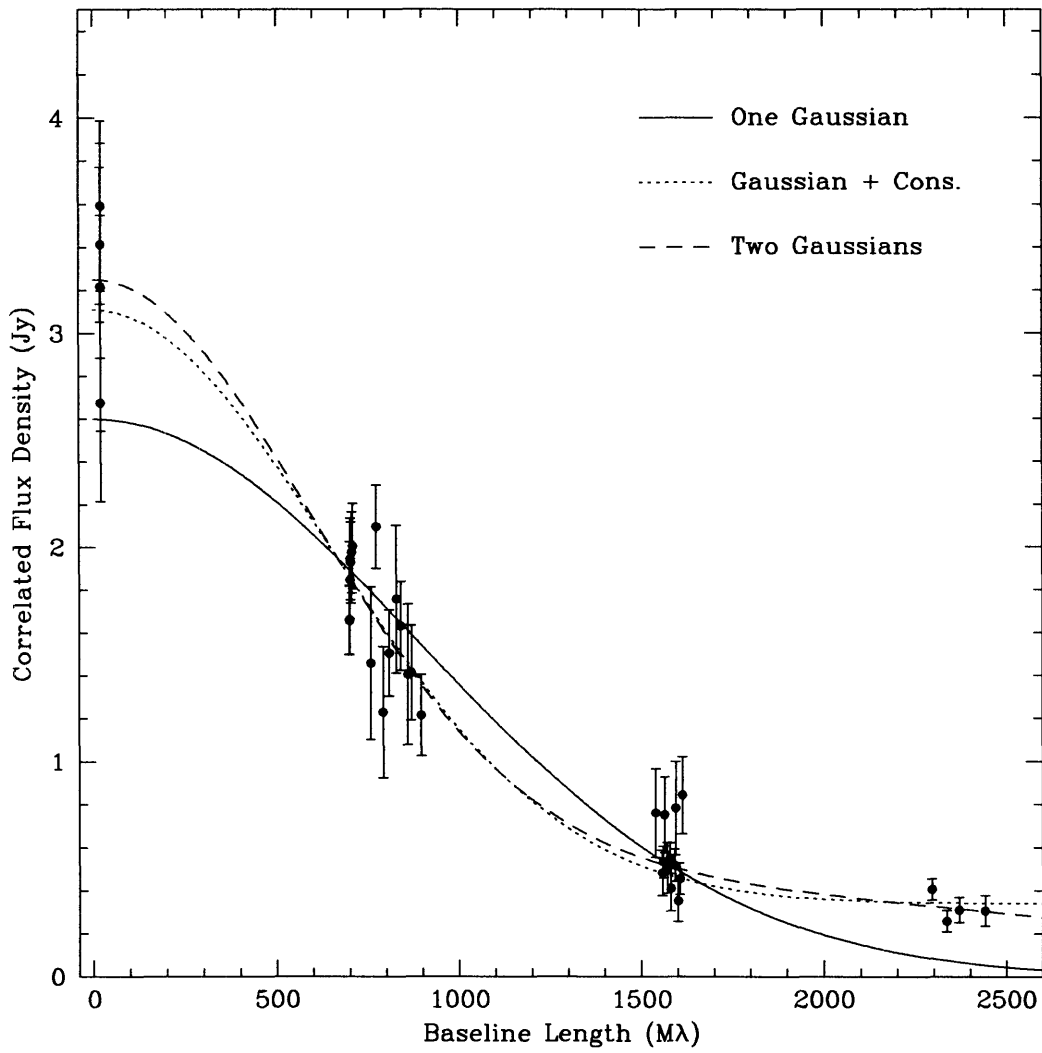
Figure 6-6: Visibility amplitude vs. Baseline Length for observations of 3C111. Three models are fit to the data : a single gaussian component, a gaussian plus unresolved source, two gaussian components.

The data are undeniably sparse but it is still valid and important to ask how a more complicated model which includes a jet might be constrained. Exploring this issue may also allow us to link the large scale jet to sub-parsec scales. To this end, we have constructed a "toy" jet model consisting of the central compact component and an offset elliptical component as shown in Fig. 6-7a. The model shown has been optimized by fitting to the data in the u,v plane which is shown along with amplitude contours of the nodel in Fig. 6-7b. We note that orientation of the jet in the u,v plane is perpendicular to the elongated structure in the image plane as expected from the fourier transform relation between the two domains. U,V plane contours illustrate the lack of sampling in two important directions : along the jet and perpendicular to it. Ideally, one would at least require good u,v coverage in both directions to rightfully claim any kind of source elongation.
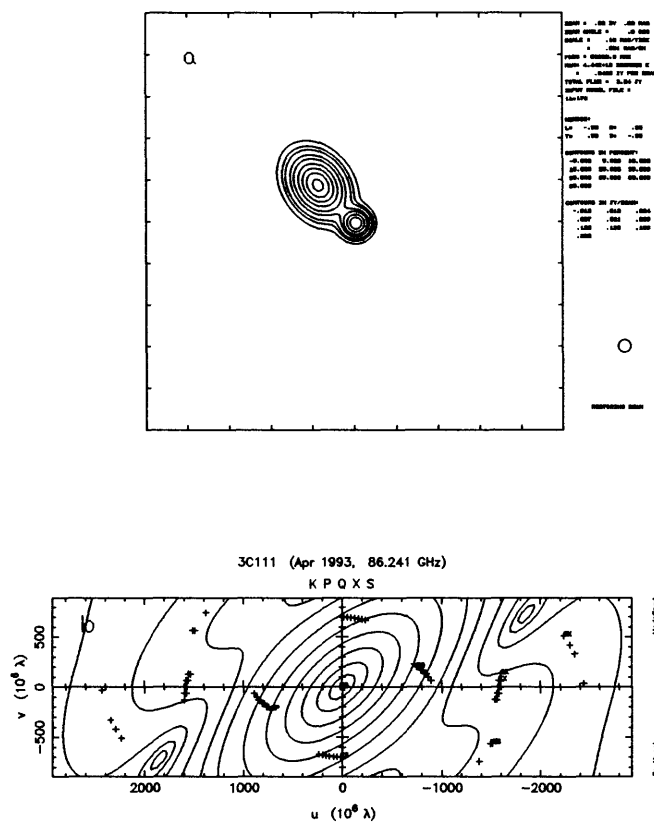
Figure 6-7: (a) model with jet for 3C111. (b) uv coverage of observations with contours of correlated flux for model superimposed. Contours are 3.0, 2.9, 2.7, 2.4, 2.0, 1.5 1.0, 0.7, 0.4, 0.2, 0.1, 0.05 Jy.

Taking this model as representative of most jet models, constraints on jet position angle can be addressed. The relevant parameters in this model include size and flux for both

components as well as separation, position angle, axial ratio and orientation for the jet component. By varying all of these parameters one finds that the jet position angle is best fit at an angle of $45 \pm 28°$ with the error being 2σ. This result brackets the 62° position angle of the VLA and 6cm-VLBI jets but we stress that this is model dependent. It is highly probable that other models (perhaps more complicated) will fit the data as well but exhibit much different brightness distributions.

Future observations of this source will probe the u,v plane in much more detail. To give some idea of what these efforts might find, a simulation of such observations using the model presented here are shown in Fig. 6-8. Wide excursions in both amplitude and closure phase are predicted and will allow us to make a direct estimation of this model's validity.

6.3.3 Physical Parameters of the Core

Though the data present a sparse view of the (u, v) plane and only a simple model can be fit, estimates of the physical environment inside the core can be made. Some estimates such as brightness temperature depend only on the fit of the model. Others including magnetic field rely on assumptions which will be stated as they arise.

Brightness Temperature

Without assuming anything more than the emission being shifted by cosmological and beaming factors, the intrinsic brightness temperature is given by

$$T_{b,intr} = 1.77 \times 10^{12} \frac{S}{\theta^2 \nu^2} \left[\frac{1+z}{\delta} \right]^{1+\alpha} \tag{6.7}$$

where θ is the angular extent in mas, S the flux density in Jy and ν the observation frequency in GHz. Note that this expression, which can be found in (Ghisillini et al,1992), takes the solid angle subtended by the source to be $\pi\theta^2/4$. If source modeling determines the total flux and the FWHM angular size then the following expression is appropriate to find the maximum intrinsic brightness temperature

$$T_{b,max} = 1.22 \times 10^{12} \frac{S}{(\theta_{FWHM})^2 \nu^2} \left[\frac{1+z}{\delta} \right]^{1+\alpha} \tag{6.8}$$

If one examines the ratio of expected x-ray emission due to comptonisation of syn-
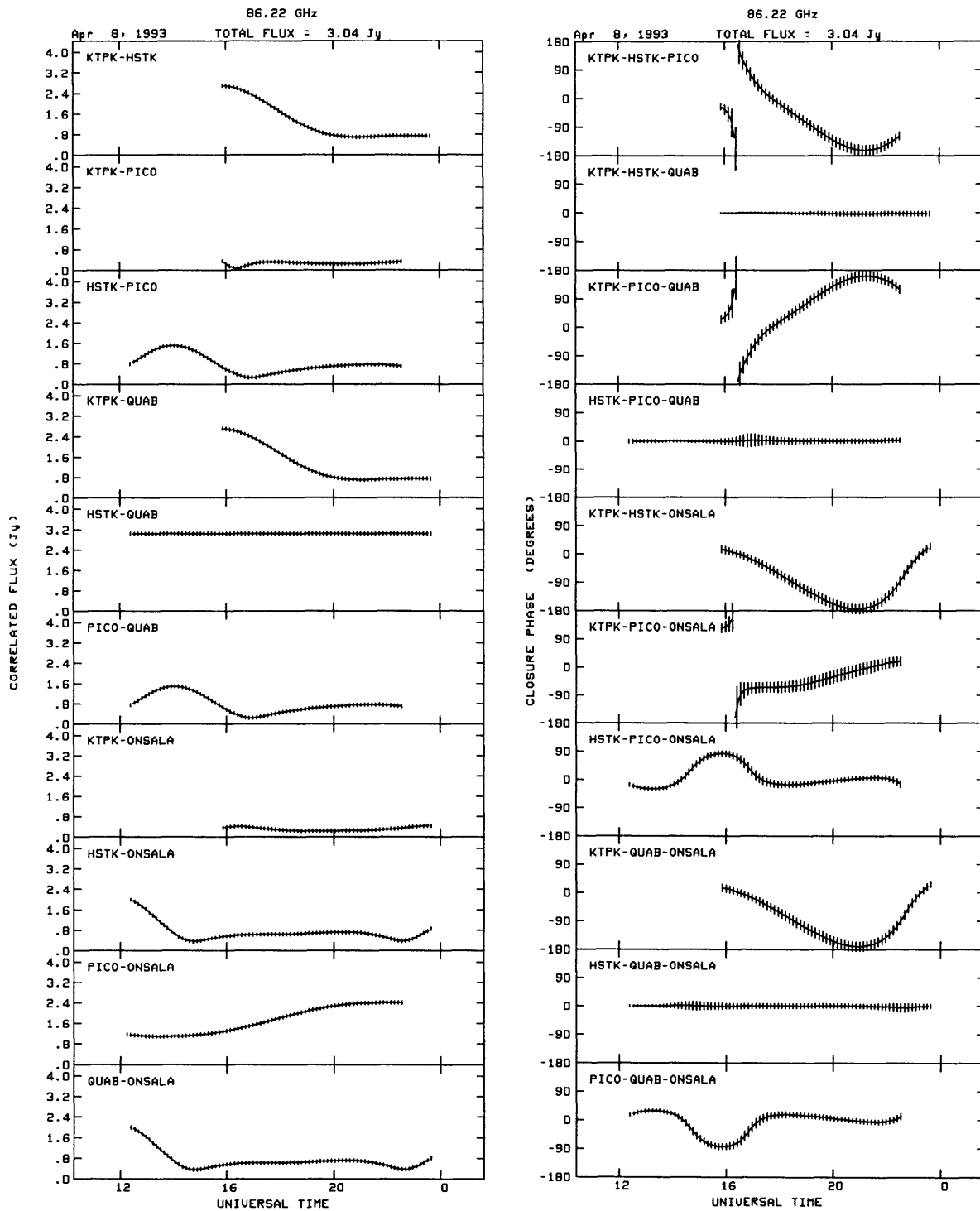
Figure 6-8: Visibility and closure phase for a model of 3C111 with a jet structure. The model shown consists of two components. The core component has FWHM of 38μas with 0.5 Jy flux. The elliptical jet component has a flux of 2.5 Jy, a center 130μas from the origin, a position angle of $45°$, a major axis of 100μas, an axial ratio of 0.3 and an orientation of $44°$.

99

chrotron photons to the radio emission a sharp temperature dependence emerges : $L_{ic}/L_{synch} \propto$ $(T_b/10^{12})^5$. Once T_b approaches 10^{12}K, the half power cooling time becomes proportional to temperature as: $t_{\frac{1}{2},cool} \propto T_b^{-2}$ (Readhead,1993). These dependencies paint a scenario in which a brightness temperature exceeding $\sim 10^{12}$K "triggers" a release of x-rays in a relatively short period of time to cool the source. Table 6.2 tabulates $T_{b,max}$ for each component in the three models (point sources excluded). None of the components exceeds the compton brightness limit except possibly for the small component in the two gaussian model. This component would have to be $< 10\mu$as in diameter for $T_{b,max} > 10^{12}K$.

Equipartition B-field

In radio components emitting synchrotron radiation there are two main sources of energy density: the relativistic particles and the magnetic field. The assumption of energy equipartition between the two is quite close to minimization of the energy and enables us to calculate a magnetic field inside each component. Readhead (1993) has reviewed three high resolution radio surveys and concludes that in most sources, equipartition is obeyed. Appendix E derives the expressions for the equipartition B-field and the associated quantities: total energy, minimum pressure and synchrotron electron lifetime. The other main assumptions that affect these values are that the synchrotron spectrum for each component extends from 10GHz to 300GHz with a spectral index of 0.5. Motivation for these values is explained in Appendix E which also shows the effects of varying the cutoff frequencies and spectral index for a specific example. The derived physical parameters are given for each component in Table 6.2.

Model	Max T_B (K)	B_{eq} (gauss)	U_{Teq} (erg)	P_{min} (dyne cm^{-2})	V (cm^3)	τ_{synch} (yrs)
1 Gaussian	5.6×10^{10}	1.35	8.6×10^{50}	0.08	6×10^{51}	0.073
1 Gaussian + 1 Pnt. Source	3×10^{10}	1.05	1.4×10^{51}	0.05	1.5×10^{52}	0.11
2 Gaussians	2.5×10^{10}	0.95	1.5×10^{51}	0.04	2×10^{52}	0.12
	7×10^{10}	1.9	1.2×10^{50}	0.16	4×10^{50}	0.04

Table 6.2: Physical parameters in core based on three best fit models.

SSC flux

To calculate the predicted SSC flux density we have had to make some assumptions due to a lack of spectral information which would allow us to find ν_m. These assumptions are made in such a way as to reduce S_ν^C as much as possible so that a firm lower bound may be found. Accordingly, ν_m is set to our observation frequency of 86GHz and S_m is then our observed flux density. Note that these assignments reduce SSC flux as much as possible and, at the same time, make 86GHz the turnover frequency to the optically thin regime. A spectral index of 0.5 is assumed which corresponds to an electron energy power law of slope -2. This is not an unreasonable value for components so near the core. Linfield (1987) adopts $\alpha = 0.6$ for 3C111's mas scale jet and one would expect the index to be no greater at the scales examined here. The SSC x-ray estimate is made even more conservative by increasing all angular sizes by a factor of 1.8. This accounts for the fact that the visibility function of a gaussian brightness distribution is very near to that of an optically thin sphere of almost twice the size (see Marscher&Broderick,1981). For comparison, the x-ray flux density at 1keV was measured by Marscher&Bloom (in press) to be $\sim 8\mu$Jy. Results of the SSC calculation for the double gaussian model are given in Table 6.3.

Component	$S_\nu^C(1keV)$ $(\mu\mathrm{Jy})$
1	$2.5 \times 10^{-4}/\delta^5$
2	$4 \times 10^{-3}/\delta^5$

Table 6.3: Estimated SSC x-ray flux density levels for both components in double gaussian model for 3C111.

From these results it appears that SSC production does not by itself place a lower limit on the doppler boosting factor. It is potentially significant though, that the smaller component in 3C111 (component 2) yields an SSC flux density only three orders of magnitude from the observed level despite our efforts to minimize the predicted values. The size of the exponents in Eq. 6.5 are such that even a modest change in peak frequency or angular diameter would drastically increase S_ν^C. If, for example, θ decreased by a factor of two and ν_m shifted to 43GHz (with a resulting shift in S_m), then we find a predicted SSC level of $500/\delta^5 \mu$Jy which would imply a $\delta \simeq 2.3$. Fig. 6-9 illustrates the extreme sensitivity of the estimated x-ray flux density to component size by graphing the SSC contributions of the two component model as the size of the smallest component is varied. It is worth repeating

that the best model fits with $\theta_2 < 36\mu$as are all within a 30% confidence contour of the chi-squared minimum.

6.4 Discussion

6.4.1 The Jet

With these limited observations there is no evidence for asymmetry on sub-parsec scales - no continuation of the kpc and pc-scale jet to the high frequency core. One possibility for this may be the timing of our observation. The lifetime for sychrotron electrons emitting at 86GHz can be written as :

$$\tau_{synch} = 0.092B^{-\frac{3}{2}} \text{ years.} \tag{6.9}$$

Assuming an order of magnitude magnetic field of ~ 1 Gauss the τ_{synch} is about 1 month. Observations within a month of 86GHz flux changes may show the innermost jet structure just as material is ejected from the core. A sample core-jet model has been shown to fit the available data but any conclusions drawn from such a structural assumption are very model dependent. The sensitivity of the array is sufficient to detect components down to a ~ 0.1Jy level. This limit comes from the most sensitive baseline (PicoVeleta-Onsala). Therefore, signatures of the sample jet modeled in this chapter should be observable in future observations. 3C111 is a relatively high-declination source and longer observations with the same stations will fill out the u,v coverage considerably.

