THE STRUCTURE OF HYDROXYL MASERS

AND CIRCUMSTELLAR ENVELOPES

OF LONG PERIOD VARIABLE STARS.

Thesis by

Mark Jonathan Reid

In Partial Fulfillment of the Requirements

For the degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1975

(Submitted October 10, 1975)

ACKNOWLEDGMENTS

Throughout the two years it took to prepare this thesis, I have been helped by many people in a variety of ways. I wish to thank those people who have been so generous with their time, knowledge and effort on my behalf.

To my wife, Katherine, I give special thanks for all the support and encouragement she has given me. Her influence and effort are evident throughout the text and in many of the drafted figures.

I wish to thank my advisor, Duane Muhleman, for supporting the observing program when it seemed to be a lost cause, and for the almost daily assistance he has offered me during my four years of graduate work.

The successful running of the VLB experiments was due to the efforts of many people. Duane Muhleman, James Moran, Kenneth Johnston, Phillip Schwartz, Jonathan Romney and Richard Schilizzi were very attentive during the actual observations. Jack Welch and the entire Hat Creek staff did an outstanding job keeping their telescope "on the air" almost every minute of the scheduled time. Peter and Holly Schloerb took a late night, impromptu "vacation" to the Owens Valley observatory to replace a broken video tape recorder and Harry Hardebeck lost that night's sleep modifying and adjusting the new recorder.

While a great deal of effort goes to recording VLB data, an equally great effort is needed to process this data on the NRAO Mk II VLB processor. I wish to thank Benno Reyher, Joe Burch, and Kris Rye for their efforts in handling and processing nearly one ton of video

ABSTRACT

Long period variable stars with envelopes of circumstellar dust
and gas often exhibit strong hydroxyl (OH) maser emission. This
emission is characterized by two emission line complexes typically
separated by 20 km/s. While the OH maser spectra suggest well defined
dynamical properties of the circumstellar material, the crucial
observations required to determine these properties have been lacking.
This thesis is primarily concerned with determining the radial velocity
of the central star and the spatial distribution of the maser emission
in the circumstellar material.

The radial velocity of the central star can be determined by
a statistical analysis of spectral line velocities in an ensemble
of stars. Such an analysis is presented in Chapter I for optical
emission and absorption lines and for radio OH maser emission lines.
The results indicate that, contrary to currently accepted theories,
the optical absorption line velocities and the high velocity OH
emission line velocities are red-shifted with respect to the stellar
radial velocity. This argues against models of the OH emission
which involve shock fronts or emission from the limb of a spherically
expanding circumstellar shell.

The spatial distribution of the maser emission can be determined
from radio interferometric observations. Spectral-line very long
baseline (VLB) interferometric observations of the OH maser emission
in long period variable stars such as IRC + 10011, U Ori and R Aql are
presented in Chapter II. The primary result of these observations is

that the apparent sizes of the OH maser components in long period variable stars are greater than or about 5×10^{15} cm. These large apparent sizes argue strongly against either gravitational collapse or rotation as the dominant dynamical process in the circumstellar envelope. Analysis of data obtained on IRC + 10011 and VY CMa (discussed in Chapter III) indicate that the two emission complexes probably are separated by distances less than, or comparable to, the extent of the emission in either complex. These and other findings suggest that OH emission from long period variable stars comes from a series of concentric, inhomogeneous, circumstellar shells expanding from the central star. The origin of these shells is probably a result of condensation of dust followed by radiative acceleration away from the star during the stellar light cycle.

A direct Fourier inversion of VLB data is presented in Chapter III for the 1612 MHz OH emission of VY CMa. This is the first attempt at such an analysis procedure for any VLB observations. This work demonstrates that Fourier inversion of spectral-line VLB data, even with very limited u-v coverage, is more efficient and less biased than direct model fitting for the resolution of complex source structures.

TABLE OF CONTENTS

FIGURES

FIGURES (Continued)

TABLES

Chapter I

The Stellar Velocity of Long Period Variables

and OH Maser Stars

I. INTRODUCTION

A rich array of spectral lines is observed in both the optical and microwave spectra of long period variable stars (Miras) with extensive circumstellar envelopes. The radial velocities of these lines are well determined observationally. However, the relationship of these velocities to the stellar radial velocity poses a significant problem, because these objects are almost never found in association with clusters of stars or as part of a binary system. Knowledge of the stellar radial velocity is one of the most important clues required to determine the structure of long period variables and their circumstellar shells. For example, the 18 cm maser emission from the hydroxyl radical (OH) observed from some of these stars is characterized by two emission complexes, separated by Doppler shifts of several tens of km/s. The current trend in the literature is to associate the higher velocity emission complex with the stellar radial velocity. This assumption imposes a severe constraint on dynamical models of the circumstellar envelopes. As a result, complex models have been developed and generally favored over simpler models.

Evidence that the high velocity OH emission complex is associated with the stellar radial velocity is due, primarily, to an optical study of 305 long period variables by Merrill (1941b). Merrill concluded that optical absorption line velocities (near maximum visible light) are randomly distributed about a mean near zero, and thus, correspond to the stellar velocity. In long period variables with OH maser emission, there is a strong correlation between the velocity of

the optical absorption lines and the high velocity OH emission complex. It is the purpose of this research to critically examine both the optical and the microwave data to determine what limits can be placed on the stellar radial velocity.

II. STATISTICAL ANALYSES

a) Techniques

The experimental data available to estimate the stellar radial velocity consist of velocities determined for many optical and some microwave lines. The radial velocity of a given line (corrected for the motion of the earth around the sun), v_ℓ, can be thought of as the sum of several velocities: a velocity due to the peculiar motion of the sun about a local standard of rest (LSR) at our position in the galaxy, $V_\odot$; a velocity due to the differential motion of two bodies in uniform circular orbits about the center of the galaxy, V_{gr}; a velocity due to the peculiar motion of the star about the LSR, v_*; and a correction term which relates the velocity determined from the spectral line to the velocity of the center-of-mass of the star, K. We hope to estimate this last term, the "K-correction," since it relates the measured line velocities to the stellar radial velocity.

The statistical procedure employed to determine the stellar radial velocity is to model the velocities determined for a group of stars from one type of spectral line as the sum of the terms discussed above:

$$v_\ell = V_\odot + V_{gr} + K + v_* \qquad . \qquad (1)$$

Given an ensemble of stars, we assume that the expectation value of v_* is zero. This assumption requires that there is no net radial

motion of the stars being analyzed about the LSR.[1] If v_* has a zero

---

[1]This assumption would be violated if, for example, a group of stars were expanding outward from the galactic plane. One could argue that a group of stars, which formed within the last $\sim 10^7$ years from interstellar materials, may have been accelerated outward from the galactic plane by a high density shock front in a spiral arm.

---

mean, we can treat the peculiar motion of each star as a noise term added on to the observed line velocities, and we have a well-defined statistical problem. Three parameters representing the solar motion, one parameter representing the differential galactic rotation (e.g. Oort's constant), and the K-correction can be estimated by minimizing the sum of the squares of the differences between the observed and modeled line velocities.

If the solar motion is represented by three cartesian galactic coordinates (Π, $\ominus$ and Z), differential galactic rotation effects by Oort's constant, (A), and the "K-correction" by K, we have a linear least-squares problem. Thus, for a given set of stars and spectral lines, a unique solution for any set of these parameters exists. Although we have significant a priori knowledge of the solar motion and Oort's constant, it is useful to simultaneously estimate these parameters to see possible correlations between the K-correction estimate and the adopted values of the other parameters. The following equations were used in the least-squares analysis for the velocity of the star due to the solar motion and differential

galactic rotation effects:

$$V_{\odot} = -\Pi \cos \ell \cos b + \Theta \sin \ell \cos b + Z \sin b$$

and

$$V_{GR} = A \, d \cos b \sin 2\ell$$

where d is the distance, ℓ the galactic longitude, and b is the galactic latitude of the star.

b) Radio Data

Table 1a presents a list of 28 stellar OH masers (i.e. type II OH/IR stars) analyzed in this section. Only those objects in which an optical or an IR identification (i.e. found in the CIT Two Micron Sky Survey, Neugebauer and Leighton, 1969) exists are analyzed. In practice, this tends to exclude distant objects ($\gtrsim$ 2 kpc) for which differential galactic rotation effects are large and whose distances are not known. A cursory examination of the last column of Table 1a reveals that the velocities of the high velocity OH complexes (with respect to a standard LSR) are almost entirely greater than zero. The mean of this group is +22 km/s $\pm$ 5 km/s (standard error of the mean). This non-zero mean velocity or K-correction is in strong disagreement with the theory that the high velocity OH complex is associated with the stellar radial velocity.

In order to more carefully investigate the K-corrections for this group of stars a least-squares analysis, as discussed above, of both the low and high velocity OH emission data was performed. The results of this analysis are presented in Table 1b. When a standard differential galactic rotation effect (Oort's constant, A = 15 km/s/kpc) and solar motion are assumed, the K-correction for the high velocity

TABLE 1a

OH Maser Stars

Name	l^{II}	b^{II}	Distance† (kpc)	Heliocentric Velocity*		LSR Velocity	
				L	H	L	H
IRC + 10011	129	-50	0.5	-7	30	-10	27
IRC + 50137	156	8	0.8	-12	21	-14	19
IRC - 20197	256	23	0.7	40	65	27	52
IRC + 40156	174	14	1.4	-22	2	-28	-4
IRC - 20540	15	-14	1.0	-4	23	8	35
IRC - 10529	36	-20	0.6	-46	-21	-32	-7
IRC + 40483	84	-10	1.1	15	43	30	58
IRC + 10365	41	8	0.5	-66	-15	-47	4
IRC + 30292	56	43	1.4	23	48	42	67
NML TAU	178	-31	0.3	29	63	17	51
WX SER	30	54	1.0	-18	-3	-2	13
R AQL	42	0	0.2	25	36	43	54
RR AQL	39	-16	0.5	8	17	23	32
V MIC	1	-46	0.5	-10	1	-4	7
R LMI	191	50	0.3	0	8	-3	5
RU HYA	323	31	0.7	-10	-3	-7	0
GY AQL	33	-16	0.5	11	22	28	39
R PEG	85	-45	0.4	16	23	22	29
U ORI	189	-2	0.2	-29	-23	-42	-36
W HYA	318	33	0.3	35	41	37	43
S CRB	49	57	0.3	-19	-13	-3	3
IRC - 20424	10	1	0.5	-12	5	1	18
R CAS	115	-11	0.3	14	21	22	29
U HER	35	40	0.3	-37	-29	-19	-11
NML CYG	80	-2	0.5	-40	6	-24	22
VY CMA	239	-5	1.0	8	72	-11	53
VX SAG	8	-1	0.5	-26	13	-14	25
PZ CAS	115	0	0.5	-75	-22	-64	-11

* L and H refer to the velocities of the outer edges of the low and high velocity emission complexes, respectively.

† Distances are from trigonometric parallax measurements by Wilson and Merrill (1942), and luminosity estimates by Hyland et al.(1972). The distances to PZ Cas, V Mic, and IRC-20424 were not available.