One can look at the model from a different standpoint and take the two observed components of the symmetric model to be the start of a jet. If this is so, then we can estimate properties of the surrounding medium needed to thermally confine the jet on small scales (see Linfield&Perley (1984)). For a gas that is 70% hydrogen we can write :

$$\frac{P_{min}}{1.95} < n_e kT \tag{6.10}$$

where n_e and T are the density and temperature of the confining medium. Using a value of $P_{min} = 0.1$dyne/cm^{-2} we can constrain the temperature and density : $n_e T > 6 \times 10^{14}$. Since the size scales of the modeled components are of order $\sim 10^{17}$cm it is natural to ask whether the broad line region (BLR) itself might serve as the collimating agent. The

BLR consists for the most part of a hot ionized gas within which are embedded the dense line emitting clouds. Hot wind models of the BLR (Weymann et al,1982) place the gas temperature near 10^9K which implies a density of $n_e > 10^5$. This density is comparable to that of the BLR and suggests that on the smallest jet scales, confinement by the BLR is possible. This is in contrast to the large scale jet structure for which it has been shown (Linfield&Perley,1984) that thermal confinement is not viable.

6.4.2 Orientation

The orientation of 3C111 can be estimated in a number of ways. Superluminal expansion at $\beta_{app} \sim 3h^{-1}$ provides a limit on beaming angle of $\theta < 36°$ using Eq. 6.3. But this value of β_{app} requires confirmation. The need for a second apparent speed measurement is best emphasized by noting that cavalier interpretation of past VLBI data on this source can lead to an estimate of $\beta_{app} > 9.6$ (Preuss et al,1987). Using the jet-counter jet ratio (R) also leads to an upper limit via the relation:

$$R = \left[\frac{1 - \beta\cos(\theta)}{1 + \beta\cos(\theta)}\right]^{2+\alpha}. \tag{6.11}$$

The ratio of 15:1 found by Linfield (1987) implies the jet orientation is within $55°$ of the line of sight. This method though is much more dependent on imaging techniques and relies on map dynamic range to set the counter jet limit.

Addition of the 3mm-VLBI results of SSC flux are inconclusive. The careful minimization that ensured a firm lower limit to δ reduced the expected x-ray flux far below the observed levels. This may mean that the mm-VLBI components are not responsible for the x-ray flux in 3C111. Since many processes contribute to x-ray flux, such as accretion disks and hot gas, this conclusion is reasonable. If we imagine that the model parameter space allows a $\delta \sim 2$ from SSC constraints then $\theta < 27°$. The model fits obtained here certainly admit this possibility but the issue cannot be pursued with the sparse data in hand. For a source as extended as 3C111, the large deprojected length at small θ can become larger than the biggest known radio galaxies. At $\theta < 6°$ its intrinsic length would be over 2Mpc and make it the largest radio galaxy in the sky.
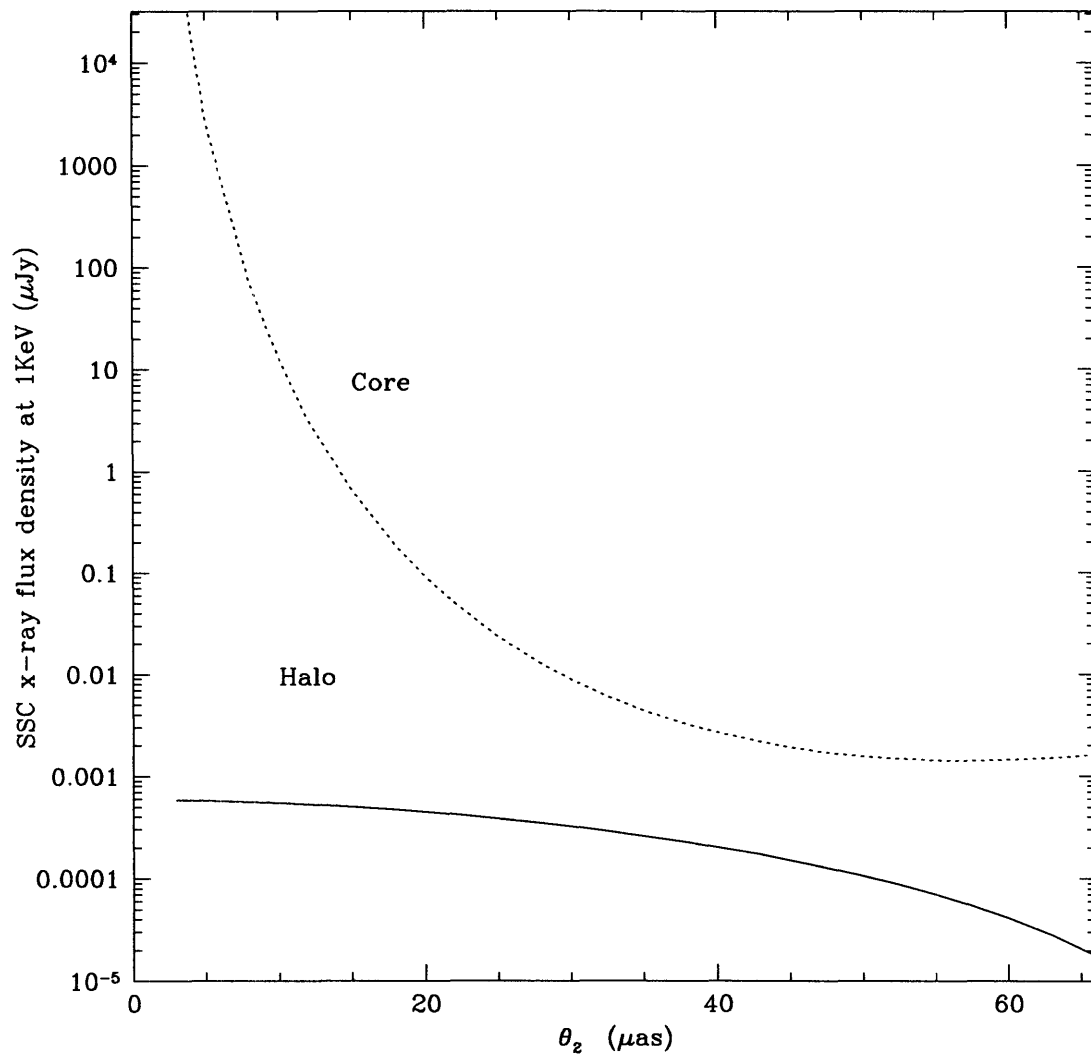
Figure 6-9: Estimated SSC x-ray emission from both components of a double gaussian model for the 3C111 brightness distribution. Size of the smaller component (Core) was varied from an upper 2-σ limit of 66μas down to 3μas with a new fit made at each value. The larger component is labeled as Halo. An optically thin spectral index of 0.5 is assumed.

Chapter 7

The Galactic Center

Study of AGN in other galaxies and at high redshifts was already well underway when indications that such an object might reside in our galaxy were noted. This is due mostly to the frustrating circumstance that lines of sight to the galactic center are obscured by a dense interstellar medium. The ISM permits only radio, IR and high energy photons ($h\nu > \sim 3\text{KeV}$) to travel unabsorbed and unscattered to us from points at low galactic latitude. But the existence of a compact non-thermal radio source in Sagittarius A was interpreted early on (Lynden-Bell&Rees,1971) as a massive black hole : the same model being applied to distant extremely luminous AGN. This source, SgrA*, was observed to be uniquely compact ($\theta < 0.1''$) and bright by Balick and Brown (1974) which supported the argument that activity in nuclei of other galaxies was taking place on a smaller scale in our own. Since that time, observations at many wavelengths and with varied resolutions have attempted to distinguish between competing models of emission in the galactic center. In this chapter a new look at SgrA* with the angular resolution of 3mm-VLBI is described and compared with predictions of emission models.

7.1 Background

The morphology of the galactic center on scales that concern these observations is quite complex and extends beyond the immediate vicinity of SgrA*. Within 10pc there are three main emission regions : SgrA East, SgrA West, and a ring of molecular emission known as the circumnuclear disk (CND). The shell-like structure of SgrA East extends over $\sim$ 8pc and emits a non-thermal spectrum which has been attributed to an internal power source

of supernovae. Its location behind the galactic center has been confirmed by absorption features in the shape of foreground objects (Goss et al,1988). SgrA West has the appearence of a mini spiral and consists of ionized gas emitting a thermal spectrum. Velocity studies of the "spiral arms" indicate that they are tidally stretched clouds infalling to the dynamical center (Serabyn&Lacey,1985). Surrounding SgrA West is the CND which shows up clearly in the HCN $(1 \rightarrow 0)$ transition - a tracer of dense molecular gas. This ring which measures $\sim$ 2pc wide, may be fed by outflow from SgrA East (Zylka et al,1994). The object of our attention, SgrA*, is located within SgrA West $\sim 1''$ from the stellar cluster IRS16. An outflow generated in IRS16 from stellar winds reaches velocities of 500-700km/s with total mass loss rate of $3-4 \times 10^{-3} M_{\odot}$/yr (Melia, 1992;Yusef-Zadeh&Melia,1992). In some models of emission this wind fuels a massive black hole in SgrA*(see Section 7.5).

SgrA* itself remains a mystery since it remains unresolved. VLBI at wavelengths ranging from 1 to 86GHz finds its size to be consistent with that expected from diffractive scattering of a point source in the ISM (see Section 7.4). Only with these VLBI studies can we unambiguously identify observed flux with the central compact source. This implies that only the radio luminosity of SgrA* is well constrained ($L_{radio} \simeq 2 \times 10^{34}$erg/sec). Still, upper limits on total luminosity are accessible if one assumes that all optical and UV emission form SgrA* gets reprocessed by the local ISM into sub-mm and IR flux (Zylka et al,1992). Adding all radio,mm and IR flux, Zylka et al (1992) find that $L_{total} \leq 10^6 L_{\odot}$ which corrects for stellar emission sources within the central $30''$. This conservative limit is on the order of 10^{-5} times the Eddington luminosity of an assumed $10^6 M_{\odot}$ black hole. If only the radio luminosity is considered then even a $1 M_{\odot}$ black hole has an Eddington luminosity 10^4 times greater than L_{radio}. Therefore, stellar mass phenomenon have not been ruled out as possible energy sources for SgrA*.

3mm-VLBI is currently the only technique with sufficient angular resolution to search for structure in SgrA*. An 86GHz experiment was carried out in 1994 (Rogers et al,1994) and showed an unresolved source or, to be more precise, a small ($\theta < 0.13$mas) upper limit on its size. This size estimate was determined by making use of non-detections on long interferometer baselines. Results depending on non-detections are always suspect but in this case there is good reason to believe that there were no instrumental problems : firm detections were made on several other target sources. The calibration of data used what was assumed to be an unresolved calibrator source with constant flux density on all

baselines. Recent re-examination of the calibrator (NRAO530) shows that it is resolved on long baselines and its structure must be compensated for in the calibration. The following section describes the observations and the methods used to recalibrate the SgrA* data. New values for the size of SgrA* are determined and found to deviate from the conservative values reported in (Rogers et al,1994) but not enough to alter their conclusions which are discussed here in Section 7.5.

7.2 Observations and Data Reduction

The observations of SgrA* are described in (Rogers et al,1994) but are briefly outlined here for completeness. An array of 3 telescopes (K,P,O in Table 5.1) observed SgrA*, NRAO530, and the SiO maser source Vxsgr for 3 consecutive days : April 2-4, 1994. OVRO used 3 antennas in its array phased together as one element. The proper phasing was determined by detecting fringes on a calibrator source between the 3 summed antennas and a separate fourth dish in the OVRO array. Most data from April 3rd is not incorporated into our analysis as OVRO experienced phasing problems on that day and could not be calibrated. At the observing frequency of 86.2GHz, the coherence time was typically $\sim$10-20 seconds. All 6.5 minute scans on SgrA* were segmented and incoherently averaged to reduce the effects of coherence loss.

Fringe searching with incoherent averaging techniques that lowered correlation amplitude detection thresholds to 0.5×10^{-4} failed to detect SgrA* on the long Haystack - OVRO and KittPeak baselines. Detection of the calibrator sources on these long baselines allowed upper limits to be placed on the SgrA* emission. NRAO530, SgrA* and Vxsgr were all easily detected on the KittPeak - OVRO baseline.

The Vxsgr data was intended to be used as the primary calibrator but it failed to adequately compensate for the effects of atmospheric attenuation. The possibility exists that a changing spectral baseline may not have been accounted for in the Vxsgr data reduction. Calibration focused instead on NRAO530 which was also observed between each SgrA* scan. NRAO530 is quite compact and has been detected with space-VLBI at 15GHz (Linfield et al,1990) and with 3mm-VLBI in April 1993.

The original calibration used correlation amplitudes for the short (PO) baselines on NRAO530 to fit a curve for SEFD as a function of elevation at each antenna. This fit

was done assuming an unresolved source and the assumed dependence of the SEFD's with elevation was :

$$\text{SEFD} \propto \left(\exp\left\{ \frac{\tau}{\sin(\theta)} \right\} - c \right) \tag{7.1}$$

with τ the atmospheric optical depth, θ the elevation angle and c a constant. A function of this form is quasi physical and comes from the definition of T_{sys} :

$$T_{sys} \simeq e^{\tau'} \left[T_{rx} + T_{atm}(1 - e^{-\tau'}) \right] \tag{7.2}$$

where $\tau' = \tau/\sin(\theta)$ and T_{rx}, T_{atm} are the receiver and atmospheric temperatures. Rearranging terms in this definition gives :

$$T_{sys} \left(\frac{1}{T_{rx} + T_{atm}} \right) \simeq e^{\tau'} - \left(\frac{T_{atm}}{T_{rx} + T_{atm}} \right) \tag{7.3}$$

So the SEFD of an antenna (the T_{sys} in Jy) is proportional to the expression in Eq. 7.1 with c being the ratio of antenna temperature to the sum of receiver and antenna temperatures. As pointed out in (Rogers et al,1994) the fit picks a τ and c that take into account not just the atmosphere but also other effects such as spillover. Thus, τ and c may not reflect the *actual* attenuation and temperature ratios described above.

Since the correlated flux density on NRAO530 was assumed to be constant, the relation between antenna SEFD and correlation amplitude was written :

$$S_{corr} = \rho\sqrt{\text{SEFD}_P\,\text{SEFD}_O} \tag{7.4}$$

The parameters for the SEFD of each station are found from a fit for each SEFD to Eq. 7.1 using the NRAO530 data. These SEFDs were then applied to convert correlation coefficient to flux density. In other words, the plot of correlated flux of NRAO530 on short baselines vs. baseline length was assumed flat and SEFDs for each antenna were found that best fit that requirement.

If it were found that NRAO530 had normalized visibility much different from 1 on long baselines (e.g. KittPeak to Haystack) then the calibration procedure would have to change. Furthermore, if the conservative Haystack sensitivity assumed in (Rogers et al,1994) is increased, the upper limit on correlated flux density found for SgrA* on long baselines would decrease and lead to a larger size estimate for SgrA*. Using an established gain curve

for Haystack, measured T_{sys} and published gain (K/Jy), the long baseline scans (OVRO and Kitt Peak to Haystack) on NRAO530 were calibrated. It was clear, using this first attempt at calibration, that the source was resolved with normalized visibility of ~ 0.5 on the Haystack baselines. In fact, the source can be modeled using a single circular gaussian with FWHM of $85\mu as$. This means the dependance of correlated flux density on baseline length will be :

$$S_{corr} \quad \propto \quad \exp\left\{-\frac{\theta^2\pi^2 B^2}{4\ln(2)}\right\} \tag{7.5}$$

$$\propto \quad e^{-B^2\ (5.9\times10^{-7})} \tag{7.6}$$

with B the baseline length in $M\lambda$. This suggests that the fit performed using Eqs. 7.1 and 7.4 needs modification to include the effects of resolution. The observed correlation amplitudes should be fit as :

$$\rho \propto \frac{e^{-B^2\ (5.9\times10^{-7})}}{\sqrt{\mathrm{SEFD}_P\ \mathrm{SEFD}_O}} \tag{7.7}$$

with the SEFDs being given by Eq. 7.1. What is actually seen when fitting Eq. 7.7 to the NRAO530 data is that the opacity term (τ) for KittPeak is very close to zero. Since that antenna shows little change of SEFD with elevation its SEFD was taken to be a constant which reduces the number of fit parameters to 3 : τ and c for OVRO and a constant of proportionality.

We immediately see that this resolution effect hardly effects the calibration since the exponential term in Eq. 7.7 ranges from 0.996 on the shortest PO baselines to 0.96 on the longest - a difference of only 3%. Figure 7-1 shows the correlation coefficient on the PO baseline for NRAO530 with one set of points assuming an unresolved source, the other a gaussian component of FWHM $85\mu as$. The two curves are fits to the data using functional forms of OVRO's SEFD as in Eq. 7.1. The solid line is the fit for NRAO530 unresolved and yields $\tau = 0.41$ and $c = 0.05$. The dotted line is a fit taking into account the core size of the calibrator with $\tau = 0.43$ and $c = -0.05$.