TABLE 1b

Statistical Analyses

Parameter*	Low Velocity OH Emission	High Velocity OH Emission	Correlation with K
K	-2 ± 7 [-3 ± 5]	20 ± 6 [19 ± 4]	1.0
Π	-2 ± 9	-11 ± 8	0.01
Θ	21 ± 11	20 ± 10	0.47
Z	10 ± 12	18 ± 11	0.18
A	11 ± 16	25 ± 14	-0.26

*
The K-correction, K, and the three parameters of the solar motion,

Π, Θ, and Z are in km/s. Oort's constant, A, is in km/s/kpc.

Solutions for the K-correction with a standard solar motion (i.e.,

Π = -10 km/s, Θ = 15 km/s, and Z = 7 km/s) and a standard Oort's

constant (A = 15 km/s/kpc) are given in brackets.

complex is 19 ± 4 km/s. If the solar motion and Oort's constant are estimated from the data, the K-correction does not change appreciably (K = 20 ± 6 km/s). The estimates of Π, Θ, Z, and A all are within one formal standard deviation of the standard values. Further, the correlations between K and Π, Θ, Z, and A are small. The largest correlation is only 0.47 between K and Θ (the component of the solar motion toward 90° galactic longitude, 0° latitude). This correlation is primarily a result of an observational bias against low declination stars which depopulates the sample in the region $180^\circ < \ell < 360^\circ$ with respect to those from $0^\circ < \ell \ 180^\circ$. Therefore, the significantly (> 3 formal standard deviations) non-zero K-correction for the high velocity OH complex argues strongly against the association of the stellar radial velocity with the high velocity OH complex.

c) Optical Data

The discovery in the radio data that the high velocity OH emission complex has a positive K-correction initiated an examination of the K-corrections for the optical lines. The correlation between the optical absorption line velocities and the high velocity OH emission complex noted by Wilson et al. (1970) seemed to imply that the traditional view that the optical absorption lines represented the stellar velocity (i.e. zero K-correction) was wrong.

Optical absorption lines have been associated with the stellar radial velocity for two reasons. First, the long period variable X Oph is reported by Merrill (1940) to be part of a binary system. Measurements on this binary indicate the radial velocity of X Oph is

the same as its absorption lines (Merrill, 1940). Although no uncertainties are mentioned, the time variability, both during the light cycle and among different cycles, of Mira absorption lines alone are ≈ 5 km/s (Merrill, 1941a) and must imply similar uncertainties for the radial velocity of X Oph. Second, a statistical study of the K-correction to the optical absorption lines of all long period variables with periods greater than 300^d indicates K-corrections of about 2 ± 3 km/s. However, only about one-third of the stars used in this analysis have measured absorption line velocities. In the majority of stars, the absorption line velocity is estimated by adding an empirically determined correction to the hydorgen emission line velocities (Merrill, 1941a). Therefore, evidence that optical absorption line velocities have no K-correction will be more critically examined below.

The optical line velocities used in this analysis are from Feast (1963). This survey of long period variables was chosen because, in addition to Merrill's sample, it includes a large number of stars in the southern hemisphere. Thus, an almost isotropic distribution of stars is analyzed which minimizes correlations of the K-corrections and solar and galactic motions. Only stars with periods longer than 300 days were analyzed. This excludes the high velocity halo population stars which are kinematically different from the longer period Miras. The K-corrections found for absorption and emission lines by analyzing many groupings of these stars are presented in Table 2. The most significant finding of this study is that, when only stars

TABLE 2

Optical Statistics on Mira Variables with Periods ≥ 300^d

Parameter*	Reported A-line Velocities (n = 58)			No Reported A-line Velocities (n = 109)		
	E-lines	A-lines	Correlation with K	E-lines	A-lines	Correlation with K
K	-5 ± 5 [-3 ± 4]	8 ± 5 [10 ± 4]	1.0	-16 ± 3 [-19 ± 3]	– –	1.0
Π⊙	-13 ± 7	-12 ± 7	0.22	-20 ± 5	–	-0.01
Θ	17 ± 7	15 ± 7	-0.20	23 ± 5	–	0.34
Z	9 ± 8	8 ± 8	-0.07	-4 ± 7	–	0.13
A	1 ± 8	1 ± 8	0.25	9 ± 3	–	-0.15

Parameter*	m_{pg} ≤ 9.5 (n = 54)			m_{pg} > 9.6 (n = 113)		
	E-lines	A-lines	Correlation with K	E-lines	A-lines	Correlation with K
K	-4 ± 4 [-3 ± 4]	9 ± 4 [9 ± 4]	1.0	-15 ± 3 [-18 ± 3]	-2 ± 3 [-5 ± 3]	1.0
Π⊙	-4 ± 7	-6 ± 7	0.22	-20 ± 5	-20 ± 5	0.12
Θ	18 ± 7	16 ± 7	-0.05	26 ± 5	26 ± 5	0.20
Z	16 ± 7	15 ± 8	0.00	-4 ± 7	-4 ± 7	-0.07
A	1 ± 9	2 ± 9	0.02	8 ± 3	8 ± 3	-0.04

Parameter*	$|b^{II}|$ > 30° (n = 63)			$|b^{II}|$ < 30° (n = 106)		
	E-lines	A-lines	Correlation with K	E-lines	A-lines	Correlation with K
K	-7 ± 4 [-8 ± 4]	6 ± 4 [5 ± 4]	1.0	-15 ± 3 [-16 ± 3]	-2 ± 3 [-3 ± 3]	1.0
Π⊙	-16 ± 8	-16 ± 8	0.23	-17 ± 5	-16 ± 5	0.13
Θ	30 ± 7	29 ± 8	0.11	24 ± 5	24 ± 5	0.12
Z	2 ± 5	1 ± 5	-0.02	12 ± 12	13 ± 12	-0.11
A	-1 ± 4	0 ± 4	0.02	11 ± 3	10 ± 3	-0.04

* see footnote on Table 1b

with reported absorption lines are analyzed, the K-correction for the absorption lines is 10 ± 4 km/s. The correlations between the K-correction and the solar motion and differential galactic rotation parameters are very small. This indicates that the positive K-correction for the absorption lines is not likely to be a result of the distribution or distances of the stars sampled.

It is interesting to note that the K-corrections for the emission lines appear different for long period variables when they are grouped by whether or not they have reported absorption line velocities. The emission line K-correction for long period variables with reported absorption lines is -3 ± 4 km/s, while for those with no reported absorption lines it is -19 ± 3 km/s. This explains the results of Merrill (1941b) that absorption lines for the entire group of long period variables have no significant K-correction, since the majority of absorption line velocities used by Merrill (1941b) are not from observations, but are inferred by adding a correction of about 13 km/s to the emission line velocities.

The difference between the emission line K-corrections for long period variables which are grouped by whether or not absorption line velocities are reported is surprising. There is no obvious physical distinction between stars in which absorption line velocities are measured or not. In general, absorption lines are seen in all long period variables and the only criterion for measurement of absorption line velocities seems to be the apparent brightness of the star

(Preston, 1975). Therefore, it is not surprising that the emission line K-corrections of long period variables grouped by apparent magnitude show similar differences as those grouped by the report of absorption line velocities (see Table 2). Similarly, since apparent brightness is correlated with distance and distance is statistically correlated with galactic latitude, stars grouped by galactic latitude (see Table 2) also show differing emission line K-corrections. Unless Mira variables have formed within the last $\sim 10^7$ years, it is unlikely that the emission line K-correction differences are due to galactic structure.

III. DISCUSSION

The primary contribution of this chapter toward the understanding
of the structure of long period (Mira) variables and stellar masers is
that the traditional theory that the stellar radial velocity coin-
cides with the optical absorption line velocities is not well founded.
This theory is primarily based upon the statistical determination
of a near zero absorption line K-correction. In order to use all
known Mira variables with periods greater than 300 days in a statisti-
cal analysis, it is necessary to estimate absorption line velocities
which are not measured for the majority of these stars. Thus, the
absorption line K-correction determined from all Miras is based upon
the assumptions that the stars are both kinematically and physically
homogeneous. Kinematic homogeneity is required because the same
emission line K-correction is assumed for all stars. Physical homo-
geneity is required because absorption line velocities are estimated
from emission line velocities by knowing this difference in only
about 1/3 of the stars.

The statistical re-analysis of the emission line K-corrections
for Miras with periods greater than 300 days presented in this chapter
casts strong doubt on the validity of the assumption of kinematic
homogeneity of this group of stars. It is not clear if this is due to
a physical or evolutionary diversity among the stars analyzed, or if
systematic observational biases have entered into the measurements of
the emission line velocities. Whatever the reason for the difference
between the emission line K-corrections of stars in which absorption

line velocities are reported or not, absorption line K-corrections should only be obtained from those stars in which the lines are actually measured. When this is done, the absorption line K-correction appears to depart significantly from zero. The value of + 10 km/s is adopted as the best estimate of the absorption line K-correction. It is determined by assuming a standard solar motion and Oort's constant (see Table 2). The most reasonable estimate of a formal one standard deviation uncertainity is 5 km/s which takes into account correlations of the K-correction with parameters of the solar motion and differential galactic rotation. This result suggests that the measured absorption line velocities (near maximum light) of long period variables with periods greater than 300 days are red shifted with respect to the stellar radial velocity.

While there are many problems associated with determining K-corrections for optical lines formed in long period variables, it may be possible to determine the stellar radial velocity from OH maser line velocities. The OH velocities show no time variability which makes them far more suitable as possible indicators of the stellar radial velocity than the highly variable optical lines. At present, however, the number of stellar OH masers which have been identified with nearby visible or infrared stars is small, which makes a statistical analysis difficult. It seems that in the future when more sensitive microwave systems are in operation , the number of known nearby stellar OH masers may grow.

With the presently available sample of identified stellar OH masers, it is still possible to make some significant observations. The K-correction for the high velocity OH emission complex estimated by assuming a standard solar motion and Oort's constant is + 19 km/s. A formal one standard deviation uncertainty for this estimate, which takes into account the correlations between the K-correction and the solar motion and galactic rotation law assumed, is about 6 km/s (see Table 1). Therefore, the high velocity OH emission complex appears red shifted from the stellar radial velocity.

The simplest model of the OH maser emission from Mira variables is that the two emission complexes originate from the approaching and receding sides of an expanding circumstellar shell (see chapter II). A more complex asymmetric model which involves emission only from the approaching side of an expanding shell, but from in front and back of a stationary shock-front, has been discussed by Wilson and Barrett (1972) and Dickinson and Chaisson (1973). In addition, a model for the super giant source NML Cyg in which the high velocity OH emission arises from the limb of an expanding shell has been proposed by Masheder et al. (1974). Both of these models have been proposed because the stellar radial velocity was assumed to be given by the optical absorption lines and, therefore, by the high velocity OH emission. The significantly positive K-correction found for the high velocity OH emission complex in stellar OH masers conflicts with this assumption, and the need for such complex models is apparently removed.

REFERENCES

Dickinson, D.F. and Chaisson, E.J. 1973, Ap.J. (Letters), 181, L135.

Feast, M.W. 1963, M.N.R.A.S., 125, 27.

Hyland, A.R., Becklin, E.E., Frogel, J.A. and Neugebauer, G. 1972,

  Astron. & Astrophys., 16, 204.

Masheder, M.R.W., Booth, R.S., and Davies, R.D. 1974, M.N.R.A.S., 166,

  561.

Merrill, P.W. 1940, Spectra of Long Period Variables (Chicago:

  University of Chicago Press).

_____ 1941a, Ap.J., 93, 380.

_____ 1941b, Ap.J., 94, 171.

Neugebauer, G., and Leighton, R.B. 1969, Two-Micron Sky Survey: A

  Preliminary Catalogue (Washington: NASA).

Preston, G. 1975, personal communication.

Wilson, R.E. and Merrill, P.W. 1942, Ap.J., 95, 248.

Wilson, W.J., and Barrett, A.H. 1972, Astr. Astrophys., 17, 385.

Wilson, W.J., Barrett, A.H., and Moran, J.M. 1970, Ap.J., 160, 545.

Chapter II

Spectral Line Very Long Baseline Observations

of OH Maser Emission from Mira Variables

I. INTRODUCTION

Since the discovery of astronomical masers, intense efforts have been directed toward understanding the physical processes responsible for the maser radiation and the environment in which the maser exists. Molecular masers have been observed imbedded in HII regions, around super-nova remnants, and in the circumstellar material of long period variable (Mira) stars. The masers associated with Mira stars are particularly amenable to analysis because of the great amount of observational data which have been accumulated in the visible, infrared, and microwave regions of the electromagnetic spectrum, and because of the stringent boundary conditions imposed on models of these objects by their close association with cool, dusty stars.

One of the most striking features of the masers associated with Mira variables, henceforth called stellar masers, is the strong 1612-MHz emission of the hydroxyl radical (OH). This radiation is observed as two complexes of emission lines separated by Doppler shifts of ~ 30 km/s. Most stars with OH masers have other masing molecules such as H_2O and SiO. Usually the Doppler velocities of the H_2O and SiO masers fall within the velocities of the two OH maser complexes. The regularity of the pattern of the maser velocities suggests that there are well defined dynamical properties of the circumstellar material common to these stars. Many dynamical models of stellar masers have been proposed but crucial observations have been lacking. The radial velocity of the central star and the spatial distribution of the circumstellar material must be determined to evaluate models involving general expansion, contraction, or

rotation of the circumstellar material. The results of the statistical
determination of the stellar radial velocity presented in Chapter I,
coupled with the spatial distribution of the OH masers around the central
star described in this chapter, strongly argue against contraction or
rotation as the dominant dynamical structure in the circumstellar mate-
rial.

The two complexes of OH emissions are Doppler shifted because the
masers probably emanate from the approaching and receding sides of cir-
cumstellar material expanding from a central star. While the general
regularities in the pattern of the maser emission can be explained in
this fashion, the detailed structure of the OH maser spectra appears to
be dominated by irregularities in the circumstellar material.

II. OBSERVATIONS AND DATA REDUCTION

This chapter presents the results of two spectral-line very long baseline (VLB) interferometric observations of stellar OH masers. The first experiment was conducted with the Owens Valley 130-foot (40 m) and the Hat Creek 85-foot (26 m) telescopes from December 2-5, 1974. The second experiment was conducted with the Greenbank 140-foot (43 m) and the Maryland Point 85-foot (26 m) telescopes from January 13-15, 1975. A description of the performance characteristics of the equipment at each station is given in Table 1. All data was recorded with standard Mk II VLB terminals and later cross-correlated on the NRAO Mk II VLB processor in Charlottesville, Virginia between December 1974 and March 1975. The post-processor data reduction consisted of several stages. The first involved time averaging the data and Fourier transforming the cross-correlation function into a "cross-correlated power spectrum." This was done on the NRAO IBM 360 computer and required approximately 1 hour of computer time for each hour of observation. The time sequence of data in each channel of the cross-correlated power spectrum was analyzed to determine the complex fringe visibility function either by a Fourier analysis or by least-squares fitting of the fringe frequency spectrum. This data analysis was done on the Caltech IBM 370 computer with a series of programs (described in Appendix I) which represent the cumulative efforts of J. Moran, J. Romney, K. Johnston and myself. The resultant fringe amplitudes and phases as a function of time and spectral frequency (or velocity) were modeled to determine the relative positions and sizes of different spectral components.

TABLE 1

Observational Parameters

	Owens Valley	Hat Creek	Maryland Point	Greenbank
Telescope diameter	40 m	26 m	26 m	43 m
Aperture efficiency	~ 0.6	~ 0.6	~ 0.6	~ 0.6
Receiver	paramp	paramp	paramp	cooled paramp
System Temperature	85°K	120°K	220°K	65°K
Polarization	linear 90° P.A.	linear 90° P.A.	RCP	90° linear/RCP
Frequency Standard	Rubidium	Rubidium	Rubidium/H maser	H maser

	Owens Valley to Hat Creek baseline	Maryland Point to Greenbank baseline
Observing frequencies	1612, 1665, 1667 MHz	1612 MHz
Polar length	~ 309303 m	~ 6030 m
Equatorial length	372827 m	227571 m
West longitude	252°.135	169°.476

At present there is a calibration problem which affects data obtained with the old 96 channel Mk II VLB processor and data reduction programs. Two sources, U Ori and W 43a, yield normalized fringe visibility amplitudes of ~ 1.6. When the U Ori data was reprocessed on the new 288 channel Mk II VLB processor and data reduction programs, a normalized fringe visibility amplitude of unity was obtained (see Fig. 4). Thus, it is possible that some or all of the old Mk II processed data may have normalized fringe visibility amplitudes which are a factor of 1.6 too high. This affects the sources IRC+10011, R Aql, OH 26.5, OH 1821-12, W 43a and VX Sgr. At the present time it is unclear if the calibration problem is due to a processor or software bug. This problem does not affect fringe phase determinations or any data obtained from the new Mk II VLB processor on the sources U Ori and VY CMa. Further, none of the conclusions reached in this research would be altered by such a calibration problem. The ultimate resolution of this problem may require a full simulation of a spectral line source and its analysis on the Mk II VLB processor and data reduction programs.

III. 1612-MHz OBSERVATIONS

a) IRC+10011

IRC+10011 (CIT 3) is one of the most important stellar OH masers because of the striking symmetry of its 1612-MHz emission complexes. This source had been observed in a previous experiment from Owens Valley to Hat Creek (Reid and Muhleman, 1975) where an upper limit to the cross-correlated flux density of 2 Jy implied that the apparent angular sizes of both emission complexes are $\gtrsim 0\overset{''}{.}07$ (FWHM circular gaussian). IRC+10011 was observed for 15 hours over two days with the Greenbank to Maryland Point baseline; nine hours were lost when one of the stations recorded data on the wrong side of the tape. The fringe amplitude, although extremely weak, was detectable. Figure 1 displays the total power (single dish) and average cross-correlated power spectrum. Figure 2 shows the fringe amplitude and the relative fringe phase of the -9 and +26 km/s components as a function of time. The time derivative of the relative fringe phase is indicated by the slope of the line passing through each phase point. The fringe amplitude of both components appears to stay within one turn (360°) over this time period. However, since each relative fringe phase point has integral turn ambiguities, the proper "phase connection" is not entirely certain. Serious "phase connection" problems occur at interferometer hour angles (IHA) of $0^h 45^m$, $2^h 15^m$, and $4^h 0^m$. The preferred phase connection is indicated by the data points with error bars. Alternative values differing by one turn are indicated without error bars.

IRC+10011   1612 MHZ
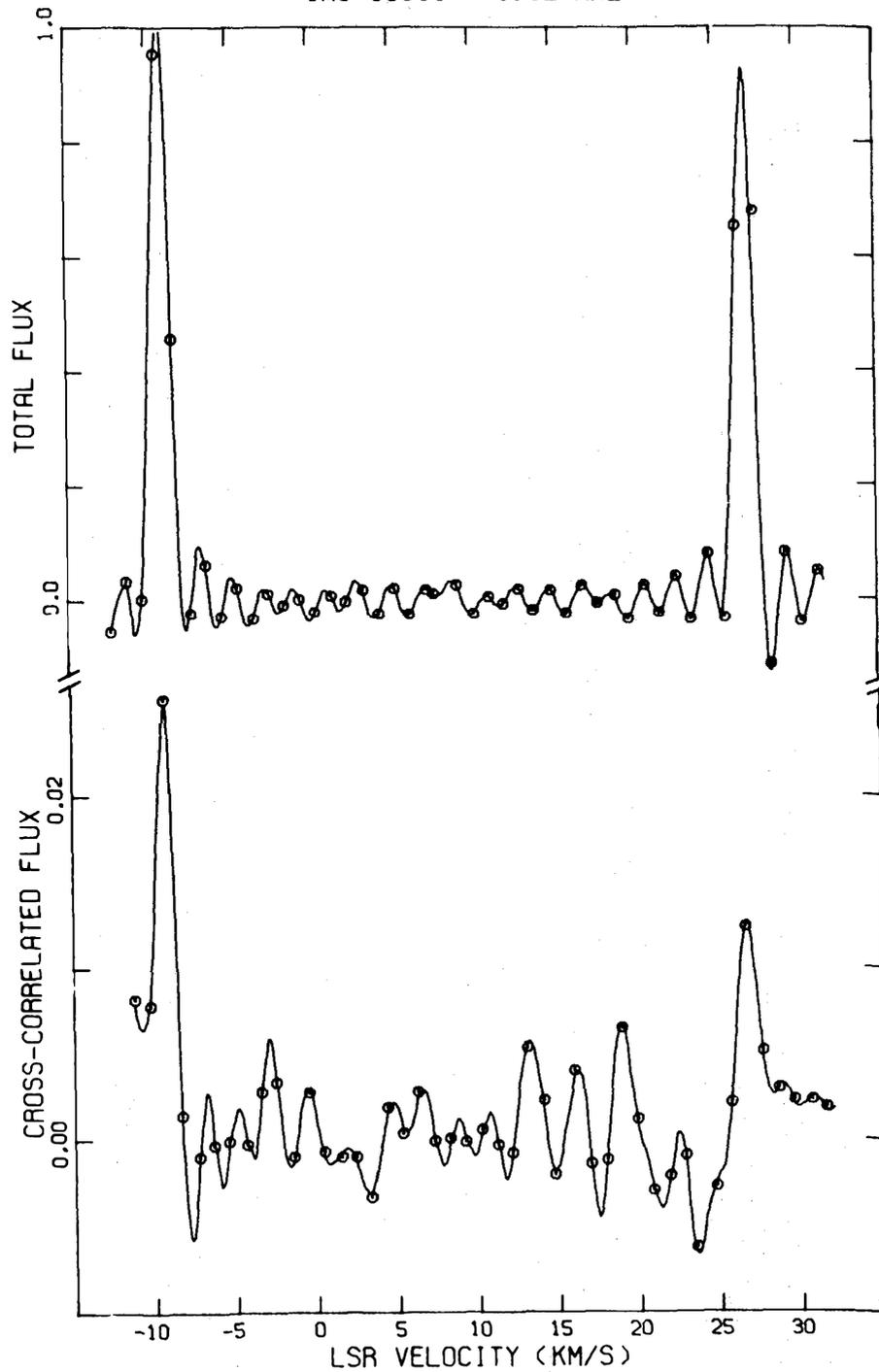
Figure 1 - The single telescope spectrum is plotted above the interfero-
spectrum for IRC+10011. The ratio of the cross-correlated flux to the
total flux at any velocity is the normalized fringe visibility. The inter-
ferometer spectrum is the sum of spectra for all scans on the Greenbank to
Maryland Point baseline. The single telescope spectrum shows considerable
"ringing" expected for narrow spectral lines and a uniform weighted spectrum.
The peak total flux of 1.0 corresponds to about 30 Jy. Note the 50 to 1
scale difference between the single telescope and interferometer spectra.
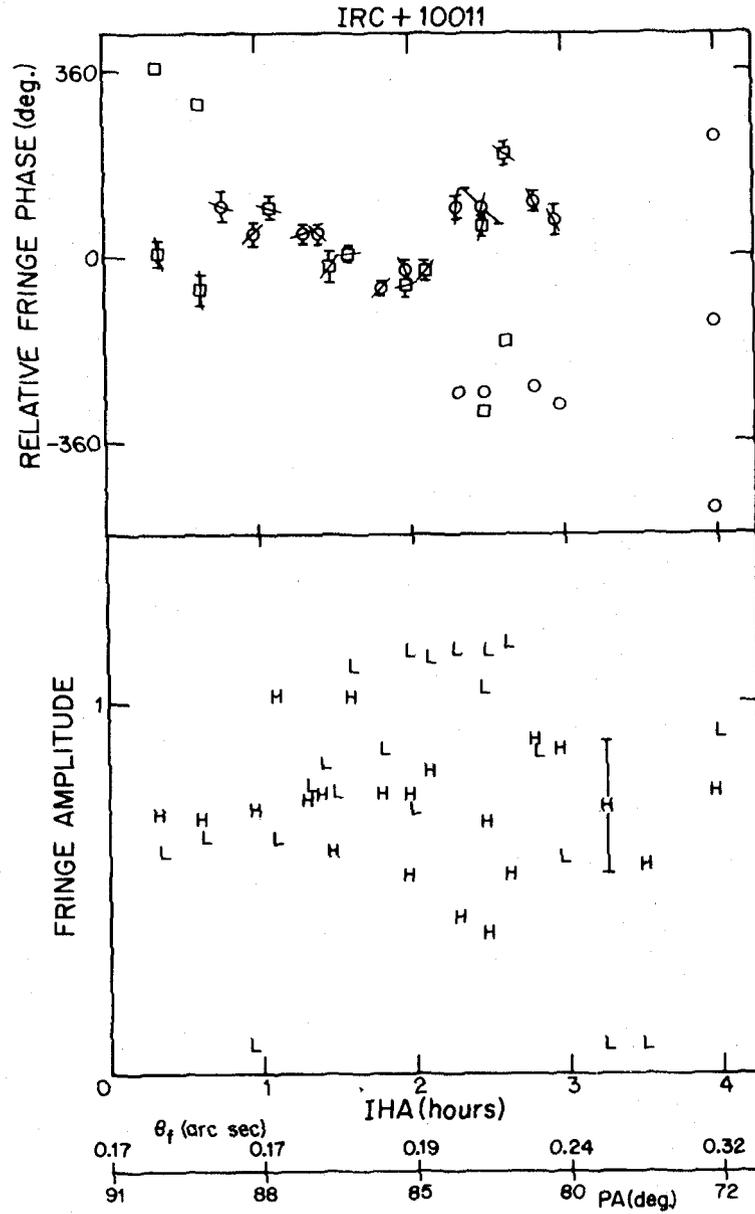
Figure 2 - The fringe phase of the 26 km/s component minus the -9 km/s component is plotted as a function of time above the fringe amplitudes H and L, of the high and low velocity peaks respectively. Alternate phase points, 360° from the "best values," are indicated without error bars. Circles and squares represent observations on different days. The slopes of the lines through the phase points give the relative fringe rates.