7.3   Modeling SgrA*

As expected , the overall calibration does not noticeably change as seen in Figs. 7-2 and 7-3 which show the two methods of calibration. What may change, however, is the upper

109
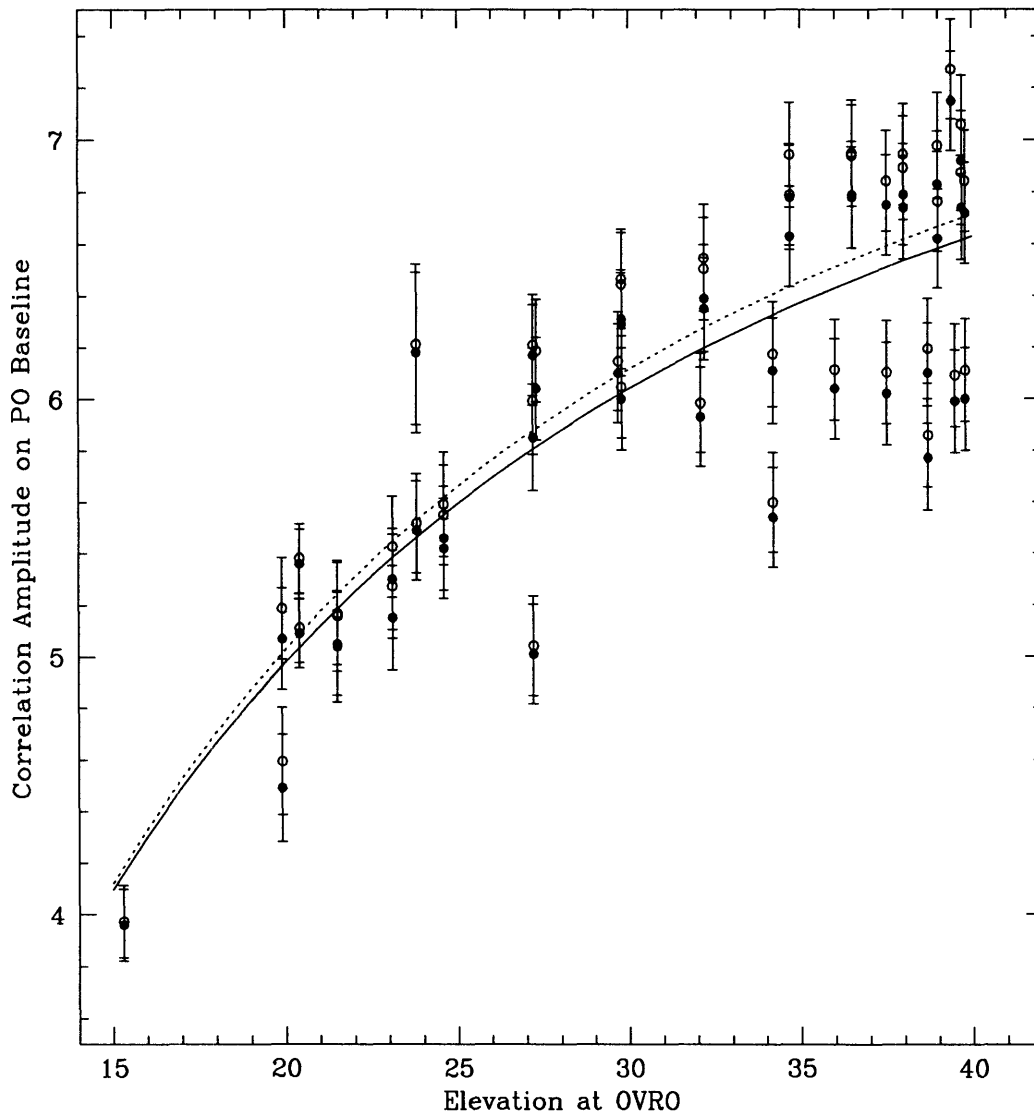
Figure 7-1: Two SEFD fits to the NRAO530 calibration data. The solid and dotted lines are for NRAO530 unresolved and resolved respectively.

limit for the correlated flux on the long baselines. The evidence that we might have to lower this limit comes from an examination of the Haystack SEFD measured during the experiment on the planet Venus. Antenna temperatures recorded on Venus are shown in Fig. 7-4. Also indicated are Haystack gains (K/Jy) for different antenna temperatures after beam corrections and taking the true flux of Venus to be 175Jy.

If we take 0.035K/Jy as the gain at Haystack then we are led to set an upper limit for the non-detection at long baselines at 0.4Jy with a σ of 0.05Jy. This is to be contrasted with the conservative limit of ~ 0.015K/Jy which one finds if assuming NRAO530 to be a point source. To find a range of aceptable sizes for SgrA* the χ^2 for a gaussian model can be mapped from the data using the flux at zero baseline length and FWHM as parameters. The resulting map (Fig. 7-5) is bounded below by a line cutting off points for which the FWHM allows too much correlated flux at 1000Mλ - the smallest Haystack baselines with elevation above 10°. The contours above this line show χ^2 increasing as the fitted curve begins to deviate from the visiblities on short baselines. Our only measured value of zero spacing flux density comes from the BIMA Array at Hat Creek (Wright&Backer,1993) nearly three months prior to the VLBI experiment. In the fit this point was not given any special weight as it might were it more current.

The $\chi^2 = 30$ and $\chi^2 = 60$ contours correspond to confidence levels of 40% and 99.9% respectively. By integrating the χ^2 probability on slices of constant FWHM we can find the probability density for any given FWHM. This was done (within the $\chi^2 = 60$ contour) and the result is shown in Fig. 7-6. The calculated means are 178±50μas for a detection limit of 0.4 Jy to Haystack and 169±50μas if the detection level rises to 0.6 Jy. This shows that using the more sensitive threshold at Haystack only moves the expected FWHM by $\sim 10\mu as$ and does not have a significant effect on conclusions drawn in the (Rogers et al,1994). This analysis does, however, point out the fact that calibration is *the* crucial matter to be addressed now that detection limits have been lowered with incoherent averaging. One feasable attack on this problem would be to use more than three antennae for the observations. While we might not hope to detect SgrA* on the long baselines, the calibrators (NRAO530, 1921-293) should show up on the new baselines. The resulting higher quality images would help calibrate SgrA*.
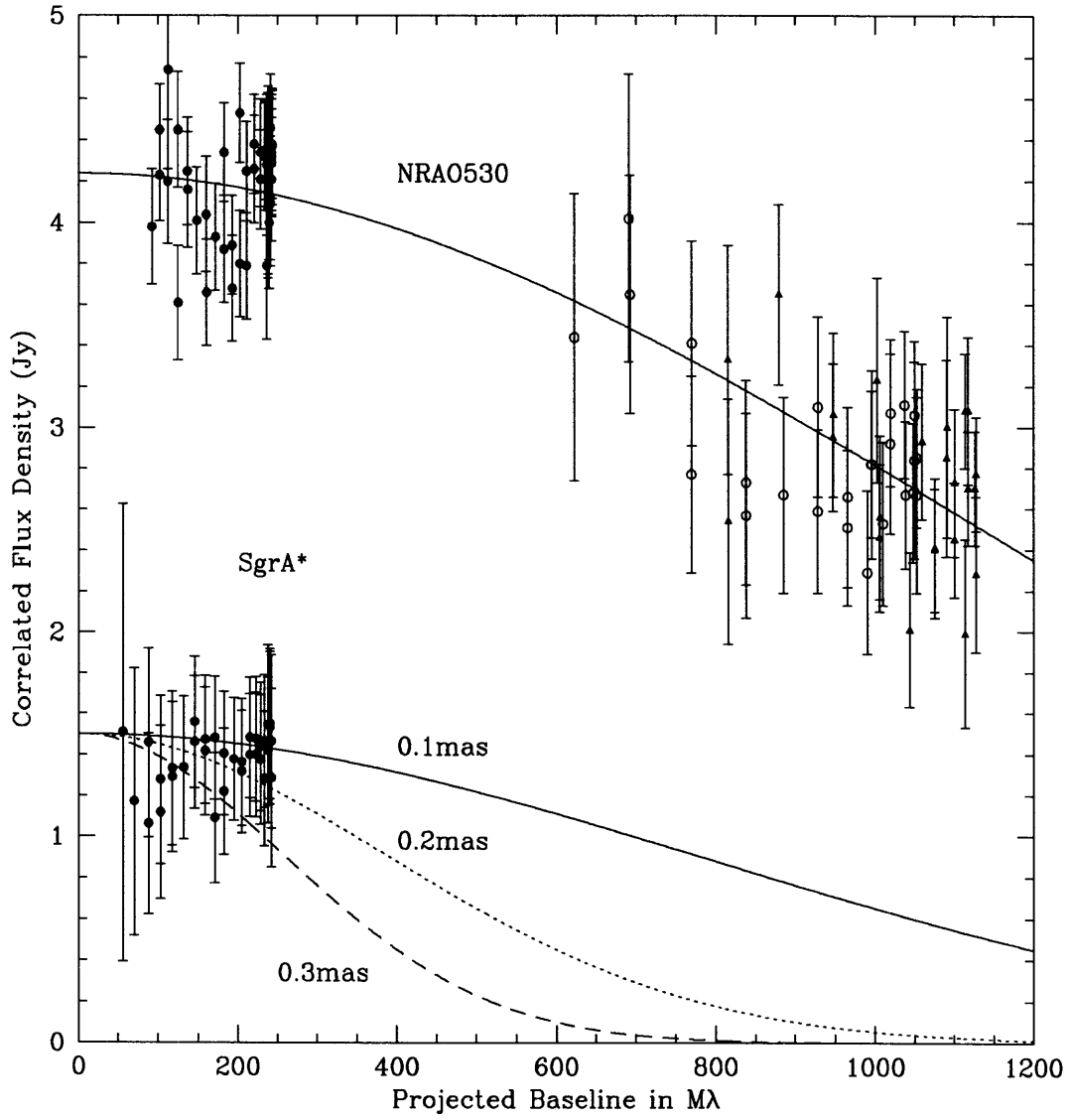
Figure 7-2: Data for NRAO530 and SgrA* using calibration of unresolved NRAO530.
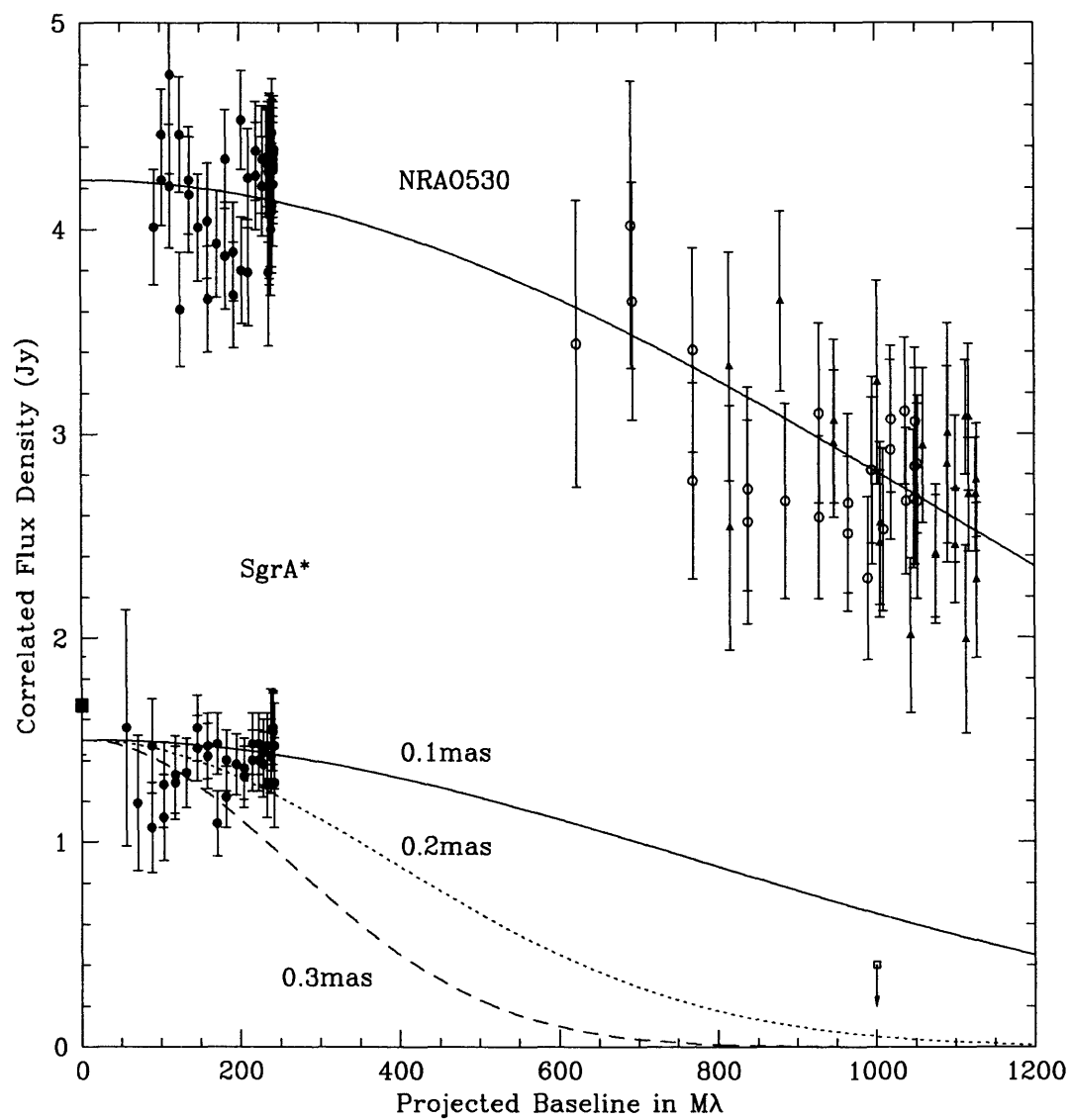
Figure 7-3: Calibrated data assuming NRAO530 has core size of $85\mu as$. The long baseline flux limit is shown as an open box at 1000Mλ.
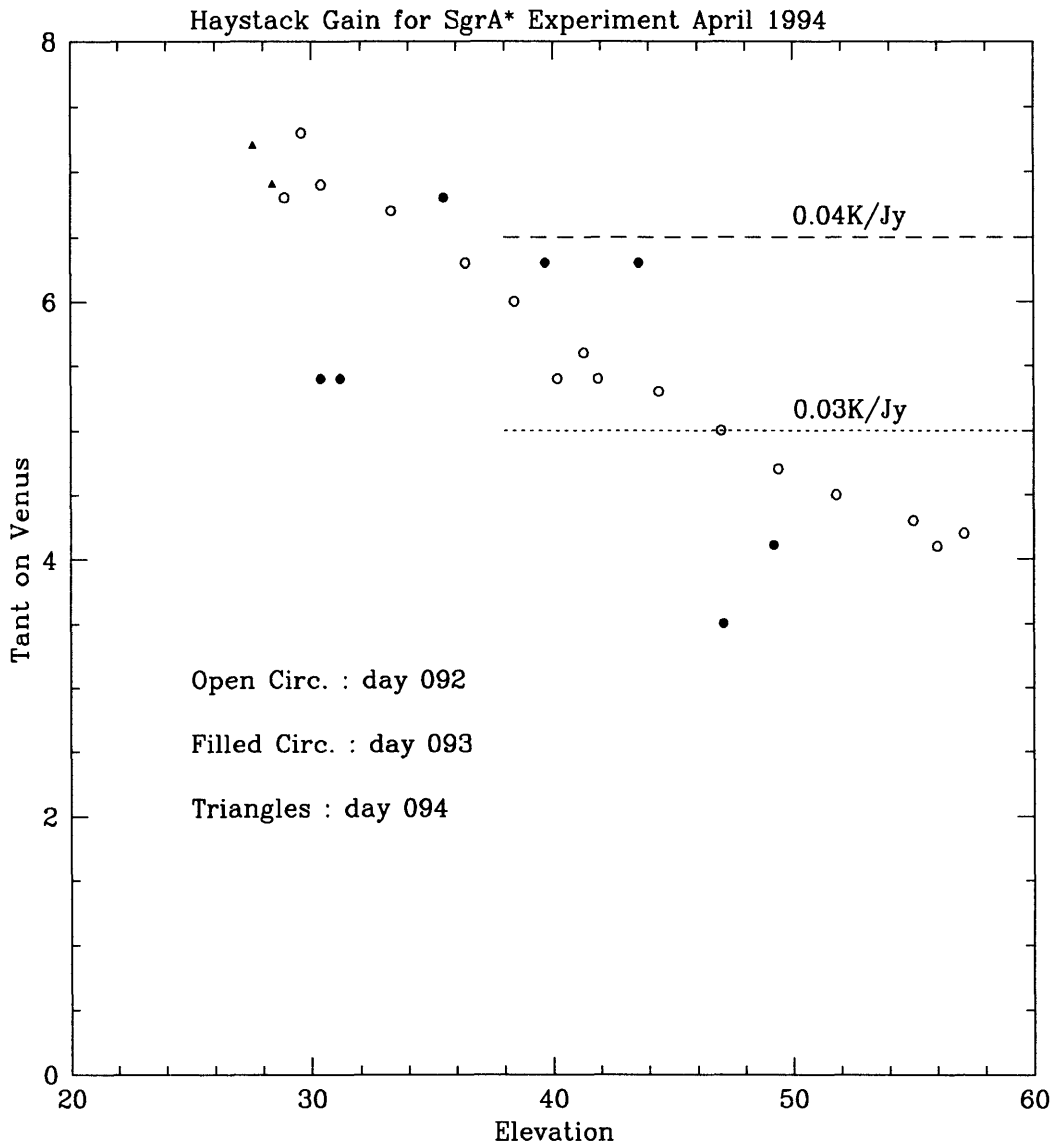
Figure 7-4: Antenna temperature of Venus at the Haystack antenna during the SgrA* experiment. Haystack calibration used an average of the values near 40° and then multiplied the gain at that elevation by a normalized gain curve determined by observing a point source (not a planet).
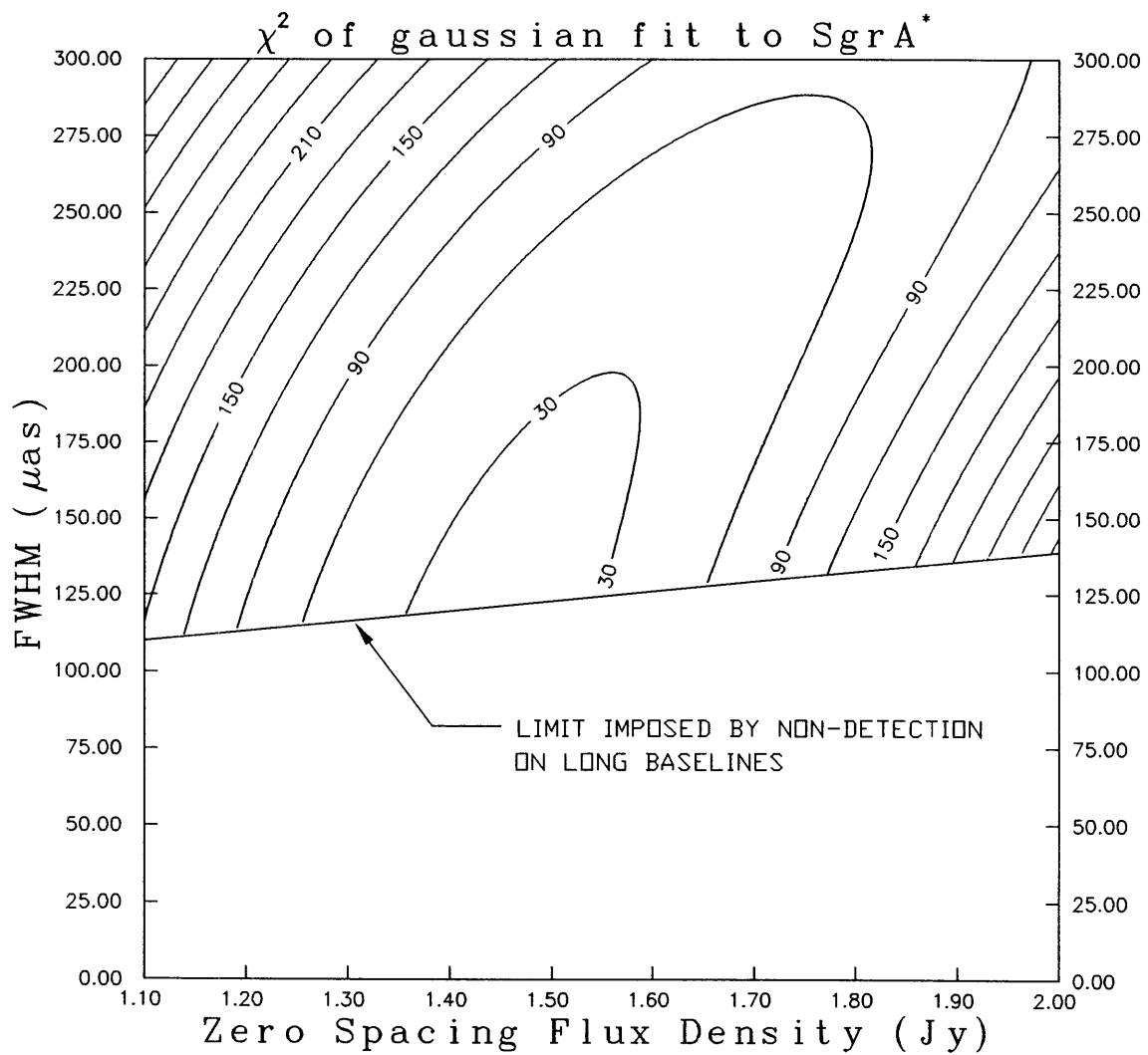
Figure 7-5: χ^2 map of two parameter gaussian fit to SgrA* visibility data. The diagonal line cutting off the lower portion of the plot is the size limit from non-detections on the Haystack - OVRO and Haystack - KittPeak baselines. The limit here is a 0.4 Jy cutoff.
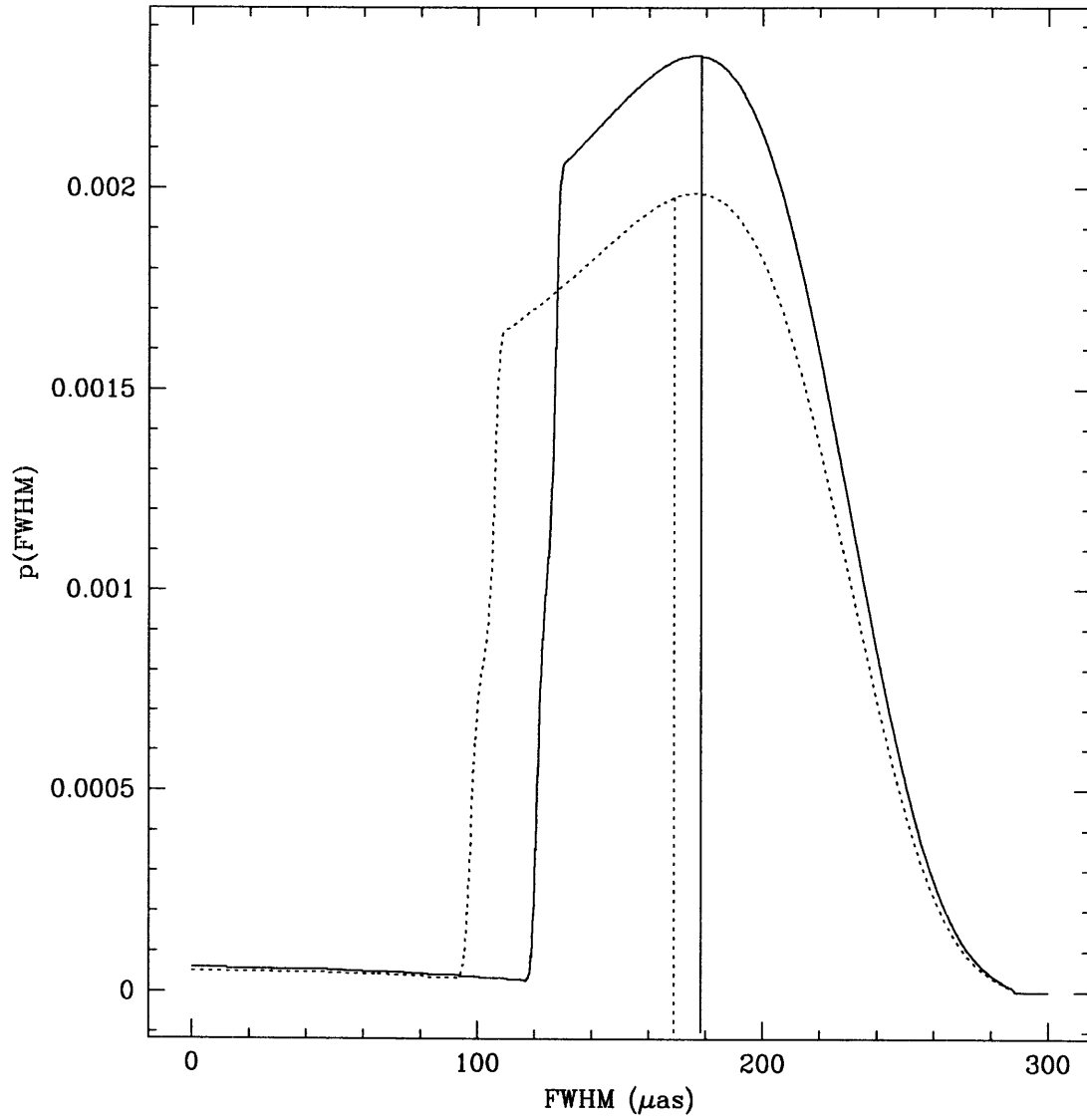
Figure 7-6: Probability densities for SgrA* FWHM using two different flux density limits on the Haystack baselines. The solid line reflects a flux threshold level of 0.4 Jy and the dotted line 0.6 Jy. Probability is calculated by integrating slices of constant FWHM through the χ^2 plane.