It is difficult to analyze such low signal-to-noise data with possible ambiguities present in the relative fringe phases. However, one can safely conclude from the low fringe amplitudes that the dominant components in the low and high velocity complexes are larger than about 0".5 (FWHM). Further, the fact that the fringe amplitudes did not increase as the projected baseline length shortened by a factor of two indicates a complex structure. One structure which is consistent with the low, non-varying fringe amplitudes is a "core-halo" whose small weak core ($\leqslant$ 0".05) is detected and whose large strong halo ($\geqslant$ 0".5) is fully resolved. If the preferred phase connection (given by the data points with error bars in Fig. 2) is accepted, the spatial separation of the components detected at -9 and +26 km/s is less than $\sim$ 0".2. The most extreme possible variation of the relative fringe phase is about two turns over the four hours of observation (i.e. from +360° at 0^h IHA to -360° at 4^h IHA) which places a very conservative upper limit of $\sim$ 1" on the separation of these components.

b) U Orionis

U Ori is the most peculiar and one of the most interesting stellar OH masers. It was reported by Wilson et al. (1972) to be a main line (1665/7-MHz) HO maser with no detectable 1612-MHz emission greater than 0.5 JY. In 1975, Pataki and Kalena reported strong 1612-MHz emission from U Ori. The main line emission occurs near -42 and -36 km/s, and has weakened since 1972 (Cimerman, 1975). This is the only case in which a dramatic change has been reported for a stellar OH maser. U Ori also has maser emission from H_2O and SiO (see Kaifu, Buhl and Snyder, 1975).

U Ori was first observed for a short period of time on the Owens

Valley to Hat Creek baseline. These observations indicated that U Ori

had components which were at least an order of magnitude smaller in

apparent size than those of most stellar OH masers. More extensive

observations were carried out on the Greenbank to Maryland Point base-

line. The first 90 minutes of observation at Greenbank were with linear

polarization at 90° P.A. (due to a temporary malfunction of one of the

two receiver channels), while the remaining observations were at right

circular polarization (RCP). The total power spectra taken with the

Greenbank telescope at these two polarizations is shown in Fig. 3. These

spectra indicate that the -39.4 km/s component is strongly ($\gtrsim 50\%$) polar-

ized and that the -45.5 km/s component is, apparently, also polarized.

This is the first report of a 1612-MHz maser with substantial polariza-

tion. Recent observations by Cimerman (1975) substantiate this

discovery and indicate that the - 39.4 km/s component is almost

100% right circularly polarized.

Figure 4 displays the total power and average cross-correlated power

spectrum (RCP) for U Ori. At any velocity, the ratios of the cross-

correlated power to the total power gives the normalized fringe visibil-

ity. Figure 5 shows the fringe amplitude of various spectral components

as a function of time. The fact that the fringe amplitudes of all com-

ponents detected in cross power did not increase as the projected base-

line length decreased by a factor of two implies that the 1612-MHz emis-

sion of U Ori is probably composed of many small maser components

($\lesssim 0\overset{''}{.}05$). In this case, the difference between the total power and

4
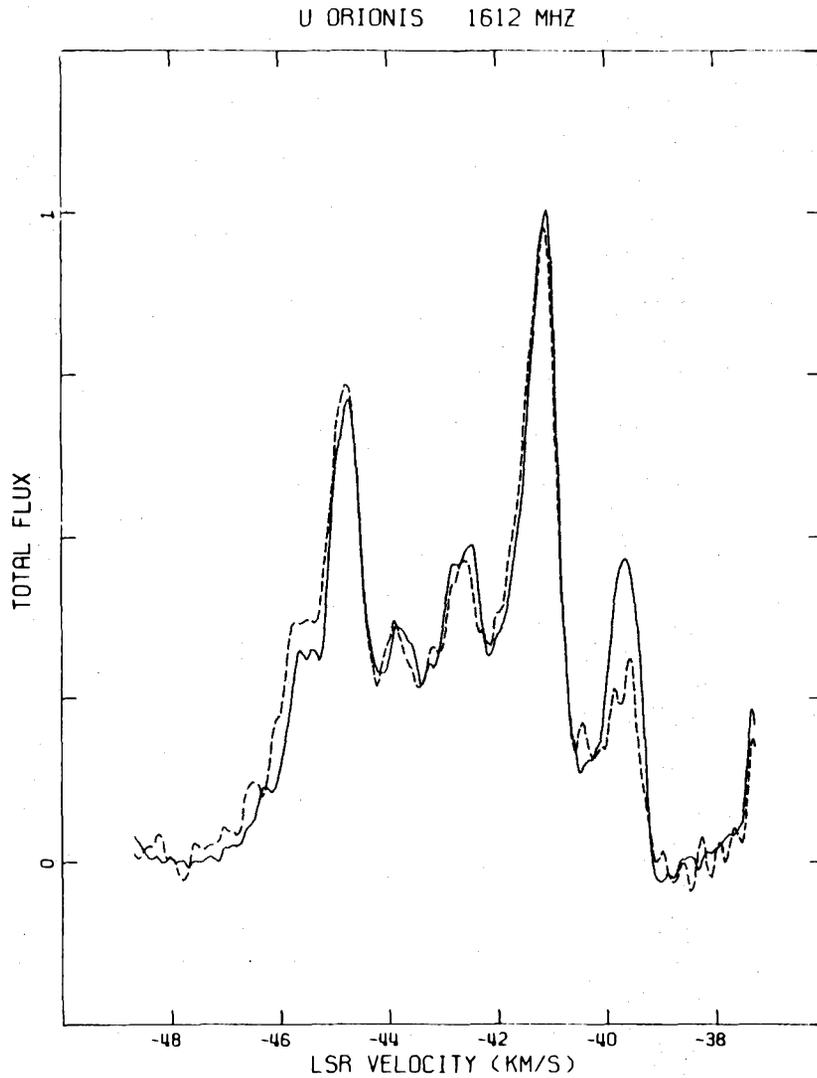
Figure 3 - Single telescope spectra in differing polarizations obtained
with the Greenbank 140' telescope. The solid line spectrum is right
circular polarization and the dashed line spectrum is (east-west) linear
polarization. A 144 pt. cosine weighted autocorrelation function was
transformed which gives a spectral resolution of 0.16 km/s. Note the
differences between the spectra at -39.4 and -45.5 km/s which indicate
a significant degree of polarization. The peak total flux of 1.0
corresponds to about 25 Jy.

U ORIONIS 1612 MHZ
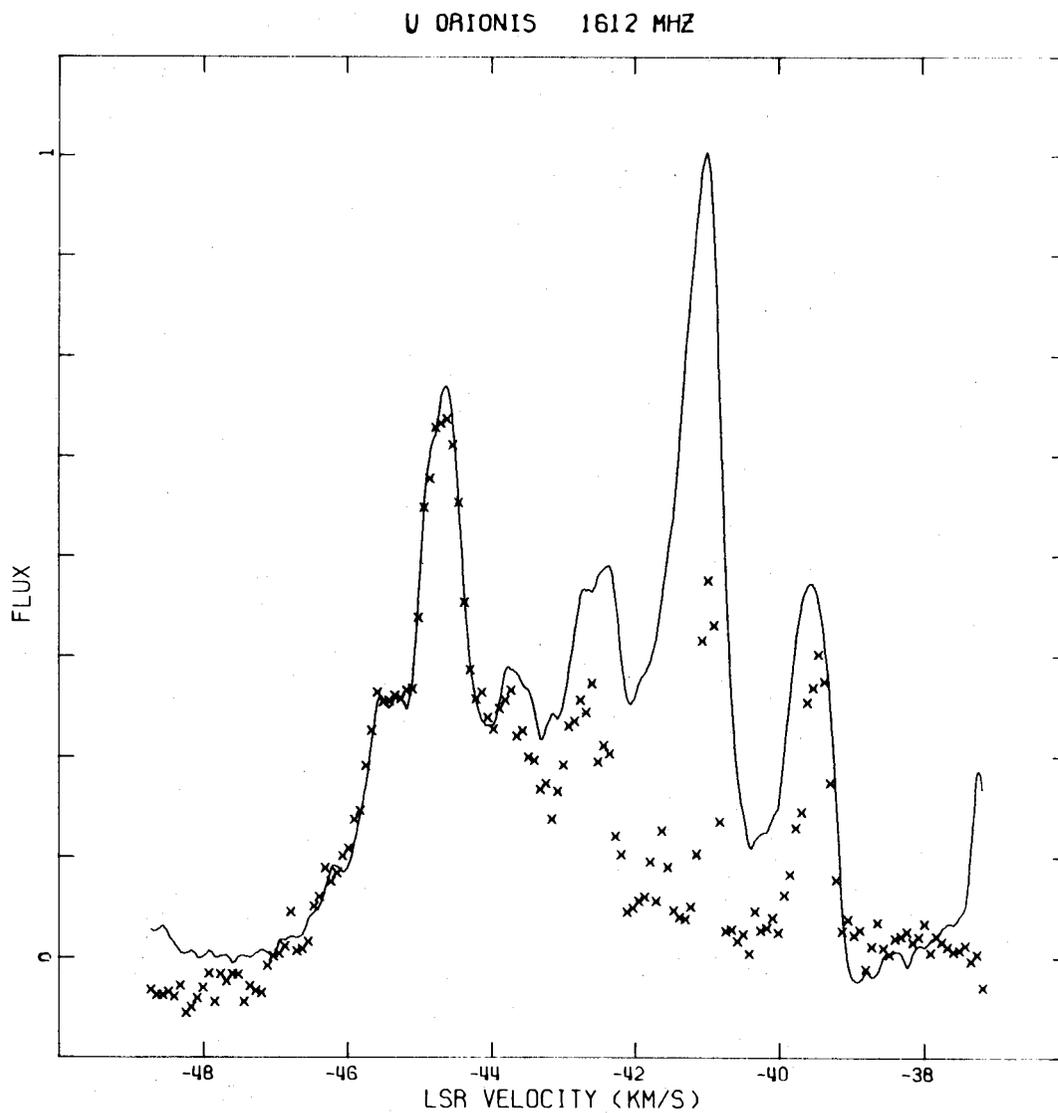
Figure 4 - The single telescope spectrum (solid line) and the interferometer spectrum (crosses) for the RCP 1612 MHz emission of U Ori. The ratio of the interferometer spectrum to the single telescope spectrum at any velocity gives the normalized fringe visibility. The spectral resolution is 0.10 km/s. Note the very narrow feature at -41.0 km/s in the interferometer spectrum.
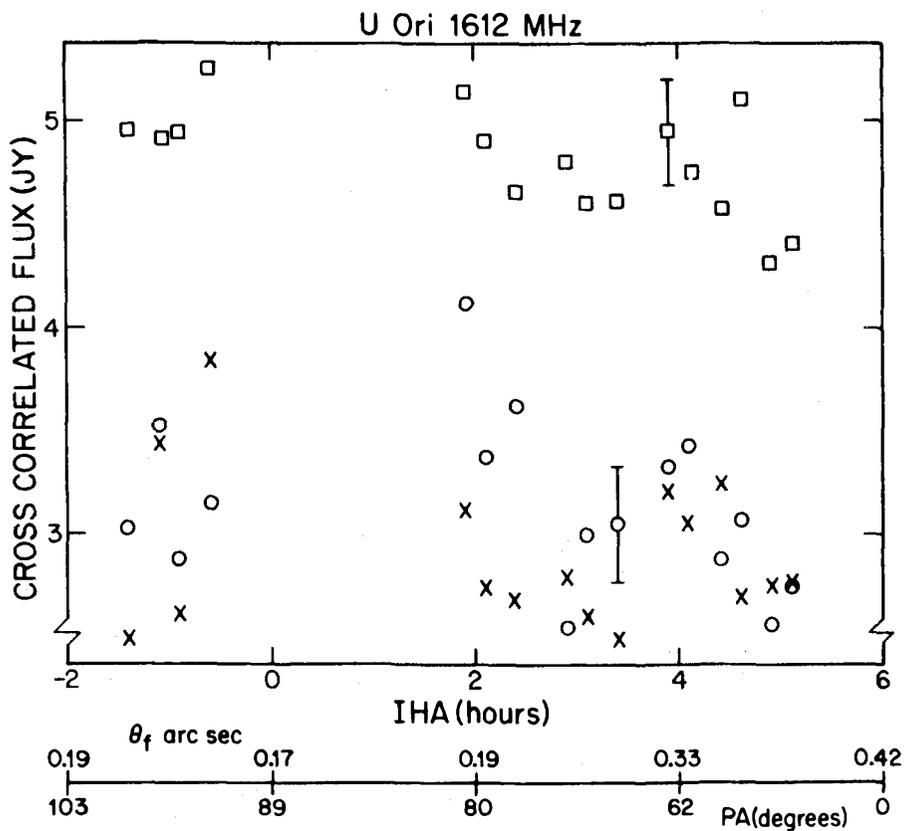
Figure 5 - The interferometer power in three spectral channels as a function of time for the Greenbank to Maryland Point baseline. The squares, circles, and crosses correspond to -44.8, -41.0, and -39.5 km/s spectral components. Note the absence of significant time variations of the interferometer power as the interferometer spacing, θ_f, and position angle, PA, change. This is best explained if the interferometer power comes from small, "unresolved" spectral components.

cross-correlation power must be accounted for by many large ($\gtrsim$ 0".3)
components. In addition, most spectral features appear to be composed
of two or more narrow spectral components. This impression is supported
by the existence of a very narrow, isolated emission line at - 41 km/s.
This line has a half-width at half-power of 650 Hz, which has been
slightly broadened by the spectral channel width of 200 Hz and by a
Doppler shift of $\approx$ 100 Hz caused by the earth's rotation over the
half-long scans. The true line width is less than 600 Hz or 0.11 km/s.
This is one of the narrowest lines ever observed in emission at 18 cm.
If the line is not narrowed by the maser process, it requires kinetic
temperatures of the gas of $\approx 5^\circ$K.

The spatial separations of the spectral components of U Ori were
determined by fitting a position offset for each component to the fringe
phase data. The spectral channel at −44.8 km/s was used as a phase
reference or calibrator, and an offset in right ascension and declina-
tion was determined relative to this maser component. A position offset
was determined by modeling the relative fringe phases, $\Delta\Phi$, by the equa-
tion

$$\Delta\Phi = 2\pi \ (u \ \cos \ \delta_s \Delta \ \alpha \ + v \ \Delta\delta \ ) \qquad \text{(radians)}$$

where δ_s is the source declination, and u and v the projected baseline
lengths (wavelengths) along the east-west and north-south directions.
The right ascension, $\Delta\alpha$, and declination, $\Delta\delta$, offsets from the reference
feature position were adjusted to minimize the sum of the squares of the
observed minus model relative phases. Relative fringe phases determined
by differencing the phase of two spectral channels are free of noise from

imperfect frequency standards and atmospheric propagation effects, and free of most systematic problems such as baseline or absolute source position errors. Typical relative fringe phases for U Ori had less than 5° uncertainty. The map of the 1612-MHz emission of U Ori determined from relative fringe phases obtained with the Greenbank to Maryland Point baseline is shown in Fig. 6. As an example of the quality of the fits, the data and model relative phases are displayed for three channels with widely differing position offsets in Fig. 7a.