7.4 Scattering in the ISM

In an ionized medium, the interaction of radiation with plasma depends on the index of refraction :

$$n = 1 - \frac{r_e n_e \lambda^2}{2\pi} \tag{7.8}$$

where r_e is the classical electron radius, n_e the free electron density. If the ISM were uniform with the density n_e a constant, then a point source at the galactic center would be observed by us to be of zero width (unresolved at all baseline lengths). But if there are inhomogeneities in the medium then rays from the source will be scattered and experience phase delays which will alter the overall appearance of the source. In the case of SgrA*, the modeled brightness distribution has a striking dependence on wavelength : over two decades of observing frequency, the source size scales as $\theta_{FWHM} \propto \lambda^2$ (Marcaide et al,1992;Backer et al,1993). This dependence is illustrated in Fig. 7-7 which shows the size of SgrA* as seen in VLBI experiments taken from the literature. The two points at $\lambda = 3.5$mm show the results of the experiment described here as well as another mm-VLBI observation conducted by a group in Germany (Krichbaum et al,1994). The fitted power law is θ_{FWHM}(mas) = $(1.19 \pm 0.07) \lambda^{1.99 \pm 0.03}$ where the errors are 95% confidence values. This λ dependence is easily explained by the refractive properties of inhomogeneities in a plasma medium.

A simple way to see this is to assume a single scale size (a) for the over/under densities and calculate the expected scattering size for a point source. This is expressed as (TMS p.453):

$$\theta_s \simeq \frac{\sqrt{2\ln 2}}{\pi} r_e \lambda^2 \Delta n_e \sqrt{\frac{L}{a}} \left(\frac{R_{scr}}{R_{sou}} \right) \tag{7.9}$$

where Δn_e is the rms variation in electron density and L is the thickness of the scattering screen. The term R_{scr}/R_{sou} is the ratio of distances from the source to screen and source to observer. This ratio acts as a magnification term so that scattering close to the source has much less effect on θ_s than scattering closer to the observer. By setting $\mathcal{E}$ = EM (emission measure) = $n_e^2 L$, one can recast Eq. 7.9 as (see Moran et al(1990)) :

$$\left(\frac{a}{cm} \right) \simeq 0.15 \left(\frac{\theta_s}{as} \right)^{-2} \left(\frac{\lambda}{mm} \right)^4 \left(\frac{\mathcal{E}}{cm^{-6}pc} \right) \left(\frac{\Delta n_e}{n_e} \right)^2 \left(\frac{R_{scr}}{R_{sou}} \right)^2 . \tag{7.10}$$

Taking $\theta_s = 180\mu$as, $\lambda = 3.5$mm, $\Delta n_e/n_e = 0.1$ and using a value of EM=1.5×10^4 pc/cm^6

(Mezger&Pauls,1985) we find

$$\left(\frac{a}{\text{cm}}\right) \simeq 1.5 \times 10^{11} \left(\frac{R_{scr}}{R_{sou}}\right)^2 \qquad (7.11)$$

which gives $a \sim 10^7$cm if the screen is close to SgrA* ($R_{scr} \sim 100$pc). This crude estimate of scale size is comparable to inner scales found by Spangler&Gwinn (1990) in an analysis of scattering data from several experiments.

It is more likely, however, that the power spectrum of density fluctuations extends over many size scales. In this case, the power spectrum is generally written

$$P_{n_e}(q) = C_n^2 q^{-\beta} \qquad (7.12)$$

where q is the wavenumber associated with the length scale l: $q = 2\pi/l$. For a spectrum with $2 < \beta < 6$, the expected dependencies of scattering angle on wavelength are (Narayan,1988)

$$\theta_{scat} \propto \lambda^{\beta/(\beta-2)} \qquad 2 < \beta < 4 \qquad (7.13)$$

$$\theta_{scat} \propto \lambda^{4/(6-\beta)} \qquad 4 < \beta < 6 \qquad (7.14)$$

which would imply that $\beta \sim 4$ for the turbulence scattering the SgrA* image. It is more physical, though, to assume that there is some lower limit to the scale of inhomogeneities reached when friction dissipates the turbulence as heat. This Kolmgorov model of turbulence predicts a β of 11/3 with some inner scale set at $q_o = 2\pi/l_o$. We can find the scattering dependence on λ for this truncated spectrum ($P_{n_e}(q) = 0$ for $q > q_o$) by noting that the phase structure function of a wave front

$$D_\phi(d) = \langle [\phi(s) - \phi(s+d)]^2 \rangle \qquad (7.15)$$

is related to the visibility measured by an interferometer by

$$\mathcal{V}(d) = \mathcal{V}_o \, e^{-D_\phi(d)/2} \qquad (7.16)$$

where in both relations d is the projected baseline length of the interferometer. The struc-

ture function can be calculated as

$$D_\phi(d) = 8\pi^2 r_e^2 \lambda^2 L \int_0^{q_o} q[1 - J_o(qd)] P_{n_e}(q) \mathrm{d}q \qquad (7.17)$$

(Spangler,1988), where J_o is a bessel function. Following Moran et al(1990), this reduces to

$$D_\phi(d) = 8\pi^4 r_e^2 \lambda^2 L C_n^2 \frac{q_o^{4-\beta}}{4-\beta} d^2 \quad \text{for} \quad d < \frac{1}{2\pi q_o} \qquad (7.18)$$

so that, again, $\theta_{scat} \propto \lambda^2$ and the visibility function is a gaussian ($\mathcal{V}(d) \propto \exp(-d^2)$) regardless of the power spectrum slope. One should emphasize that the validity of Eq. 7.18 depends on the baselines remaining less than the inner scale size. In general this requirement may not be met and the visibility function will be non-gaussian with $\mathcal{V}(d) \propto \exp(-d^{\beta-2})$. A Kolmgorov spectrum, therefore, can either agree or not with the observed λ^2 behavior of the SgrA*; it depends on interferometer baselines and specifics of the scattering medium. Obtaining visibility data for a wide range of baselines allows a fit and determination of β to be made in order to constrain the choice of q_o. Mutel&Lestrade (1988) have pointed out that with high SNR VLBI data, one could look for the baseline at which the exponential dependence of visibility changes from 2 to $\beta - 2$ and so determine q_o. This method was attempted (Mutel&Lestrade,1990) on the source 2005+403 but the expected spectral break was too weak to detect.

The turbulence scales and location of the scattering media in the case of SgrA* are still open questions. Studies of OH masers in the vicinity (25') of the galactic center show angular broadening comparable to that of SgrA* (van Langevelde et al,1992). This and the strength of the scattering argue for a single region of scattering along our line of sight. Only two other galactic paths show scattering as strong : an extragalactic source seen through NGC6334 and the pulsar PSR1849+00. Both are identified with single specific scattering sites. Frail et al (1994) suggest that a hot ($\sim 10^8 K$) diffuse ($0.03 - 0.06 \mathrm{cm}^{-3}$) medium very near the galactic center may be the cause. This scenario agrees with temperature and density constraints derived from the lack of free-free absorption towards SgrA* (van Langevelde et al,1992).

We also mention here an apparent anisotropy in the scattered SgrA* image that is noticeable at longer wavelengths. Yusef-Zadeh et al (1994) point out that at 20.7,3.6,1.35 and 0.7 cm, SgrA* can be modeled as an elliptical gaussian with major axis position angle $\sim 85°$

and axial ratio ~ 0.5. This has been interpreted to be a symptom of the elongation of turbules caused by a directed magnetic field. Such stretched overdensities would preferentially scatter in a direction perpendicular to the imposed field. The observations reported here sample, for the most part, the longer of the two axes. The 3mm results of Krichbaum et al (1994) sample a direction also close to the major axis but they find a larger size of 0.33mas. The difference between these two sizes will have to be resolved with future, well calibrated observations. Fig. 7-8 shows the u,v coverage of the April 1994 experiment described here along with position angles of the galactic plane and for the u,v coverage from Krichbaum et al (1994).

An extrapolation of the expected scattering size of SgrA* from lower frequencies is 0.16 ± 0.01mas. Assuming that the scattering and intrinsic sizes add in quadrature, our largest size of 0.25mas sets a 0.19mas upper limit on the underlying structure. With a total flux of 1.5Jy, the maximum brightness temperature of a gaussian model is 7×10^9K. Thus the brightness temperature is a factor of 0.5 less and the size 30% greater than the corresponding limits given in Rogers et al (1994). It is also possible to set a lower limit to the size by considering the expected intensity fluctuations due to refractive scattering. Modeling the turbulence as a screen near the galactic center, Gwinn et al (1991) find that expected scintillation is not present over a range of time scales spanning 0.1s to 24 hours. This measurement sets a lower limit on the size of 0.1AU or 0.012mas at a wavelength of 0.8mm.

7.5 Emission Models for SgrA*

Emission models for SgrA* tend to include as centerpiece a massive black hole (MBH). The reasons for this trend are dynamical arguments rather than energetic considerations. Stellar and gas motions at the galactic center both point to a central mass of a few times $10^6 M_\odot$ (Genzel&Townes,1987). In addition, a recent proper motion measurement (Backer&Sramek,1987) confirms that SgrA* is not an extragalctic source and has a transverse velocity much lower than that expected for a stellar mass object. But stellar phenomenon such as neutron stars can easily supply the radio luminosity seen from SgrA* (Lo,1988) and canot be discounted altogether. The difficulty with these models is that they require special conditions or predict characteristics which make them unlikely can-

120

didates. A spinning neutron star, for example, matches the observed compactness if it moves through a dense medium at high velocity and emits from a compressed bow shock (Reynolds&McKee,1980). The observed low proper motion, however, argues against this scenario. Brown, Lo, and Johnston (1978) suggest a binary stellar system but this model predicts much higher radio variability than has been seen in SgrA*. For these reasons, we limit ourselves to considering MBH models and discuss two of them below in light of the recent 86GHz observations.