The map of U Ori can be characterized as three clusters of emission separated by about 30 AU. (The distance to U Ori determined by trigonometric parallax measurements is 190 pc (Wilson & Merrill, 1942).) It is interesting to note that if one draws lines connecting the three clusters of emission, they closely resemble a 45° right triangle aligned along the east-west and north-south axes. This is almost certainly coincidental (unless it is the work of extraterrestrial intelligence), and the coincidence is not as remarkable as it may appear. The chance of three random points forming some regular geometric shape is reasonably large (~10%). It is almost impossible for any symmetric errors in the data reduction to produce such effects since all of the data were taken within the same observing bandpass and reduced simultaneously. As an internal check on the model fitting program, the -39.5 km/s component was tested as a phase or position reference and the relative positions of the components were not changed.
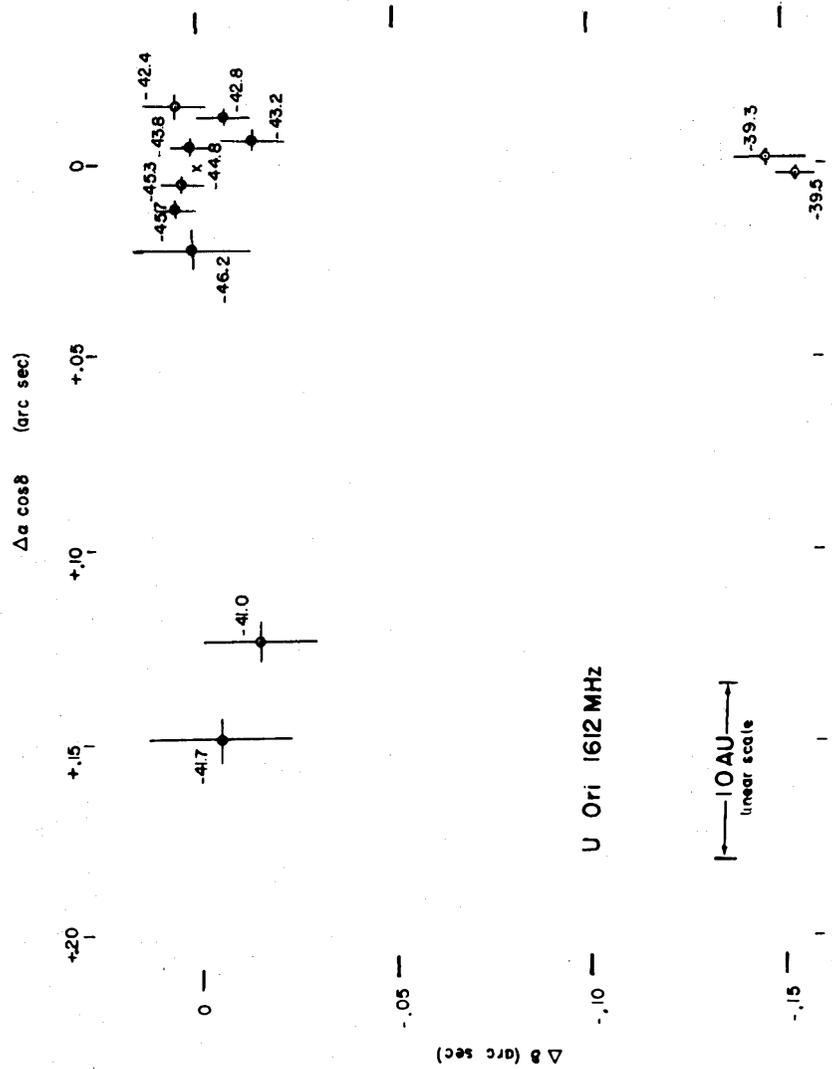
Figure 6 - The relative positions of the spectral components of the 1612 MHz emission of U Ori determined by modeling the relative fringe phases obtained from the Greenbank to Maryland Point baseline are shown. The reference component is at -44.8 km/s. Error bars are ±2 formal standard deviations and the correlation coefficient between the right ascension and declination offsets is -0.45.
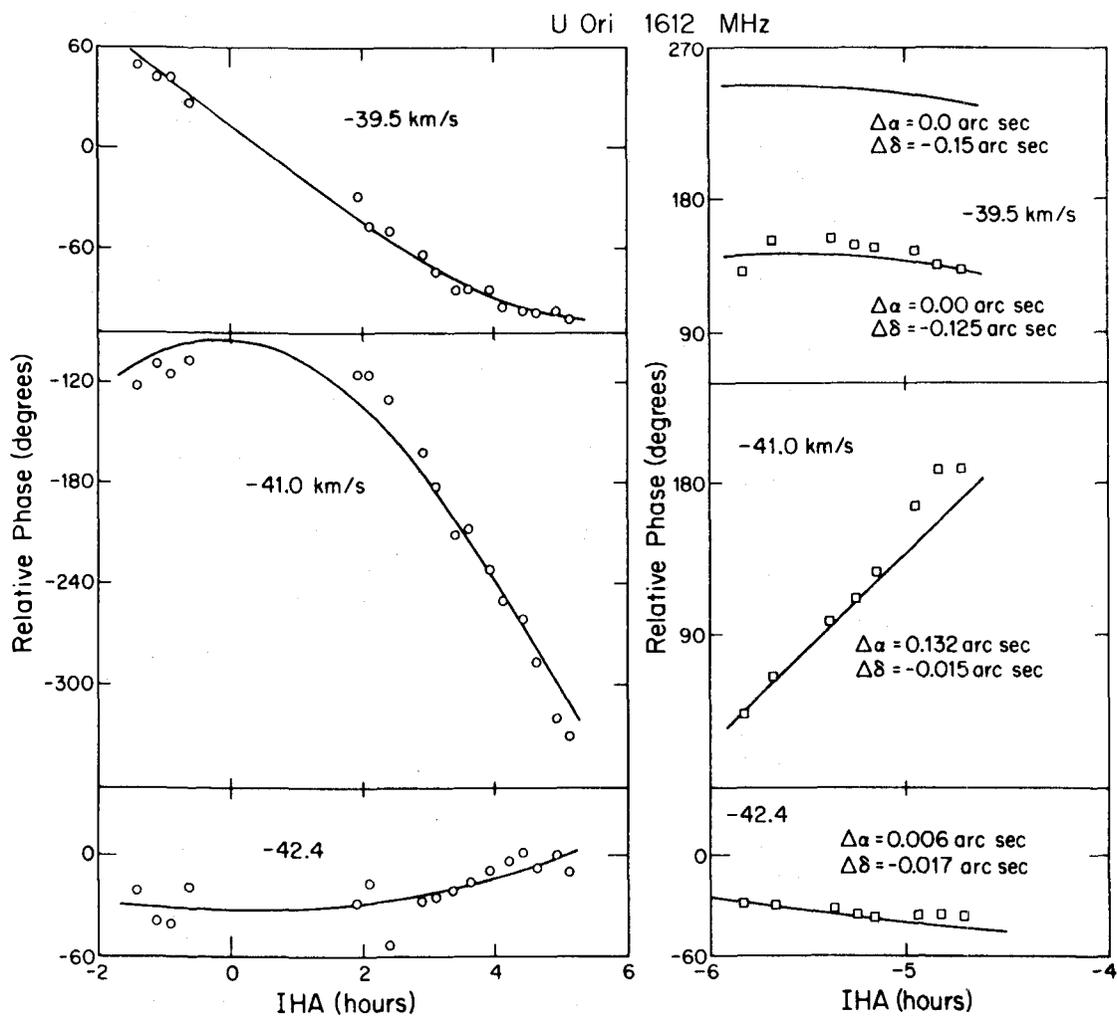
Figure 7a - The relative fringe phases as a function of time of three representative spectral components obtained with the Greenbank to Maryland Point baseline are plotted with circles. The least squares fits to these phases are indicated by the solid lines.

Figure 7b - The relative fringe phases as a function of time of the three spectral components in (a) obtained with the Owens Valley to Hat Creek baseline are plotted with squares. The solid lines indicate the phases predicted by the position offsets determined with the Greenbank to Maryland Point baseline and used in the map shown in Figure 6.

As a final check of the relative fringe phase map of U Ori (Fig. 6),
the position offsets determined from the Greenbank to Maryland Point
baseline data were used to predict the relative fringe phases expected
for the one and one-half hours of data taken one month earlier with the
Owens Valley to Hat Creek baseline. These relative fringe phases are
shown in Fig. 7b. The -41.0 km/s component phases from the Owens Valley-
Hat Creek baseline agree exceedingly well with the offset of 0.''13 in the
right ascension and -0.''015 in declination determined from the Greenbank
to Maryland Point baseline data. The -39.5 km/s component phases from
the Owens Valley to Hat Creek baseline indicate a pure declination off-
set, but do not agree with the -0.''150 declination offset determined from
the Greenbank to Maryland Point baseline. Instead, it appears to be

$$-0.125 \pm .09N \qquad \text{for} \qquad N = 0, \pm 1, \pm 2, \ldots \ldots ,$$

which represents integral fringe spacing ambiguities present due to the
lack of north-south motion of the projected baseline during the observa-
tions. There are several possible explanations for the discrepancy
between the -0.''15 offset of the -39.5 km/s component determined from the
Greenbank to Maryland Point data and the -0.''125 offset implied from the
Owens Valley to Hat Creek data. First, the difference between the
clocks at the two telescopes was not precisely calibrated for the Owens
Valley to Hat Creek baseline. This can introduce a constant relative
phase offset (see the discussion of the 3rd term in eq. (A-4) in the
Appendix) which would exactly mimic a declination offset for the very
limited U-V coverage obtained from this baseline. This could account
for the apparent 0.''025 discrepancy between the data from the two

baselines. Second, the two experiments were over one month apart, and time variability of the source structure on a scale of $\sim$ 0".025 cannot be ruled out. Thus, the detailed agreement of the relative fringe phases for the two baselines of the -41.0 km/s component, and the qualitative agreement for the -39.5 km/s component, strongly supports the map of U Ori displayed in Fig. 6.

c) OH 26.5 + 0.6

OH 26.5 is a very strong stellar OH maser and infrared source at 20 microns (Evans, 1975). No optical identification has been made, probably because of the very thick dust shell surrounding the central star. The strength and width of its low and high velocity emission complexes closely resemble the "super-giant" stellar OH masers - NML Cyg and VY CMa - while the 30 km/s separation of these complexes more closely resembles a normal Mira OH spectrum. It is difficult to classify it either as a new type of stellar OH maser or as an intermediate form between the Mira (giants) and the "super-giants."

OH 26.5 was detected with the Greenbank to Maryland Point baseline with a fringe spacing of 0".4 (see Fig. 8). The fringe visibilities of the strongest components in each of the emission complexes were $\sim$ 0.05, indicating apparent sizes of $\sim$ 0".4. Thus, the apparent angular sizes of the strongest components are similar to those of the Mira type, and significantly larger than the "super-giant" sources (although the distances to the super-giants and OH 26.5 are not known well enough to compare linear sizes).

OH 26.5   1612 MHZ

TOTAL FLUX

1.0

0.04 0.0

CROSS-CORRELATED FLUX

0.00

10   15   20   25   30   35   40   45

LSR VELOCITY (KM/S)
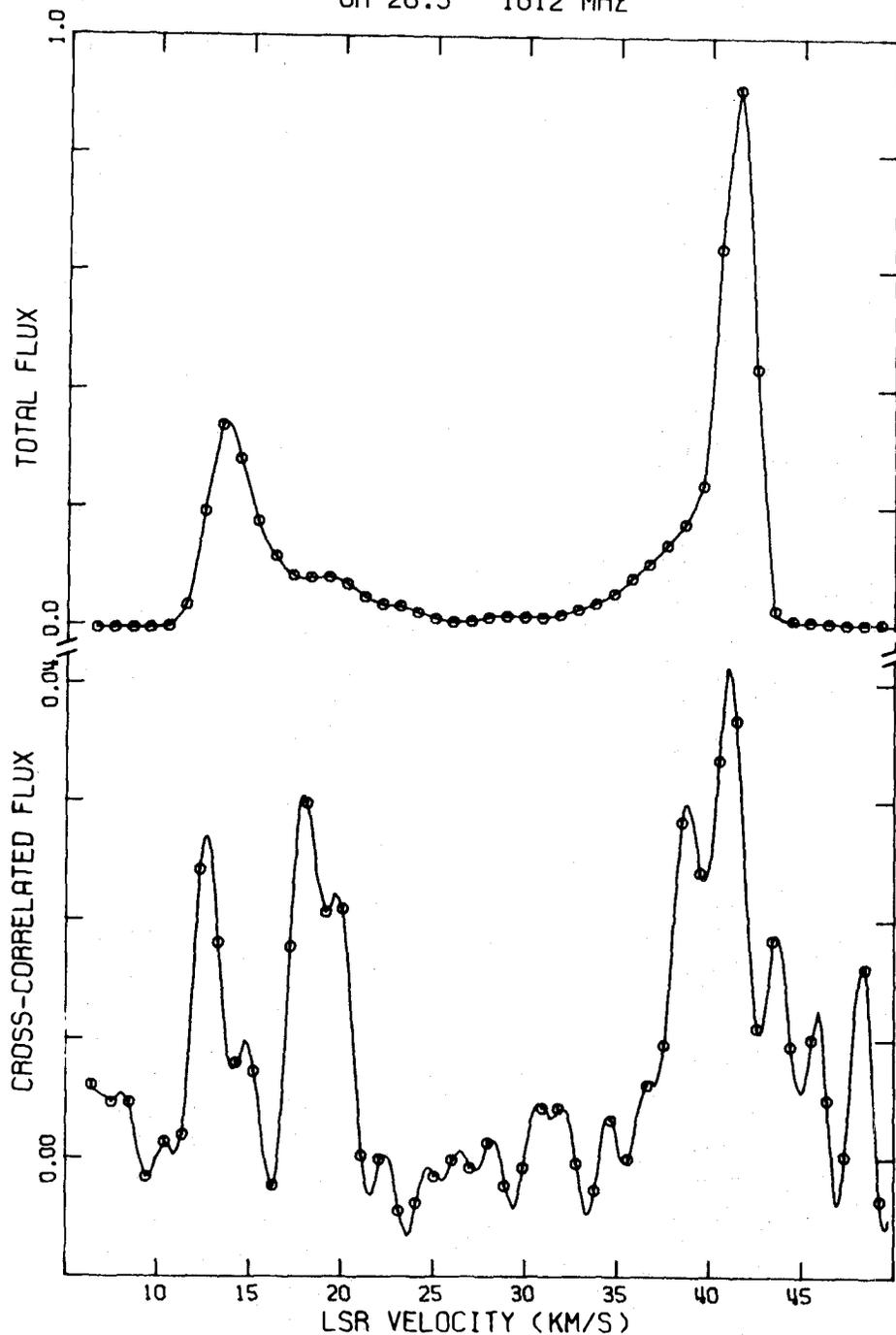
Figure 8 - The single telescope spectrum is plotted above the interferometer
spectrum for OH 26.5. The ratio of the cross-correlated flux to the total
flux at any velocity is the normalized fringe visibility. This single
telescope spectrum was obtained with the Greenbank autocorrelator, rather
than from the Mk II VLB processor, and thus the relative calibration of
the total flux and the cross-correlated flux could be in error by as much
as a factor of two. The peak total flux of 10 corresponds to about 250
Jy. Note the spectral component near 18 km/s which has a fringe visibil-
ity which is an order of magnitude greater than the other features.

OH 26.5 has one very interesting spectral feature at 19 km/s. In the cross-correlated spectrum this feature is comparatively strong, while in the total power spectrum this feature is very weak. The fringe visibility of this spectral feature is $\geqslant$ 0.5, indicating that its apparent size is $\leqslant$ 0".1. In addition, on this half-hour scan, the fringe rate of the 19 and 12 km/s features differs by more than 1 mHz which probably implies a spatial separation of these features of more than 3".0. The presence of a weak, but unresolved component, among very strong, but heavily resolved components, is not unique to this OH 26.5; it has been reported in the 1612-MHz emission of VY CMa by Moran et al. (1975). The strongest components in the total power spectrum (13 km/s and 42 km/s) have fringe rates which differ by 0.4 ± 0.1 mHz which probably imply separations of greater than 1".0.

d) R Aquilae

R Aql is one of the strongest 1612-MHz and (1667 MHz) stellar OH masers. Its 1612-MHz spectrum (see Fig. 9) is one of the most asymmetric ever observed. It is characterized by a strong, narrow, high-velocity complex at 54 km/s and a broader but very weak low velocity complex from 39-45 km/s. The circumstellar dust shell surrounding R Aql is one of the thinnest of all stellar masers as indicated by its low infrared excess. The 0.8μ-2.2μ (I-K) index of R Aql is 5.0 (Hyland et al., 1972).

The fringe visibility amplitude of the strongest component (+53.8 km/s) as a function of the projected baseline length is shown in Fig. 10. The isolated data point at 2×10^6 wavelengths is from the December, 1973, Owens Valley to Hat Creek experiment (Reid and Muhleman, 1975), while

R AQUILAE   1612 MHZ
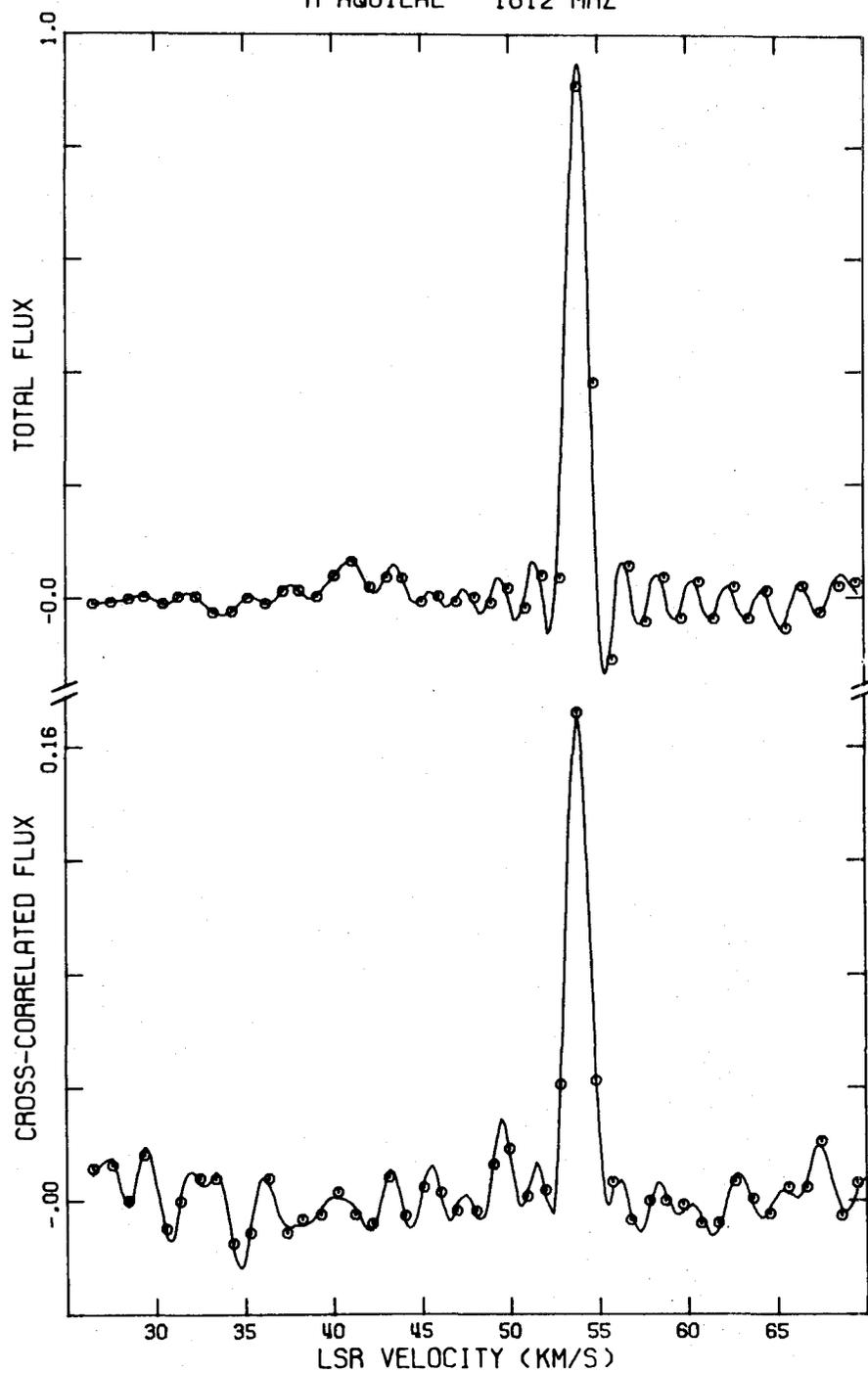
Figure 9 - The single telescope spectrum is plotted above the interfero-
meter spectrum (from one 25 minute scan) for R Aql. The ratio of the
cross-correlated flux to the total flux at any velocity is the normalized
fringe visibility. The ringing in the single telescope spectrum is a
result of the narrow spectral line and the uniformly weighted spectrum.
The peak total flux of 1.0 corresponds to about 65 Jy.

41
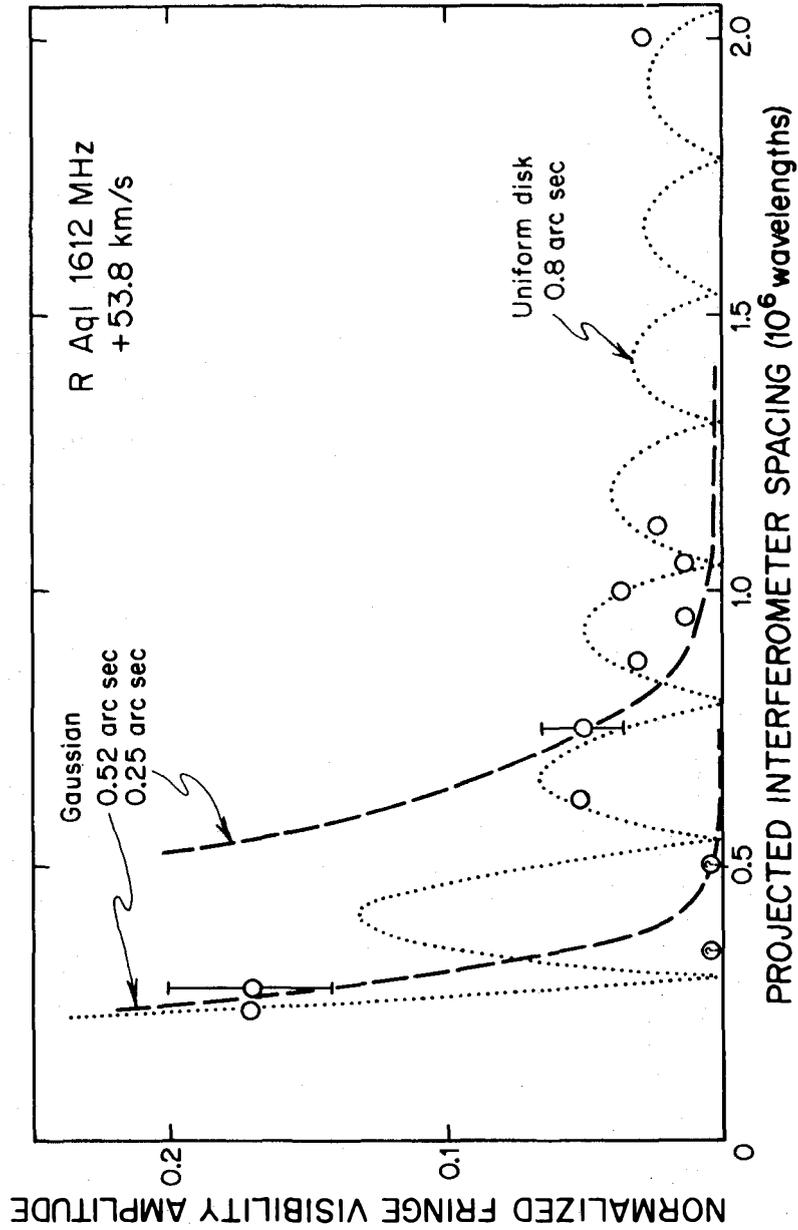
Figure 10 - The normalized fringe visibility as a function of the projected interferometer spacing. The two data points at zero fringe visibility (with question marks) indicate that no corss-correlated power was detected. It is not clear if this represents source structure or instrumental malfunction. Note that the data cannot be fit by a single gaussian component. It can be roughly fit by a uniform disk with a diameter of 0".8 or a two component "core-halo" described in the text

the rest of the data is from the Greenbank to Maryland Point experiment. The structure of this maser component is complex and cannot be explained by a single gaussian component. The fringe amplitudes of U Ori, IRC+10011, and R Aql all show a similar complex structure which can be explained with a "core-halo" model. For R Aql, this model requires a 3 Jy component with a size (FWHM) $\lesssim$ 0".05 which is seen with projected interferometer spacings less than 0".5, and a ~60 JY component with a size (FWHM) $\approx$ 0".5. Clearly more detailed observations of these sources with both longer and shorter baselines would be desirable to better test the validity of the core-halo type of structures proposed in this chapter.

e) OH 1821-12

OH 1821-12 was observed in one twenty minute scan with the Greenbank to Maryland Point baseline. Figure 11 displays the total power and the cross-correlated power spectrum at a fringe spacing of $\approx$ 0.4 arc sec. The total power spectrum was obtained from the Greenbank on-line auto-correlation reduction programs. (The Mk II video tapes have not yet been auto-correlated.) The scaling of the cross-correlated spectrum to the total power spectrum is not as precise as obtained with the normal calibration procedure, but must be accurate within a factor of two.