7.5.1 Starved AGN

In the introduction to this chapter we note that discovery of SgrA* inevitably led to comparisons with active nuclei of other galaxies. Work in this area has included scaling down MBH-jet models for AGN to account for galactic center observations. But a general difference in accretion regimes seems to separate other bright nuclei from SgrA*. To generate a total luminosity of 10^{46} erg/sec, an MBH with $M = 10^9 M_\odot$ must accrete matter very near to the critical Eddington rate ($L_{edd} \simeq 10^{47}$erg/sec). SgrA*, on the other hand, emits only about 10^{-5} of its Eddington luminosity. So scaled versions of AGN must address the following question : how can emission from the galactic center remain so low when there is clearly enough wind to generate $\sim 10^{-2} L_{edd}$. In the AGN model proposed in (Falcke,Mannheim,Biermann, 1993) and (Falcke&Heinrich,1994) the answer is the existence of a "fossil" disk left over from a previous era of nuclear activity. It presently serves to limit the accretion rate by controlling the dynamics of infalling matter.

Three main components make up this version of an AGN : a MBH, a jet emitting synchrotron radiation, and an accretion disk feeding the singularity and visible in the IR, optical and UV. The starting point in modeling SgrA* is to recognize that the accretion disk serves two purposes. Its primary responsibility is to interact with the galactic wind reducing the rate of actual accretion to $\sim 10^{-7} M_\odot$/yr to match the total observed output. The cold outer portion of this molecular disk extends to a radius of $\sim 3 \times 10^{15}$cm to catch infalling debris. A much smaller inner portion is hot and fulfills the second purpose of the disk by radiating enough to account for the IR seen in the direction of SgrA* (Eckart et al,1994). No elaborate mechanism for the formation of this "fossil" disk is worked out but the authors claim it may be a remnant of galactic evolution (Falcke&Heinrich,1994). Viscous time scales in their disk are $\sim 10^7$ years and since the present infall to SgrA* is

121

much higher than the disk accretion, they anticipate an increase in nuclear activity in the next few million years. The authors, clearly, are not opposed to taking the long view.

In the radio portion of the spectrum, SgrA* has a spectral index of ~ 0.25 from 1GHz to 200GHz. In keeping with the general AGN picture, a relativistic jet is assumed to originate in the inner diameter of the accretion disk. This jet consists of a power law of high energy electrons which, along with an equipartition magnetic field, generates the observed $\sim$ 1Jy flux in the radio. The general assumptions in deriving characteristics of this jet are that it cannot release more energy or mass than is accreted. These restrictions and the form of the jet (Blandford&Konigl,1979) result in a size of its optically thick base (z):

$$z \simeq 2 \times 10^{13} cm \left(\frac{43GHz}{\nu} \right) \tag{7.19}$$

Unfortunately, the dependence on ν in this expression means that the increased resolution at higher frequencies will be searching for a correspondingly smaller jet structure. Since the jet was not definitively resolved at 43GHz (Krichbaum et al,1993), there should be no surprise that it was not detected with the new 86GHz observations.

Overall, then, there is an unease about this particular model which cannot be readily quieted with current VLBI technology. Despite an increasing number of widely spaced mm-antenna, the weak flux density from SgrA* precludes detection on all but the short ($\sim$ 500km) baselines. and these do not have the resolution to distinguish the jet proposed by Falcke et al. Jets in extragalactic sources *are* observed at lower wavelengths and these may be expected to continue to the core, but the low activity in our galaxy may mean the SgrA* jet is not only starved by also stunted. So, while the scaled AGN model does not conflict with any major observational fact, it : (1) is forced to postulate a fossil disk, (2) predicts an observable increase in nuclear activity on a timescale far too large to be monitored, and (3) uses a jet model which cannot be resolved with current instrumentation.

7.5.2 Spherical Accretion onto Black Hole

The recent observations of outflow from the IRS16 cluster provide a natural source of fuel for a MBH at the galactic center. Melia (1992a) uses this wind for a model in which spherical accretion onto a black hole of mass $\sim 10^6 M_\odot$ emits a spectrum that matches observations of SgrA* over 12 decades of frequency. Essentials of such a spherical infall and accretion

122

were first discussed by Shvartsman (1971), later refined by Ipser & Price (1977,1982) and applied specifically to SgrA* by Melia (1992a,1992b,1994).

The attraction of this model lies in the small number of necessary free parameters due to assumptions of energy equipartition in the accretion process. Specifying just a BH mass (M), wind speed (v_g), and wind mass loss $(\dot{M}_g)$, one can start by writing down a radius of accretion (r_a) inside which all the wind is captured and drawn into the MBH's gravitational well:

$$r_a = \frac{2GM}{v_g^2}.$$ (7.20)

The accretion rate $(\dot{M})$ depends on $\dot{M}_g$ and the angle subtended by the accretion radius at the distance (D) between the wind source (taken to be IRS16) and SgrA*. These parameters $(\dot{M}, r_a, M)$ determine the radiative properties of the accreting gas by further invoking equipartition of magnetic field, gravitational, and kinetic energy densities (Shvartsman,1971). An assumption of steady flow conserves baryon accretion rate and defines the density profile (n) of the infalling gas. Enforcing these guidelines leads to the following dependencies:

$$\dot{M} \propto v_g^{-4}\dot{M}_g D^{-2}M^2$$ (7.21)

$$n \propto \dot{M}r^{-\frac{3}{2}}$$ (7.22)

$$B \propto \dot{M}^{-\frac{1}{2}}r^{-\frac{5}{4}}$$ (7.23)

These quantities, in turn, determine the heating (due to magnetic compression and reconnection of field lines) and cooling (synchrotron and bremsstrahlung) factors which yield a characteristic emission spectrum. To appreciate why equipartition is necessary, we note that if the magnetic field were frozen into the infalling plasma, the magnetic energy density would grow as r^{-4} and quickly dominate all forms of energy density. Ipser and Price (1977) argue that turbulence in the plasma would cause reconnection of field lines and a heating of the gas. They estimate magnetic heating per baryon as the difference between the "frozen magnetic field" and equipartition cases. For small radii, this magnetic term dominates the overall heating rate.

As the gas heats it also cools via two main processes : synchrotron and bremsstrahlung radiation. A signature of this model is the thermal synchrotron radiation we expect to see

due to the effective condition of local thermodynamic equilibrium. For frequencies below 10^{12}Hz, this mechanism provides most of the luminosity whereas bremstrahlung accounts for the energy output in the x-ray regime. The thermal temperature of the electrons increases towards the singularity until the cooling rate (Λ) becomes approximately equal to the heating rate (Γ). Since both vary as r^{-4}, an inner isothermal sphere exists which becomes optically thick to synchrotron radiation at some point. This specific radius where the optical depth is ~ 1 at 86GHz should be observed with 3mm-VLBI as an opaque disk of diameter 0.55mas (Rogers et al,1994).

It is this prediction that signals a fatal shortcoming of the spherical accretion model in its current form. Recall that the 2σ upper limit to the intrinsic size of SgrA* is 0.22mas including the effects of scattering. Deconvolving the expected scattering size lowers this limit to 0.13mas - well below the predicted 0.55mas. This inplies at least an incorrect radial dependence on the critical parameters of B, n, T which could, in part, stem from the spherical accretion hypothesis itself. Another way to see the failure of this model to match the 3mm observations is to work backwards from the observed size to a likely BH mass as is done in (Melia,Jokipii,Narayanan, 1992b). These authors compute the scattering size of the opaque disk as a function of BH mass using the spherical accretion model. They then use observed VLBI sizes to measure the SgrA* mass within the assumptions of this model. At the time of their 1992 paper, VLBI results were consistent with the changing radius of the disk being scattered by the ISM and they were able to set an upper limit on the MBH mass of $2 \times 10^6 M_\odot$. The new measurements correspond to a BH mass in this model of $< 1 \times 10^5 M_\odot$. This low estimate directly conflicts with the dynamical estimates of mass mentioned in the introduction to this section.

In Melia (1994), residual angular momentum of the infalling gas is included as a free parameter to investigate the possible formation of an accretion disk. Though most angular momentum is lost as the galactic wind passes through an accretion shock, it is shown to create a small inner disk within $3r_g$ of the BH ($r_g = 2GM/c^2$). Restrictions imposed by IR observations mean that the best fit disk must be inclined at an angle of 85° from the face-on orientation. But, as pointed out earlier, much of the disk flux in the UV and optical should wind up as diffuse IR after scattering in the galactic center medium. A simple radial integration of Melia's disk yields a total flux of $> 2 \times 10^8 L_\odot$ which is in excess of the total flux from a 30″ region surrounding SgrA*. This excess disk flux and the optically thick

124

inner core size both conspire to rule out this version of events in SgrA*.
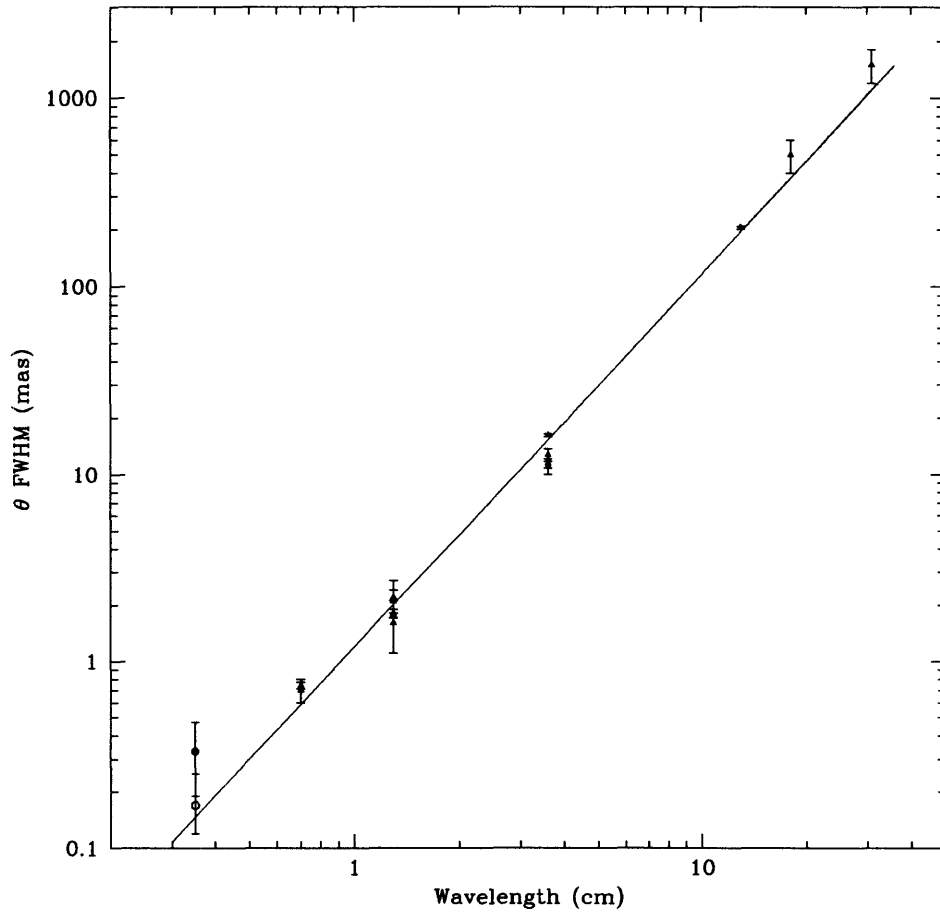
Figure 7-7: Observed size of SgrA* as a function of wavelength. Triangles represent measurements for $\lambda < 7$mm and are from Marcaide et al(1993), Backer et al(1993) and Krichbaum et al(1993). The points for $\lambda = 3.5$mm are from Krichbaum et al(1994) and the values estimated in this work.
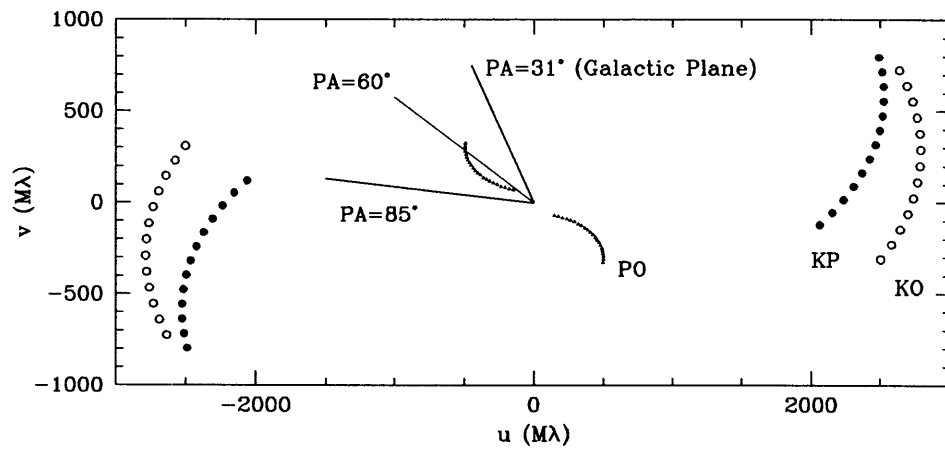
Figure 7-8: u,v coverage on SgrA* in the April 1994 observations. Shown also are the position angles for the noted major axis of ellipticity in the low frequency SgrA* images (85°), the galactic plane (31°) and the observations of Krichbaum et al(1994) (60°).

Appendix A

Form of $p(A^2)$

We would like to consider the probability density function of our averaged, noise-bias subtracted, visibility amplitude :

$$A^2 = \sqrt{\frac{1}{N}\left(\sum Z_i^2\right) - 2\sigma^2} \tag{A.1}$$

where N is the number of segments we average, Z_i is the segment amplitude, and σ is stnd. var. of one component of the noise.

For now, assume that the probability density of Z is given by an approximation to the Rice distribution (TMS p. 261):

$$p(Z) \simeq \frac{Z}{\sigma^2}\exp(-Z^2/2\sigma^2)\left[1 - \frac{V^2}{2\sigma^2} + \frac{1}{4}\left(\frac{ZV}{\sigma^2}\right)^2\right] \tag{A.2}$$

with V being the true visibility. This is an approximation for small V/σ. To find $p(A^2)$, the first step is to write the probability density for Z^2. Letting $x = Z^2$ and following Davenport & Root p.34 :

$$p(x) = \frac{1}{2\sigma^2}\exp(-x/2\sigma^2)(a + bx) \tag{A.3}$$

where

$$a = \left(1 - \frac{V^2}{2\sigma^2}\right) \text{ and } b = \frac{V^2}{4\sigma^4} \tag{A.4}$$

The next density to calculate is that describing a sum of these squared variables which will involve a convolution. To make the work easier, this convolution will be done after fourier transforming $p(x)$ to its characteristic function. The characteristic function in this

case is :

$$M_x(ju) = \frac{1}{2\sigma^2} \int_0^\infty \exp(-x/2\sigma^2)(a + bx)exp(jux)dx \qquad \text{(A.5)}$$

which becomes

$$M_x(ju) = \frac{1}{2\sigma^2} \left(\frac{1}{2\sigma^2} - ju\right)^{-1} \left(a + b\left(\frac{1}{2\sigma^2} - ju\right)^{-1}\right) \qquad \text{(A.6)}$$

so the characteristic function of the sum $y = \sum Z_i^2$ will be :

$$M_y(ju) = \left(2\sigma^2\left(\frac{1}{2\sigma^2} - ju\right)\right)^{-N} \left(a + b\left(\frac{1}{2\sigma^2} - ju\right)^{-1}\right)^N \qquad \text{(A.7)}$$

We can expand $M_y(ju)$ in a binomial series :

$$M_y(ju) = \left(2\sigma^2\left(\frac{1}{2\sigma^2} - ju\right)\right)^{-N} \sum_{k=0}^N \frac{a^k b^{N-k}}{\left(\frac{1}{2\sigma^2} - ju\right)^{N-k}} \begin{pmatrix} N \\ k \end{pmatrix} \qquad \text{(A.8)}$$

Performing the inverse transform, we recover the density function for y:

$$p(y) = \left(\frac{1}{2\sigma^2}\right)^N \sum_{k=0}^N \left(\frac{a^k b^{N-k}}{2\pi} \begin{pmatrix} N \\ k \end{pmatrix} \int_{-\infty}^\infty \frac{exp(-juy)}{\left(\frac{1}{2\sigma^2} - ju\right)^{2N-k}} dy\right) \qquad \text{(A.9)}$$

The integral in the above sum can be calculated (see D&R p. 72-3) and terms rearranged to give:

$$p(y) = \left(\frac{1}{2\sigma^2}\right)^N \left(\sum_{k=0}^N \frac{a^k b^{N-k} y^{2N-k-1}}{(2N-k-1)!} \begin{pmatrix} N \\ k \end{pmatrix} \exp(-\frac{y}{2\sigma^2})\right) \qquad \text{(A.10)}$$

In the specific case of pure noise we have $a = 1$ and $b = 0$ (i.e. Eq. A.2 reduces to the Rayleigh case) so that :

$$p(y) = \left(\frac{1}{2\sigma^2}\right)^N \frac{y^{N-1}}{(N-1)!} \exp(-\frac{y}{2\sigma^2}) \qquad \text{(A.11)}$$

Finally, we can look at the probability density function of A^2 by normalizing the sum and subtracting $2\sigma^2$:

$$p(A^2) = \left(\frac{N}{2\sigma^2}\right)^N \frac{(A^2 + 2\sigma^2)^{N-1}}{(N-1)!} \exp\left(-\frac{N(A^2 + 2\sigma^2)}{2\sigma^2}\right) \qquad \text{(A.12)}$$

We will often want to find the cumulative probability $P(A^2)$ so that a probability of error can be calculated. It is possible to find an analytic form for this integral as follows :

The cumulative probability (that $A^2 < A_o^2$) is

$$P(A_o^2) \;=\; \int_{-2}^{A_o^2} p(A^2)\,\mathrm{d}A^2 \tag{A.13}$$

$$\tag{A.14}$$

$$=\; \frac{1}{(N-1)!}\int_0^c t^{(N-1)}\,e^{-t}\,\mathrm{d}t$$

$$\tag{A.15}$$

$$=\; \frac{c^N}{(N-1)!}\int_0^1 w^{(N-1)}\,e^{-cw}\,\mathrm{d}w$$

$$\tag{A.16}$$

$$=\; \left[1 - e^{-c}\sum_{r=0}^{N-1}\frac{c^r}{r!}\right]$$

where $c = N(A_o^2 + 2\sigma^2)/2\sigma^2$ and we have changed variables twice : $t = N(A^2 + 2\sigma^2)/2\sigma^2$ and $w = t/c$.

We can then write a probability density for the maximum :

$$p_{max}(A_o^2) = m\frac{c^{(N-1)}}{(N-1)!}\frac{N}{2}\,e^{-c}\left[1 - e^{-c}\sum_{r=0}^{N-1}\frac{c^r}{r!}\right]^{(m-1)} \tag{A.17}$$

So to get the theoretical expectation for the maximum of A^2 we write :

$$< \mathrm{Max}\ A^2 >= \int_{-2}^{\infty} mA_o^2\frac{c^{(N-1)}}{(N-1)!}\frac{N}{2}\,e^{-c}\left[1 - e^{-c}\sum_{r=0}^{N-1}\frac{c^r}{r!}\right]^{(m-1)}\mathrm{d}A_o^2 \tag{A.18}$$

130
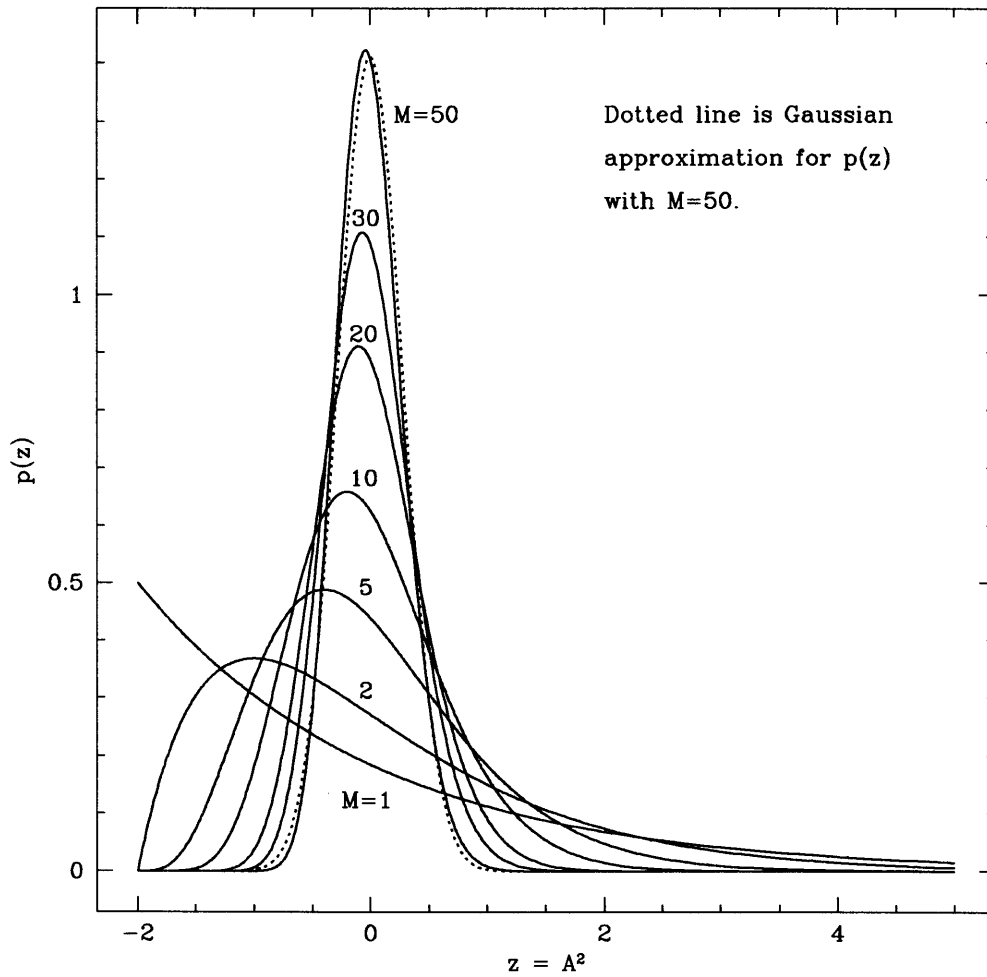
Figure A-1: Probability density of A^2 in the no-signal case for increasing numbers of segments in the average. For one segment (M=1) the density is an exponential probability density with an offset of 2. As M increases, the density approaches a Gaussian

131

Appendix B

Optimum weighting for segments of varying length

For computational and bookeeping convenience, coherently averaged intervals are marked by fixed time boundaries. This means that some "10 second" segments may contain more or less data than others. If the correlator outputs 1 second AP's, then a formal 10 second segment may contain from 0 to 10 AP's and should be weighted based on its contribution to the incoherent estimate of amplitude and the overall SNR. In most cases, the coherent segments will have nearly the same amount of data and weighting is a minor issue. If, however, recording at the antenna or playback at the correlator is of bad quality then the amount of data per segment may vary quite a bit. In this case, all segments cannot be treated equally. In this appendix, we find an optimum weighting scheme for the segments within the incoherent sum.

We start with an unknown weighting scheme for the noise bias corrected amplitudes :

$$A^2 = \frac{1}{\sum w_i} \sum w_i (Z_i^2 - 2\sigma_i^2) \tag{B.1}$$

where Z_i is the coherent measurement of amplitude for segment i and σ_i is one component of the noise ($\sigma_i = 1/\sqrt{2Bt_i}$). The expectation of the amplitude squared is given by : $\langle Z_i^2 \rangle = |V|^2 + 2\sigma_i^2$. We first make sure that this estimate remains unbiased :

$$\langle A^2 \rangle = \frac{1}{\sum w_i} \sum w_i |V|^2 = |V|^2. \tag{B.2}$$

Now the SNR_A is calculated including the weights and will be maximized to determine the exact weightings. $\mathrm{SNR}_A = \langle A^2 \rangle / \langle A^4 \rangle$ so we first examine $\langle A^4 \rangle$.

$$\langle A^4 \rangle = \frac{1}{(\sum w_i)^2} \left\langle \sum_i^M w_i^2 (Z_i^2 - 2\sigma_i^2)^2 + \sum_{i \neq j}^{M,M} w_i w_j (Z_i^2 - 2\sigma_i^2)(Z_j^2 - 2\sigma_j^2) \right\rangle \qquad (\text{B.3})$$

The first term within the expectation brackets simplifies using $\langle Z_i^4 \rangle = |V|^4 + 8|V|^2 \sigma_i^2 + 8\sigma_i^4$:

$$\text{1st Term} = \sum_i w_i^2 (|V|^4 + 4|V|^2 \sigma_i^2 + 4\sigma_i^4). \qquad (\text{B.4})$$

The second term makes use of the relation:

$$\langle Z_i^2 Z_j^2 \rangle = \left\langle \left((|V| + \sigma_i)^2 + \sigma_i^2 \right) \left((|V| + \sigma_j)^2 + \sigma_j^2 \right) \right\rangle \qquad (\text{B.5})$$

$$= \left\langle \left(|V|^2 + 2|V|\sigma_i + 2\sigma_i^2 \right) \left(|V|^2 + 2|V|\sigma_j + 2\sigma_j^2 \right) \right\rangle \qquad (\text{B.6})$$

$$= |V|^4 + 2|V|^2 \sigma_i^2 + 2|V|^2 \sigma_j^2 + 4\sigma_i^2 \sigma_j^2 \qquad (\text{B.7})$$

In Eq. B.7, all terms with single powers of σ_i and σ_j have been averaged to zero.

Case I: Low SNR

It turns out that the second term contains only terms dependent on $|V|$ and so will collapse to zero when we find $\langle A^4 \rangle$ in the no-signal ($|V| \ll \sigma$) case. We then write :

$$\langle A^4 \rangle_{\text{no-sig}} = 4 \frac{\sum w_i^2 \sigma_i^4}{(\sum w_i)^2} \qquad (\text{B.8})$$

So that

$$\mathrm{SNR}_A = \frac{|V|^2}{2} \frac{\sum w_i}{\sqrt{\sum w_i^2 \sigma_i^4}} \qquad (\text{B.9})$$

We can simplify things by adding only two segments such that $w_1 = 1$ and $w_2 = w$ and expressing the SNR as:

$$\mathrm{SNR}_A = \frac{|V|^2}{2} \frac{(1 + w)}{\sqrt{\sigma_1^4 + w^2 \sigma_2^4}}. \qquad (\text{B.10})$$

A maximum in SNR is found by requiring that

$$\frac{\mathrm{d}\,\mathrm{SNR}_A}{\mathrm{d}\,w} = 0 \qquad (\text{B.11})$$

133

which leads to the relation

$$1 = \frac{w(1+w)\sigma_2^4}{\sigma_1^4 + w^2\sigma_2^4} \tag{B.12}$$

and reduces, finally, to

$$w = \left(\frac{\sigma_1}{\sigma_2}\right)^4 = \left(\frac{\sqrt{2Bt_2}}{\sqrt{2Bt_1}}\right)^4 = \left(\frac{t_2}{t_1}\right)^2. \tag{B.13}$$

So the weighting in the sum given in Eq. B.1 should be $w_i = t_i^2$ to optimize the SNR_A.

Case II: Full Range of SNR

If the signal is large enough that it cannot be neglected with in comparison with the noise, then all terms involving $(|V|)$ must be kept in Eq. B.3. Including these terms, we reduce the second term in B.3 to :

$$\left\langle \sum_{i\neq j}^{M,M} w_i w_j (Z_i^2 - 2\sigma_i^2)(Z_j^2 - 2\sigma_j^2) \right\rangle = \sum_{i\neq j}^{M,M} w_i w_j |V|^4. \tag{B.14}$$

So that the full form of $\langle A^4 \rangle$ is

$$\langle A^4 \rangle = \frac{1}{(\sum w_i)^2} \left\{ \sum_i w_i^2 \left(|V|^4 + 4|V|^2\sigma_i^2 + 4\sigma_i^4 \right) + \sum_{i\neq j}^{M,M} w_i w_j |V|^4 \right\}. \tag{B.15}$$

The SNR must now include the effects of the signal itself :

$$\text{SNR}_A = \frac{|V|^2}{\sqrt{\langle A^4 \rangle - \langle A^2 \rangle^2}} \tag{B.16}$$

$$= \frac{|V|^2 \sum w_i}{2\sqrt{\sum w_i^2 \left(\sigma_i^2 |V|^2 + \sigma_i^4\right)}} \tag{B.17}$$

which reduces to the SNR in Eq. B.9 in the low signal case. If we, again, simplify the analysis by considering only two segments ($w_1 = 1$ and $w_2 = w$) then

$$\text{SNR}_A = \frac{|V|^2}{2} \frac{(1+w)}{\sqrt{\sigma_1^2(|V|^2 + \sigma_1^2) + w^2\sigma_2^2(|V|^2 + \sigma_2^2)}} \tag{B.18}$$

And enforcing the maximum condition as in Eq. B.11 leaves us with:

$$1 = \frac{w(1+w)\sigma_2^2(|V|^2 + \sigma_2^2)}{\sigma_1^2(|V|^2 + \sigma_1^2) + w^2\sigma_2^2(|V|^2 + \sigma_2^2)} \tag{B.19}$$

134

So that the proper weighting on the second segment is

$$w = \left(\frac{\sigma_1^2}{\sigma_2^2} \right) \left[\frac{|V|^2 + \sigma_1^2}{|V|^2 + \sigma_2^2} \right] \tag{B.20}$$

In the two extremes we see that

$$w = \frac{\sigma_1^4}{\sigma_2^4} \quad |V| \ll \sigma \tag{B.21}$$

$$w = \frac{\sigma_1^2}{\sigma_2^2} \quad |V| \gg \sigma \tag{B.22}$$

so for low SNR we want to weight with t_i^2 and in the high SNR case we would just use t_i. In the intermediate case it is less clear what to do since the weight depends on the SNR of the segment :

$$w_i = \frac{1}{\sigma_i^4 \left(\frac{|V|^2}{\sigma_i^2} + 1 \right)} \tag{B.23}$$

which means we would need, a priori, information about the signal. In any case, we have assumed in Eq. B.2 that the weights do not change the expectation of A^2. If the exact w_i directly depend on the Z_i then the unbiased estimation of $|V|^2$ may be compromised. One algorithm of assigning weights which preserves the high-low SNR distinction is to check the SNR of each segment and use B.21 or B.22 depending on some SNR cutoff. This will not be the optimum weighting in the most general sense but probably just a "better" weighting (than nothing at all). The SNR given in B.17 is the correct SNR for all cases regardless of weighting.

Appendix C

Running Mean

The incoherent averaging process requires that a VLBI scan be parsed into segments which are comparable to the coherence time of the interferometer. This sole requirement leaves open the question of exactly where we should break the data into segments. For example, the easiest interpretation is to decide on a coherence length and divide the data into non-overlapping segments, each of the coherence time. It is possible, however, to increase the sensitivity of incoherent searches by creating as many different segments as the scan will allow and averaging all of these. As a start, we could segment a scan into non-overlapping sections then, using the same coherence interval, re-segment the scan with each segment offset from its original spot by one-half the segment length. We would then have twice the number of segments, *all of them different*. In this scheme we can recalculate the incoherent SNR and compare the new detection threshold with that obtained with only one set of segments. We do this first for the case of two sets of segments then generalize to a true running mean. Both cases are worked out for the low SNR case.

Case I: Two sets of segments with offset = one-half segment

We want to compute

$$\mathrm{SNR}_A = \frac{\langle A^2 \rangle}{\sqrt{\langle A^4 \rangle}}.$$ (C.1)

Now if we do this for the case of M non-overlapping segments we find that

$$\langle A^4 \rangle = \frac{4}{M}\sigma^4 \quad and \quad \mathrm{SNR}_A = \frac{|V|^2\sqrt{M}}{2\sigma^2}.$$ (C.2)

with a flux threshold of

$$V_M = \frac{2\sigma^2 \sqrt{\text{SNR}_A}}{M^{\frac{1}{4}}} \tag{C.3}$$

So, in our case, the number of segments has grown to 2M and some of the segments will be correlated with each other. This will affect the calculation of $\langle A^4 \rangle$ but not $\langle A^2 \rangle$ which remains unchanged. Starting with the expression for A^2:

$$A^2 = \frac{1}{M} \sum (|Z_i|^2 - 2\sigma^2) \tag{C.4}$$

we expand A^4 to find :

$$\langle A^4 \rangle = \frac{1}{4M^2} \left\{ \sum_{i=1}^{2M} \langle (|Z_i|^2 - 2\sigma^2)^2 \rangle + \sum_{|i-j|=1}^{2M,2M} \langle (|Z_i|^2 - 2\sigma^2)(|Z_j|^2 - 2\sigma^2) \rangle \right. \tag{C.5}$$

$$\left. + \sum_{|i-j|>1}^{2M,2M} \langle (|Z_i|^2 - 2\sigma^2)(|Z_j|^2 - 2\sigma^2) \rangle \right\}$$

The first group of terms are products of two identical quantities (perfectly correlated). The second group are terms which involve the product of "nearest neighbor" segments which will have some correlation. The last group are products of uncorrelated segments. The first and third groups are straightforward to evaluate leaving us with :

$$\langle A^4 \rangle = \frac{1}{2M}(V^4 + 4V^2\sigma^2 + 4\sigma^4) + \frac{(2M-1)}{2M^2}(\langle |Z_i|^2 |Z_j|^2 \rangle - 4V^2\sigma^2 - 4\sigma^4) + \frac{(2M^2 - 3M + 1)}{2M^2}V^4 \tag{C.6}$$

where we have used $\langle |Z_i|^2 \rangle = V^2 + 2\sigma^2$ and it is understood that i,j represent nearest neighbor segments. We can evaluate $\langle |a_i|^2 |a_j|^2 \rangle$ by noticing that half of the noise in segment i is shared by segment j. So, in each segment, there is a signal vector $\vec{V}$ and two noise vectors:

$$\vec{Z_i} = \vec{V} + \vec{n_1} + \vec{n_2} \tag{C.7}$$

$$\vec{Z_j} = \vec{V} + \vec{n_2} + \vec{n_3} \tag{C.8}$$

Note that $\vec{n}_2$ is shared by both i and j. For now we will group the shared components of the signals together, writing $\vec{V}' = \vec{V} + \vec{n}_2$. Then

$$|Z_i|^2 = V'^2 + n_1^2 + 2\vec{n}_1 \cdot \vec{V}' \tag{C.9}$$

$$|Z_j|^2 \;=\; V'^{\,2} + n_3{}^2 + 2\vec{n}_3 \cdot \vec{V}' \tag{C.10}$$

And, neglecting all terms that cancel in the average :

$$\langle |Z_i|^2 |Z_j|^2 \rangle \;=\; \langle V'^{\,4} + V'^{\,2} n_1{}^2 + V'^{\,2} n_3{}^2 + n_1{}^2 n_3{}^2 \rangle \tag{C.11}$$

$$=\; \langle V'^{\,4} \rangle + \langle V'^{\,2} \rangle \langle n_1{}^2 \rangle + \langle V'^{\,2} \rangle \langle n_3{}^2 \rangle + \langle n_1{}^2 \rangle \langle n_3{}^2 \rangle \tag{C.12}$$

The total noise vector for one segment has a standard deviation of $\sqrt{2}\sigma$ so each of the noise vectors in Eq. C.12 will have a variance of σ^2. Therefore,

$$\langle |Z_i|^2 |Z_j|^2 \rangle = \langle V'^{\,4} \rangle + 2\langle V'^{\,2} \rangle \sigma^2 + \sigma^4 \tag{C.13}$$

Evaluation of the above expectations follows directly from TMS p. 264 so that

$$\langle |Z_i|^2 |Z_j|^2 \rangle = V^4 + 6V^2\sigma^2 + 5\sigma^4 \tag{C.14}$$

Now we are in a position to evaluate $\langle A^4 \rangle$ in the absence of signal ($V = 0$):

$$\langle A^4 \rangle = \left[\frac{2}{M} + \frac{(2M-1)}{2M^2} \right] \sigma^4 \tag{C.15}$$

and

$$\lim_{Large\,M} \langle A^4 \rangle = \frac{3}{M} \sigma^4 \tag{C.16}$$

with a resulting SNR_A of

$$\mathrm{SNR}_A = \frac{V^2 \sqrt{M}}{\sigma^2 \sqrt{3}} \tag{C.17}$$

If we convert this SNR to a flux density threshold which we call V_{2M}

$$V_{2M} = \left(\frac{3}{M} \right)^{\frac{1}{4}} \sqrt{\mathrm{SNR}_A}\, \sigma^2 \tag{C.18}$$

So comparing the two thresholds (non-overlapping vs. 2 sets overlapping) we find that

$$\frac{V_{2M}}{V_M} = \left(\frac{3}{4} \right)^{\frac{1}{4}} \tag{C.19}$$

which means a detection threshold lowered by approximately 7%.

Case II: Maximum number of segments as derived from running mean.

For incoherent averaging, the smallest possible segment size is determined by the accumulation period (AP) used in the correlator. And, if we specify segment length, the AP also fixes the maximum number of different segments into which the data can be parsed. For example, a scan might be separated into M segments of length n AP's. But we could also create $(Mn - n + 1)$ different segments if we use the n AP size and offset each segment from the next by *only one* AP. This scheme is exactly equivalent to computing a coherent running mean of the data as it exits the correlator.

To make the same calculation here as we did in Case I, we have to examine $\langle A^4 \rangle$. The starting point will be M non-overlapping segments of n AP's each. This yields Mn total AP's and $(Mn - n + 1)$ possible segments using the running mean configuration. As in Case I, $\langle A^4 \rangle$ will involve products of uncorrelated, partially correlated, and totally correlated segments. Looking at how many products share k AP's we find:

$$\text{number of products sharing k AP's} = 2(Mn - 2n + 1 + k) \tag{C.20}$$

And, as before, we will separate the components of two segments into signal, shared noise, and unshared noise. Generally,

$$\langle |Z_i|^2 |Z_j|^2 \rangle = \langle V'^{\,4} \rangle + 2\langle V'^{\,2} \rangle \langle n_o^{\,2} \rangle + \langle n_1^{\,2} \rangle^2 \tag{C.21}$$

where $\vec{n_o}$ is an unshared noise vector and $\vec{V'} = \vec{V} + \vec{n_{sh}}$ with V the signal and n_{sh} the shared noise. If we still keep k as the number of AP's in common between two segments then the variances of the two noise components will be

$$\langle n_o^2 \rangle = 2 \left(\frac{n - k}{n} \right) \sigma^2 \tag{C.22}$$

$$\tag{C.23}$$

$$\langle n_{sh}^2 \rangle = 2 \left(\frac{k}{n} \right) \sigma^2$$

so that

$$\langle |Z_i|^2 |Z_j|^2 \rangle = \langle V'^{\,4} \rangle + \frac{4(n - k)}{n} \langle V'^{\,2} \rangle \sigma^2 + \frac{4(n - k)^2}{n^2} \sigma^2 \tag{C.24}$$

139

with

$$\langle V'^4 \rangle = V^4 + 8V^2\sigma^2 \left(\frac{k}{n}\right) + 8\sigma^4 \left(\frac{k}{n}\right)^2 \qquad \text{(C.25)}$$

and

$$\langle V'^2 \rangle = V^2 + 2\left(\frac{k}{n}\right)\sigma^2 \qquad \text{(C.26)}$$

Now, if we take the no-signal case and set $V = 0$, then

$$\langle |Z_i|^2 |Z_j|^2 \rangle = \sigma^4 \left[8\left(\frac{k}{n}\right)^2 + 8\left(\frac{k}{n^2}\right)(n-k) + \left(\frac{4}{n^2}\right)(n-k)^2 \right] \qquad \text{(C.27)}$$

Our expression for $\langle A^4 \rangle$ then becomes

$$\langle A^4 \rangle = \frac{4\sigma^4}{Mn - n + 1} + \frac{\sigma^4}{(Mn - n + 1)^2} \sum_{k=1}^{n-1} 2(Mn - 2n + k + 1)\left(\frac{4k^2}{n^2}\right) \qquad \text{(C.28)}$$

Now we use the two sum formulas :

$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4} \quad \text{and} \quad \sum_{i=1}^{n} i^2 = \frac{n(2n+1)(n+1)}{6} \qquad \text{(C.29)}$$

and simplify our expression to

$$\langle A^4 \rangle = \frac{4\sigma^4}{Mn - n + 1} + \frac{4(n-1)\sigma^4}{n(Mn - n + 1)^2} \left(\frac{(Mn - 2n + 1)(2n - 1)}{3} + \frac{n(n-1)}{2} \right) \qquad \text{(C.30)}$$

so that in the limit of large M and n we have

$$\lim_{Large\, M,n} \langle A^4 \rangle = \frac{8}{3M}\sigma^4 \qquad \text{(C.31)}$$

and our new SNR is

$$\text{SNR}_A = \frac{V^2}{\sigma^2}\sqrt{\frac{3M}{8}} \qquad \text{(C.32)}$$

with a corresponding flux threshold we call V_{run} of

$$V_{run} = \sigma^2 \left(\frac{8}{3M}\right)^{\frac{1}{4}} \qquad \text{(C.33)}$$

140

The final comparison between thresholds (non-overlapping vs. running mean) gives

$$\frac{V_{run}}{V_M} = \left(\frac{2}{3}\right)^{\frac{1}{4}} \tag{C.34}$$

with a threshold reduction of $\sim 10\%$.

Appendix D

Closure Relations

The measured visibility phase in VLBI is always corrupted by uncertain atmospheric and instrumental phase offsets. Atmospheric inhomogeneities change path length from the source to antenna and effectively steer the fringe responnse pattern to a lightly different part of the sky. This changes the interferometer phase in an undetermined way assuming the baseline is large enough that the atmospheric variations above each antenna are uncorrelated. On the ground, instrumental effects such as frequency standard drift, frequency offset and polarization angle mismatch can corrupt the phase. We can write the measured fringe phase as :

$$\phi_{ij} \simeq \theta_{ij} + (\Phi_i - \Phi_j) + (E_i - E_j) \pm (\psi_i - \psi_j) \tag{D.1}$$

where θ_{ij} is the true phase, Φ_i is the atmospheric phase contribution from station i, E_i is the total error including rate offsets in the time standard at station i and ψ_i is the polarization angle measured at the ith antenna. If we then sum the measured phase around a triangle of baselines we find that most of the error terms cancel :

$$\phi_{ijk}^c = \phi_{ij} + \phi_{jk} - \phi_{ik} \tag{D.2}$$

$$\simeq \theta_{ij} + \theta_{jk} - \theta_{ik}. \tag{D.3}$$

This leaves the closure phase dependent only on intrinsic source structure. Note that ϕ_{ijk}^c is unaffected by source translation on the sky since any such offset can be decomposed into a phase offset at each antenna and cancels in the closure sum. This closure cancellation also occurs for any baseline errors that may occur. For the same reasons that leave closure

phase immune to sky position offsets, we can regain no information about absolute source position from closure phase.

Fringe rate and delay are related to the fringe phase by :

$$r_{ij} \;=\; \frac{\mathrm{d}\,\phi_{ij}}{\mathrm{d}\,t} \tag{D.4}$$

$$\tau_{ij} \;=\; \frac{\mathrm{d}\,\phi_{ij}}{\mathrm{d}\,\nu}. \tag{D.5}$$

Which says that as long as the closure phasse is not changing too fast in time or across the passband of the interferometer, the rates and delays will close to zero. The fringe fitting process takes out significant variation of phase with frequency and segmentation will prevent large temporal excursions. It is, therefore, permissible to use the zero closure of rate and delay on two legs of a baseline triangle to narrow the search for fringes on the third baseline.

Appendix E

Source parameter estimation from VLBI observables

This section gives details of the calculations used to find the equipartition magnetic field in a source and associated quantities such as minimum pressure, particle energy, and synchrotron lifetime. The analysis closely follows that of Conway et al (1983) and also draws from the treatment by Moffet (1975). A number of assumptions are made and these will be clearly stated.

We start by outlining the basic idea. VLBI observables (flux density and angular size) will be used to estimate the total luminosity of the source. This luminosity can also be found separately by integrating the synchrotron power per electron over an assumed electron number density. Comparison of these two luminosities leads to an estimate of the total energy in relativistic particles (U_p) which will depend on magnetic field. An assumption of energy equipartition between U_p and the magnetic field energy (U_b) then fixes the magnetic field and allows us to find values for the other quantities mentioned above.

Least squares fitting of visibility data in the (u, v) plane will yield sizes and fluxes of components in the observers frame. We can convert this information to the source (emitted) frame by taking into account all factors of cosmological redshift and evolution. If we imagine that all radiation in a bandpass $\Delta \nu_{ob}$ is collected from a source at redshift z, then

$$4\pi d_L^2 S_{ob}(\nu_{ob})\Delta\nu_{ob} = P_{em}\left(\nu_{ob}(1+z)\right)\Delta\nu_{em} \qquad \text{(E.1)}$$

where d_L is the luminosity distance at z:

$$d_L = \frac{c}{q_o^2 H_o} \left\{ z q_o + (1 - q_o) \left[1 - \sqrt{2 z q_o + 1} \right] \right\},$$ (E.2)

$S_{ob}(\nu)$ is the observed flux density at ν, $\Delta \nu_{em}$ is the corresponding bandpass in the source frame and $P_{em}(\nu)$ is the power emitted per unit bandwidth. We will suppose that we observe the source in the optically thin regime so that $P_{em} \sim \nu^{-\alpha}$. Furthermore, the bandwidth of emission and observation scale as $(1 + z)$: $\Delta \nu_{em} = \Delta \nu_{ob}(1 + z)$ so that

$$P_{em}(\nu_{ob}) = 4\pi d_L^2 S_{ob}(\nu_{ob}) (1 + z)^{\alpha - 1}.$$ (E.3)

If we fix some reference frequency (ν_o) different from ν_{ob} then :

$$S_{ob}(\nu_o) = S_{ob}(\nu_{ob}) \left(\frac{\nu_{ob}}{\nu_o} \right)^{\alpha}.$$ (E.4)

The total source luminosity between endpoint frequencies ν_1, ν_2 is

$$
\begin{aligned}
L_{em} &= \int_{\nu_1}^{\nu_2} P_{em}(\nu) \, d\nu \\
&= 4\pi d_L^2 (1 + z)^{\alpha - 1} S_{ob}(\nu_{ob}) \left(\frac{\nu_{ob}}{\nu_o} \right)^{\alpha} \frac{\nu_o}{1 - \alpha} \left[\left(\frac{\nu_2}{\nu_o} \right)^{1 - \alpha} - \left(\frac{\nu_1}{\nu_o} \right)^{1 - \alpha} \right].
\end{aligned}
$$ (E.5)(E.6)

Conway et al define

$$f(\alpha) = \left[\left(\frac{\nu_2}{\nu_o} \right)^{1 - \alpha} - \left(\frac{\nu_1}{\nu_o} \right)^{1 - \alpha} \right] / (1 - \alpha)$$ (E.7)

so that

$$L_{em} = \nu_o P_{em}(\nu_o) f(\alpha).$$ (E.8)

Now it remains to find the L_{em} predicted from synchrotron theory for comparison. We will make the approximation that an electron of energy E in a magnetic field $B_\perp$ perpendicular to its motion emits power $p = C_2 B_\perp^2 E^2$ tightly distributed around a frequency $\nu = C_1 B_\perp E^2$. Here we use cgs units so that $C_1 = 6.265 \times 10^{18}$ and $C_2 = 2.367 \times 10^{-3}$ and $B_\perp$ is in gauss, E in ergs. Taking the electron distribution to be

$$N(E) = N_o \left(\frac{E}{E_o} \right)^{-(2\alpha + 1)}$$ (E.9)

145

with E_o corresponding to ν_o we find

$$L_{em} = \int_{E_1}^{E_2} p(E)N(E)\,\mathrm{d}E = \frac{1}{2}C_2 B_\perp^2 N_o E_o^3 f(\alpha).$$
(E.10)

Setting the two forms of L_{em} equal in Eq. E.8 and Eq. E.10:

$$P_{em}(\nu_o) = \frac{1}{2}\frac{C_2}{C_1} B_\perp N_o E_o$$
(E.11)

We can now determine the total energy in relativistic electrons by

$$U_e = \int_{E_1}^{E_2} N(E)E\,\mathrm{d}E = \frac{1}{2}N_o E_o^2 g(\alpha)$$
(E.12)

where we define

$$g(\alpha) = \left[\left(\frac{\nu_2}{\nu_o}\right)^{\frac{1}{2}-\alpha} - \left(\frac{\nu_1}{\nu_o}\right)^{\frac{1}{2}-\alpha}\right] / (\frac{1}{2} - \alpha)$$
(E.13)

Keeping in mind that $P_{em}(\nu_o)$ is a quantity that follows directly from our VLBI observations we relate U_e and $P_{em}(\nu_o)$ by

$$U_e = C_3 P_{em}(\nu_o)\nu_o^{\frac{1}{2}} B_\perp^{-\frac{3}{2}} g(\alpha)$$
(E.14)

where $C_3 = \sqrt{C_1}/C_2$. A small correction should be made for the randomly directed field lines and electron velocity vectors. Since the power per electron is proportional to the sin of the pitch angle between field direction and electron velocity, we multiply U_e by $\langle \sin\theta \rangle = \pi/4$. Eq. E.14 can then be rewritten

$$U_e = C_4 P_{em}(\nu_o)\nu_o^{\frac{1}{2}} B^{-\frac{3}{2}} g(\alpha)$$
(E.15)

where $C_4 = C_3(4/\pi)^{\frac{3}{2}}$. The total energy U_p in relativistic particles includes the electron component in Eq. E.15 and an additional amount from protons (U_{prot}). These heavier particles do not radiate synchrotron radiation in the same frequency range due to their mass but do contribute to the total energy density. If both species of particle were accelerated to the same velocity via the same mechanism then the energy densities would scale with the mass ratio $(U_{prot}/U_e = k_E \sim 1800)$. Burbidge (1956,1959), however, argues for a k_E of 50-100 based on a different model of particle acceleration in which the protons are primary

particles with the electrons produced via proton collisions with surrounding media. In general, the value of k_E depends on the acceleration mechanism and is not well constrained. We take k_E to be 1 in our calculations but keep this dependence in the formulae for reference.

Equipartition between particle energy and energy stored in the magnetic field can be expressed as

$$U_p = U_b = \frac{B^2}{8\pi}V \tag{E.16}$$

with V the volume of emission. Note that this is *nearly* the same as requiring that the total energy be minimized. The total energy is

$$U_T = U_p + U_b = \frac{B^2}{8\pi}V + KB^{-\frac{3}{2}} \tag{E.17}$$

with $K = (1 + k_E)C_4 P_{em}(\nu_o)\nu_o^{\frac{1}{2}} g(\alpha)$ and it has a minimum when

$$B_{min} = \left[\frac{6\pi K}{V}\right]^{\frac{2}{7}} \tag{E.18}$$

whereas the equipartition field is

$$B_{eq} = \left[\frac{8\pi K}{V}\right]^{\frac{2}{7}} \tag{E.19}$$

so that B_{eq} is $\sim 8\%$ greater than B_{min}. The total energies in both cases differ by only a small amount :

$$U_T(B_{eq}) = 2\left(\frac{V}{8\pi}\right)^{\frac{3}{7}} K^{\frac{4}{7}} \tag{E.20}$$

$$U_T(B_{min}) = \left(\frac{V}{\pi}\right)^{\frac{3}{7}} K^{\frac{4}{7}} \left[6^{-\frac{3}{7}} + \frac{6^{\frac{4}{7}}}{8}\right] \tag{E.21}$$

so that $U_T(B_{eq})$ is greater than $U_T(B_{min})$ by $\sim 1\%$ and equipartition is effectively minimization. Assuming equipartition, the field is given by

$$B_{eq} = \left[\frac{8\pi}{V}(1 + k_E)C_4 P_{em}(\nu_o)\nu_o^{\frac{1}{2}} g(\alpha)\right]^{\frac{2}{7}} \tag{E.22}$$

$$= \left[32\pi^2(1 + k_E)C_4\nu_o^{\frac{1}{2}} g(\alpha)(1 + z)^{3+\alpha}\frac{S_{ob}(\nu_o)}{l\theta^2}\right]^{\frac{2}{7}} \tag{E.23}$$

$$= 1.57 \times 10^4 \left[\frac{(1 + k_E)\nu_o^{\frac{1}{2}} g(\alpha)(1 + z)^{3+\alpha} S_{ob}(\nu_o)}{l\theta^2}\right]^{\frac{2}{7}} \tag{E.24}$$

where $l = V^{\frac{1}{3}} = \theta d_L/(1+z)^2$ and θ is the component size estimated from VLBI measurements. In more useful units :

$$B_{eq} = 0.013 \left[\frac{(1 + k_E)\nu_o^{\frac{1}{2}} g(\alpha)(1 + z)^{5+\alpha} S_{ob}(\nu_o)}{\theta^3[(1 + z) - \sqrt{1 + z}]} \right]^{\frac{2}{7}} h^{-\frac{2}{7}} \qquad (E.25)$$

with ν_o in GHz, S_{ob} in Jy, B_{eq} in Gauss and θ in milliarcsec.

The value and relative contribution of the $g(\alpha)$ factor merits some discussion. It depends on the high and low synchrotron cutoff frequencies which have customarily been assigned the values 100GHz and 10MHz respectively. The lower limit is usually set to the synchrotron self-absorption turn-over frequency if this feature is present since the $\sim \nu^{\frac{5}{2}}$ dependence below this point will cause the energy flux ($\sim \nu S(\nu)$) to drop precipitously. This self-absorption frequency is approximately (Moffet,1975):

$$\nu_m \approx \left[\frac{900 S_m B^{\frac{1}{2}}}{\theta^2} \right]^{\frac{2}{5}} \text{ MHz} \qquad (E.26)$$

so that for a source of dimension 0.5mas, flux density of 10Jy and a field of 1 Gauss, $\nu_m \approx 15$GHz. This agrees with the observation that many 3mm-VLBI sources exhibit self-absorbed cores at 6 and 3cm wavelengths. We therefore take $\nu_1 = 10$GHz for our analysis. Conway et al choose an upper cutoff of 100GHz - the point beyond which a lack of highly relativistic electrons reduces the synchrotron emission. At an observing frequency of 86GHz one is looking at the most compact components, those closest to the actual central engine and we expect the ν_2 to move higher for this specific region. AGN spectra generally show a high frequency turnover at $\sim$ 1mm wavelength (Landau,1986; Bloom et al,1994) so we have adopted a ν_2 of 300GHz for the components observed in this thesis.

These points being made, the question of overall dependence of B_{eq} on $g(\alpha)$ can be summarized by writing Eq.(xx) for a specific example : $\nu_o = 86$GHz, $S_{ob}(\nu_o) = 1$Jy, $\theta = 0.5$mas, $z = 0.5$. Fig. E-1 shows the behavior of B_{eq} for different low frequency cutoff values and for a range of spectral index. Even as ν_1 changes over 5 decades of frequency and α varies from -0.3 to 0.7, the maximum shift in B-field is only a factor of 2. B_{eq} is thus fairly insensitive to parameters contributing to $g(\alpha)$; it is much more sensitive to the observed angular size which has led to a demand for better modeling of compact high frequency VLBI components.
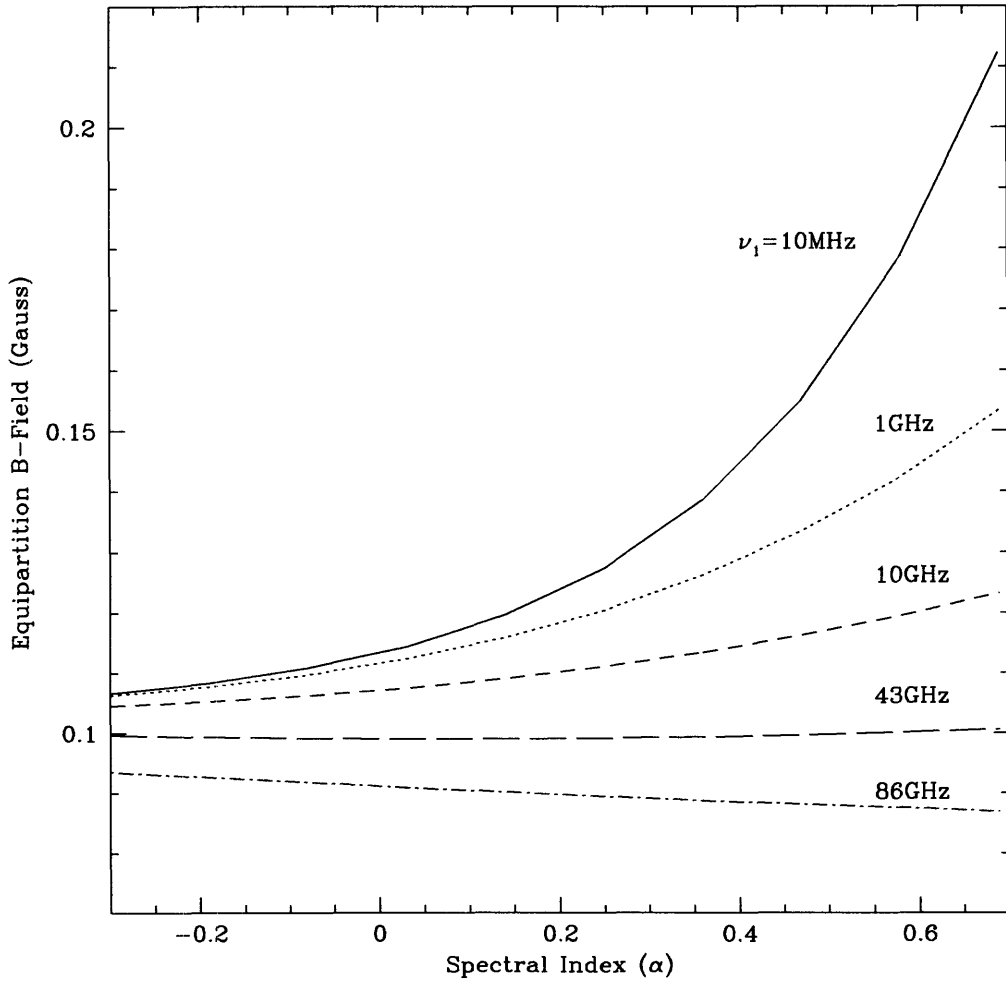
Figure E-1: Values of B_{eq} for an example source. Curves show the effects on B_{eq} of changing α and ν_1 keeping ν_2 fixed at 300GHz.

The total energy (U_T), the total energy density (u_T) and the synchrotron lifetime at the reference frequency (T_{synch}) can now be evaluated:

$$U_T(B_{eq}) = VB_{eq}^2/4\pi \tag{E.27}$$

$$u_T(B_{eq}) = B_{eq}^2/4\pi \tag{E.28}$$

$$T_{synch} = C_3(B_{eq}^3\nu_0)^{-\frac{1}{2}} \tag{E.29}$$

Contributions to pressure due to the energy density above breaks down into two parts : one from the relativistic particles and the other from the tangled B-field so that

$$P = \frac{1}{V}\left(\frac{U_p}{3} + U_b\right) \tag{E.30}$$

149

which, when minimized with respect to the magnetic field yields

$$P_{min} \simeq 0.53 \left(\frac{U_T(B_{eq})}{V} \right) \qquad (E.31)$$

Bibliography

Alef, W. & Porcas, R. 1986, A & A, **168**, 365.

Antonucci, R. & Miller, J. 1985, ApJ, **297**, 621.

Antonucci, R. 1993, Ann. Rev. Ast. Astr., **31**, 473.

Backer, D. C. et al. 1993, Science, **262**, 1414.

Backer, D. C., Sramek, R. A. 1987, in *The Galactic Center* (AIP Conference Proceedings 155), ed. D. Backer (New York:AIP), p.163.

Backer, D. C. 1988 in *Radio Wave Scattering in the Interstellar Medium* (AIP Conference Proceedings 174), ed. J. Cordes, B. Rickett, and D. Backer (New York:AIP),p.111.

Balick, B., Brown, R. 1974, ApJ, **194**, 265.

Bevington, P. 1969, *Data Reduction and Error Analysis for the Physical Sciences*, (New York:McGraw-Hill).

Blandford, R. & Konigl, A. 1979, ApJ, **232**, 34.

Blandford, R. 1990 in *Active Galactic Nuclei* (Berlin:Springer-Verlag), p.161.

Blitz, L., Binney, J., Lo, K. Y., Bally, J. & Ho, P. T. P. 1993, Nature, **361**, 417.

Bloom, S., et al. 1994, AJ, **108**, 398.

Born, M. & Wolf, E. 1987, *Principles of Optics*, (New York:Pergamon).

Brown, R. L., Lo, K. Y. & Johnston, K. J. 1978, AJ, **83**, 1594.

Bruyn, A. G. de 1976, A & A, **52**, 439.

Burbidge, G. 1956, ApJ, **124**, 48.

Carilli, C.L, Bartel, N. & Linfield, R.P., 1991, AJ, **102**, 1691-1695.

Carlson, J. & Hines, A. 1947, Quarterly of App. Math., **Vol. IV**, 313.

Clark, B. G. 1968, IEEE Trans. Ant. Prop., **AP-16**, 143.

Cohen, M. & Unwin, S. 1984, in *VLBI and Compact Radio Sources*, (IAU Symposium 110), ed. R. Fanti, et al, (Boston:D.Reidel), p.95.

Conway, R., et al. 1983, M.N.R.A.S., **202**, 813.

Davenport, W. & Root, W. 1958, *Random Signals and Noise*, (New York:McGraw-Hill).

Eckart, A. et al. 1994 in *The Nuclei of Normal Galaxies*, (Kluwer), p.305.

Falcke, H., Mannheim, K. & Biermann, P. L. 1993, A & A, **278**, L1.

Falcke, H. & Heinrich, O. M. 1994, A & A, **292**, 430.

Frail, D. A., Diamond, P. J., Cordes, J. M. & van Langevelde, H. J. 1994, ApJL, **427**, L43.

Genzel, R. & Townes, C. H. 1987, Ann. Rev. Ast. Astr., **25**, 418.

Ghisellini, G., et al. 1992, in *Testing the AGN Paradigm* (AIP Proceedings 254), ed. S. Holt, S. Neff, C. Urry, (NewYork:AIP), p.398.

Goss, W. M. et al. 1988, in The Center of the Galaxy (IAU Symposium 136), ed. M. Morris, (Boston:Kluwer), p.345.

Gotz, M., Alef, W., Preuss, E. & Kellerman, K. 1987, A&A, **176**, 171.

Gould, R. J. 1979, A & A, **76**, 306.

Gwinn, C.R., et al. 1991, ApJL, **381**, L43.

Hine, R. G., Scheuer, P. 1980, M.N.R.A.S., **193**, 285.

Hughes, P., Aller, H., Aller, M. 1992, ApJ, **396**, 469.

Ipser, J. R. & Price, R. H. 1977, ApJ, **216**, 578.

Ipser, J. R. & Price, R. H. 1982, ApJ, **255**, 654.

Jauncey, D. L. 1989, AJ, **98**, 44.

Jennison, R., 1958, M.N.R.A.S., **118**, 276.

Landau, R., et al. 1986, ApJ, **308**, 78.

Kirschbaum, H. S. & Chen, S. 1957, IRE Trans. Mic. Th. and Tech., **MTT-5**, 199.

Krichbaum, T. P. 1993, A & A, **274**, L37.

Krichbaum, T. P., et al. 1994, in *Compact Extragalactic Radio Sources* (NRAO Workshop No.23), ed. J. Zensus & K. Kellerman, (GreenBank,WV:NRAO), p.39.

Lengyel, B. 1951, J. of App. Physics, **22**, 265.

Linfield, R. & Perley, R. 1984, ApJ, **279**, 60.

Linfield, R. 1981, ApJ, **244**, 436.

Linfield, R. 1987, ApJ, **317**, 121.

Linfield, R., et al. 1990, ApJ, **358**, 350.

Lo, K. Y., Schilizzi, R. T., Cohen, M. H. & Ross, H. N. 1975, ApJL, **202**, L63.

Lo, K. Y., Cohen, M. H., Readhead, A. S. C. & Backer, D. C. 1981, ApJ, **249**, 504.

Lo, K. Y., 1988, in The Center of the Galaxy (IAU Symposium 136), ed. M. Morris, (Boston:Kluwer), p.527.

Lo, K. Y. et al. 1993, Nature, **362**, 38.

Lynden-Bell, D., Rees, M. 1971, M.N.R.A.S., **152**, 461.

Marcaide, J. M. et al. 1992, A & A, **258**, 295.

Marscher, A., Broderick, J. 1981, ApJ, **249**, 406.

Marscher, A. 1983, ApJ, **264**, 296.

Marscher, A. 1992, in *Sub-Arcsecond Radio Astronomy*, ed. R. Davis & R. Booth, (Cambridge:Cambridge University Press), p.297.

Melia, F. 1992a, ApJL, **387**, L25.

Melia, F., Jokipii, J. R., & Narayanan, A. 1992b, ApJL, **395**, L87.

Melia, F. 1994, ApJ, **426**, 577.

Mezger, P. & Pauls, T. 1985, in *The Large-Scale Characteristics of the Galaxy* (IAU Symposium 84), ed. W. Burton, (Dordrecht:Reidel), p.357.

Moffet, A. 1975, in *Stars and Stellar Systems vol.IX : Galaxies and the Universe*, ed. A. Sandage, M. Sandage, J. Kristian, (Chicago:U. of Chicago), p.211.

Moran, J. M., 1975, in *Methods of Experimental Physics*, ed. M. Meeks, (New York:Academic Press), **12C**, p.228.

Moran, J. M., Rodriguez, L. F., Greene, B. & Backer, D. C. 1990, ApJ, **348**, 147.

Mutel, R. & Lestrade, J. 1988, in *Radio Wave Scattering in the Interstellar Medium* (AIP Proceedings 174), ed. J. Cordes, B. Rickett & D. Backer, (NewYork:AIP), p.122.

Mutel, R. & Lestrade, J. 1990, ApJL, **349**, L47.

Narayan, R. 1988, in *Radio Wave Scattering in the Interstellar Medium* (AIP Proceedings 174), ed. J. Cordes, B. Rickett & D. Backer, (NewYork:AIP), p.17.

Pearson, T. & Readhead, A. 1984, Ann. Rev. Ast. Astr, **22**, p.97.

Penzias, A. & Burrus, C. 1973, Ann. Rev. Ast. Astr, **11**, p.51.

Pauliny-Toth, I. I., et al 1976, A & A, **52**, 471.

Preuss, E. et al. 1987 in *The Impact of VLBI on Astrophysics and Geodetics* (IAU Symposium 129), ed. M. Reid & J. Moran, (Boston:Kluwer), p.105.

Preuss, E. et al. 1990 in *Parsec-scale radio jets* (NRAO workshop, Socorro, N.M.), ed. J. Zensus & T. Pearson, (Cambridge:Cambridge University Press), p.120

Readhead, A. 1994, ApJ, **426**, 51.

Reynolds, S. P. & McKee, C. F. 1980, ApJ, **239**, 893.

Robson, E.I., et al., 1993, M.N.R.A.S., **262**, 249-272, 1993.

Rogers, A. E. E. et al. 1974, ApJ, **193**, 293.

Rogers, A. E. E. et al. 1983, Science, **219**, 51.

Rogers, A. E. E. et al. 1984, Radio Science, **19**, 1552.

Rogers, A. E. E. & Doeleman S. 1992, Haystack Obs. mm-wave memo series, **1**.

Rogers, A. E. E. et al. 1994, ApJL, **434**, L59.

Rogers, A. E. E., Doeleman, S., Moran, J. M. 1995, AJ, **109**, 1391.

Ruffert, M. & Melia, F. 1994, A & A, **288**, L29.

Sargent, W. 1977, ApJL, **212**, L105.

Serabyn, E. & Lacy, J. H. 1985, ApJ, **293** 445.

Shvartsman, V. F. 1971, Soviet Astronomy – AJ, **15**, No. 3, 377.

Sofue, Y., Reich, W. & Reich, P. 1989, ApJL, **341**, L4.7

Spangler, S. 1988, in *Radio Wave Scattering in the Interstellar Medium* (AIP Proceedings 174), ed. J. Cordes, B. Rickett & D. Backer, (NewYork:AIP), p.32.

Spangler, S. R. & Gwinn, C. R. 1990, ApJL, **353**, L29.

Terasranta, H., et al. 1992, A & AS, **94**, 121.

Terasranta, H., & Valtaoja, E. 1994, A & A, **283**, 51.

Thompson, A. R., 1980, VLA Electronics Memo No. 192.

Thompson, A. R. & D'Addario, L. R. 1982, Radio Science, **17**, 357.

Thompson, A. R., Moran, J. M. & Swenson, G. W. (TMS) 1986, *Interferometry and Synthesis in Radio Astronomy* (New York:Wiley).

Treuhaft, R. & Lanyi, G. 1987, Radio Science, **22**, 251.

van Langevelde, H. J., Frail, D. A., Cordes, J. M. & Diamond, P. J. 1992, ApJ, **396**, 686.

Weymann, R. J., et al 1982, ApJ, **262**, p.497.

Wilkinson, P. 1988, in *Very Long Baseline Interferometry : Techniques and Applications*, ed. M. Felli & R. Spencer, (Dordrecht:Kluwer), p.69.

Wills, B. 1975, ApJL, **202**, L59.

Wright, M. & Backer, D. 1993, ApJ, **417**, 560.

Yusef-Zadeh, F. & Melia, F. 1992, ApJL, **385**, L41.

Yusef-Zadeh, F., Cotton, W., Wardle, M., Melia, F. & Roberts, D. A. 1994, ApJL, **434**, L63.

Zylka, P., Mezger, P. G. & Lesch, H. 1992, **261**, 119.

Zylka, P., et al. 1994, in *The Nuclei of Normal Galaxies*, ed. R. Genzel, A. Harris, (Dordrecht:Kluwer), p. 161.