This source is interesting because the low velocity emission complex appears strong, while the high velocity complex is not detected, in cross-correlation. Although the absolute calibration of the cross- to auto-correlation spectra is not very good for this source, the fringe visibility of the -2 km/s component is at least a factor of 5 greater
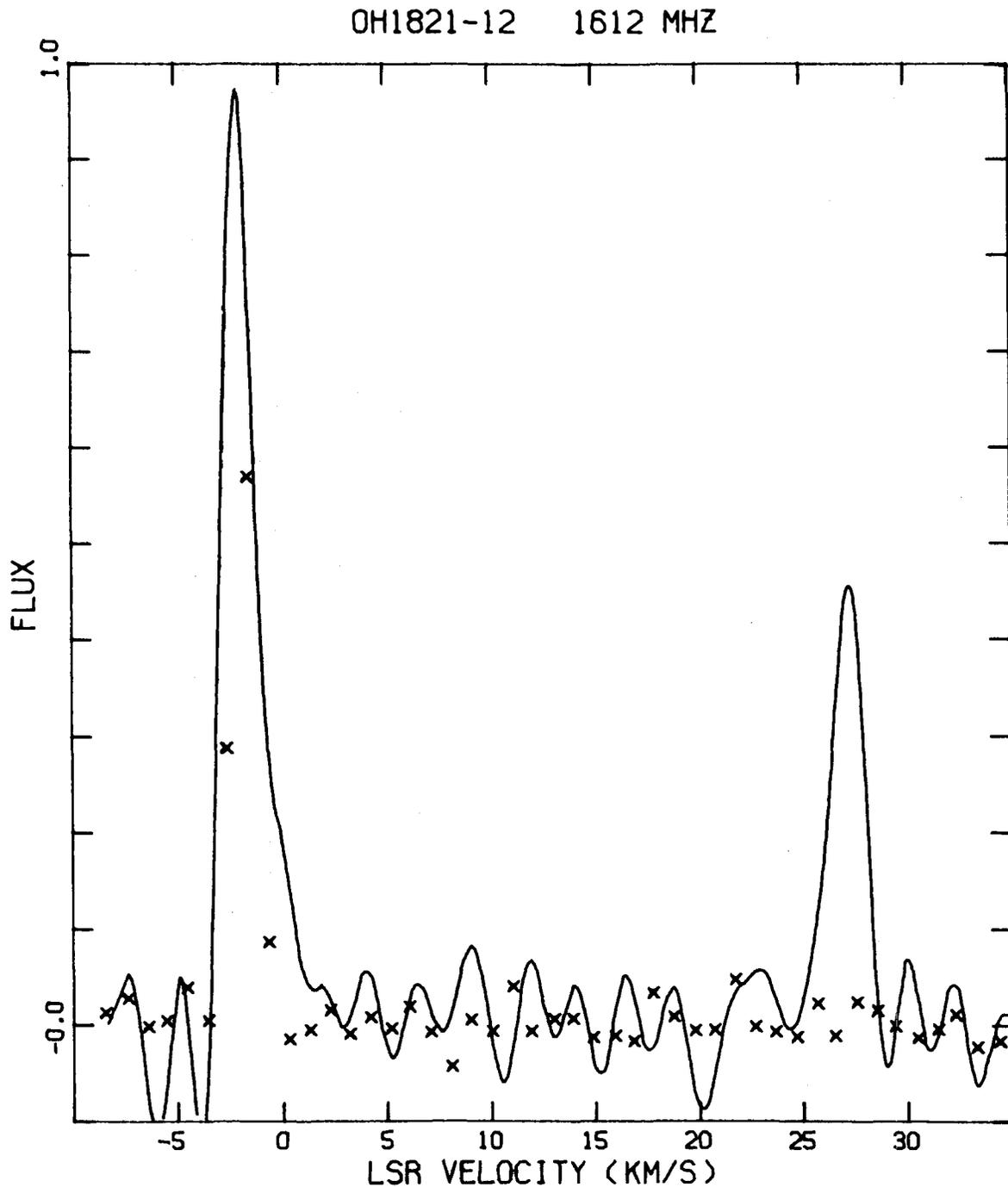
Figure 11 - The single telescope spectrum (solid line) and the inter-
ferometer spectrum (crosses) at a fringe spacing of ~0".4 for the
1612 MHz emission of OH 1821-12. The ratio of the interferometer
spectrum to the single telescope spectrum at any velocity gives
the normalized fringe visibility. The spectra are uniformly
weighted and the peak flux of 1.0 corresponds to about 25 Jy.
Note the absence of the 27 km/s component in the interferometer
spectrum.

than the visibility of the +27 km/s component. Assuming a visibility > 0.5 for the -2 km/s component, its apparent size is < 0".2 (gaussian FWHM). This implies a visibility of < 0.1 for the 27 km/s component, and thus, an apparent size > 0".3. These are very conservative size limits. The real difference in size may be much greater.

f) W 43a

W 43a has not been identified with an optical or IR star. It is located 10 arc min from the HII region W 43, but it is not clear if it is directly associated,since its velocity is 60 km/s less than the HII region (Wilson and Barrett, 1972). W 43a has a 1612-MHz spectrum which appears similar to other stellar OH masers, and thus, has been considered a type II OH/IR star.

W 43a was observed on one twenty minute scan with the Greenbank to Maryland Point baseline at a fringe spacing of $\sim$ 0".4. Figure 12 displays the total power and cross-correlated power spectrum. The calibration of the fringe visibility of this source is questionable, since the data was processed on the old 96 channel Mk II processor (see Section II of this chapter). A fringe visibility of unity was assumed for the 41 km/s component. This implied fringe visibilities very close to unity for all other spectral components which supports the adopted calibration. A fringe visibility, γ, near unity was expected for this source (at a fringe spacing of $\approx$ 0".4),   since Moran et al. (1975) found apparent sizes of the spectral features to be $\approx$ 0".05.
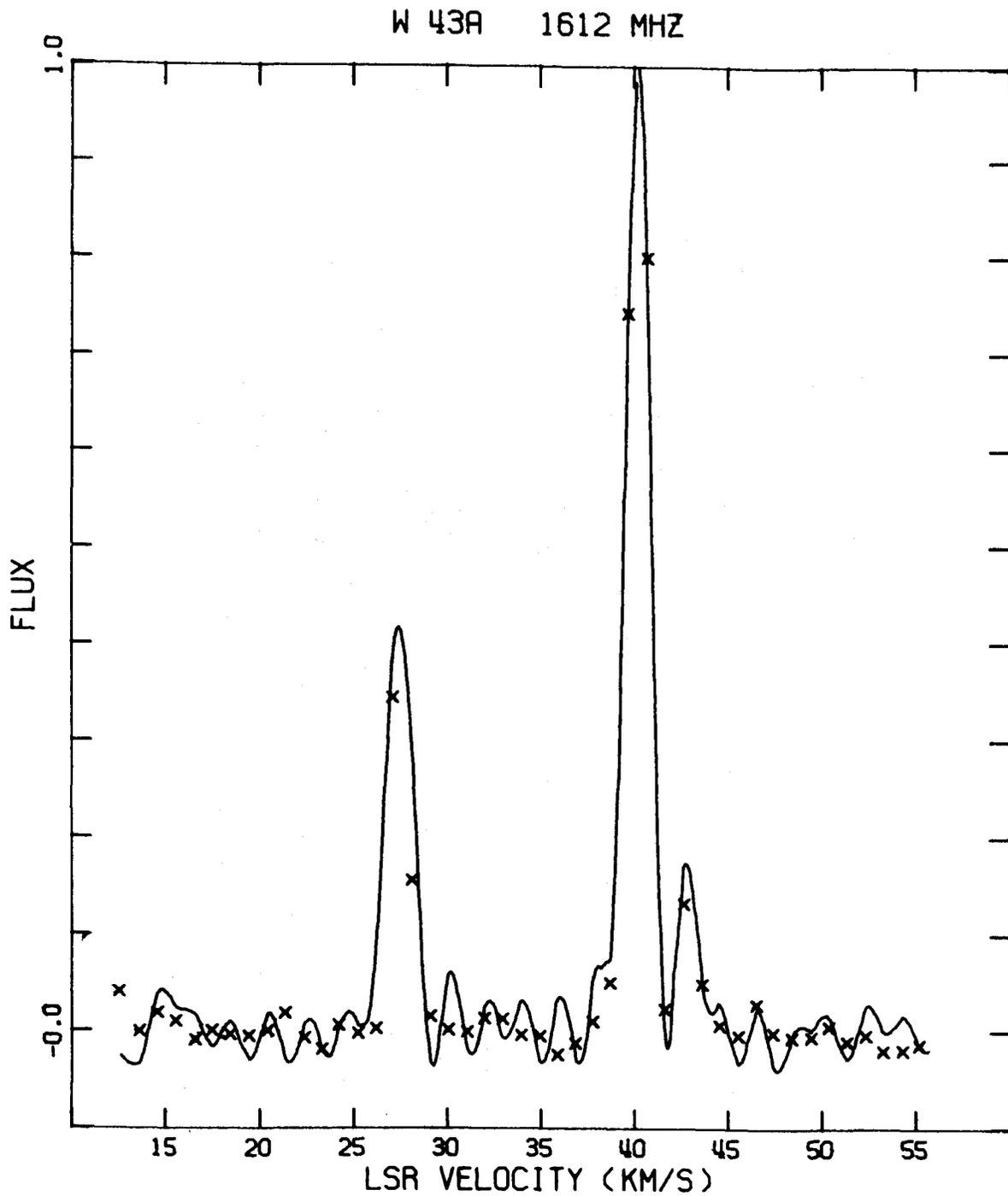
Figure 12 - The single telescope spectrum (solid line) and the inter-
ferometer spectrum (crosses) at a fringe spacing of ~0."4 for the
1612 MHz emission of W43a. The spectra are uniformly weighted
and the peak flux of 1.0 corresponds to about 45 Jy. Note that
all components in W43a appear spatially unresolved.

46

g) VX Sgr

VX Sgr is considered a supergiant stellar OH maser in some ways similar to NML Cyg and VY CMa. It has been observed in two previous experiments. Mutel (1974) detected VX Sgr with an Owens Valley to Hat Creek baseline, and Moran et al. (1975) detected and mapped VX Sgr with an Algonquin Park to Greenbank baseline. VX Sgr was observed on two twenty minute scans with the Greenbank to Maryland Point baseline at a fringe spacing of $\approx 0\rlap{.}''4$. The total power and cross-correlated power spectrum are presented in Fig. 13.

h) OH 1837-05 and IRC+50137

No fringes were detected on one scan of OH 1837-05 with the Greenbank to Maryland Point baseline or on several scans with the earlier Owens Valley to Hat Creek baseline (Reid and Muhleman, 1975). If the negative result (from the single 30 minute scan) on the Greenbank to Maryland Point baseline is correct, this places a lower limit on the apparent size (FWHM gaussian) of the strongest maser component of $\sim 0\rlap{.}''2$. No fringes stronger than 2 Jy were found on IRC+50137 with many scans on the Greenbank to Maryland Point baseline at fringe spacings of from $0\rlap{.}''2$ to $0\rlap{.}''4$. This implies a lower limit on the apparent angular size of the strongest maser components of $\sim 0\rlap{.}''3$.
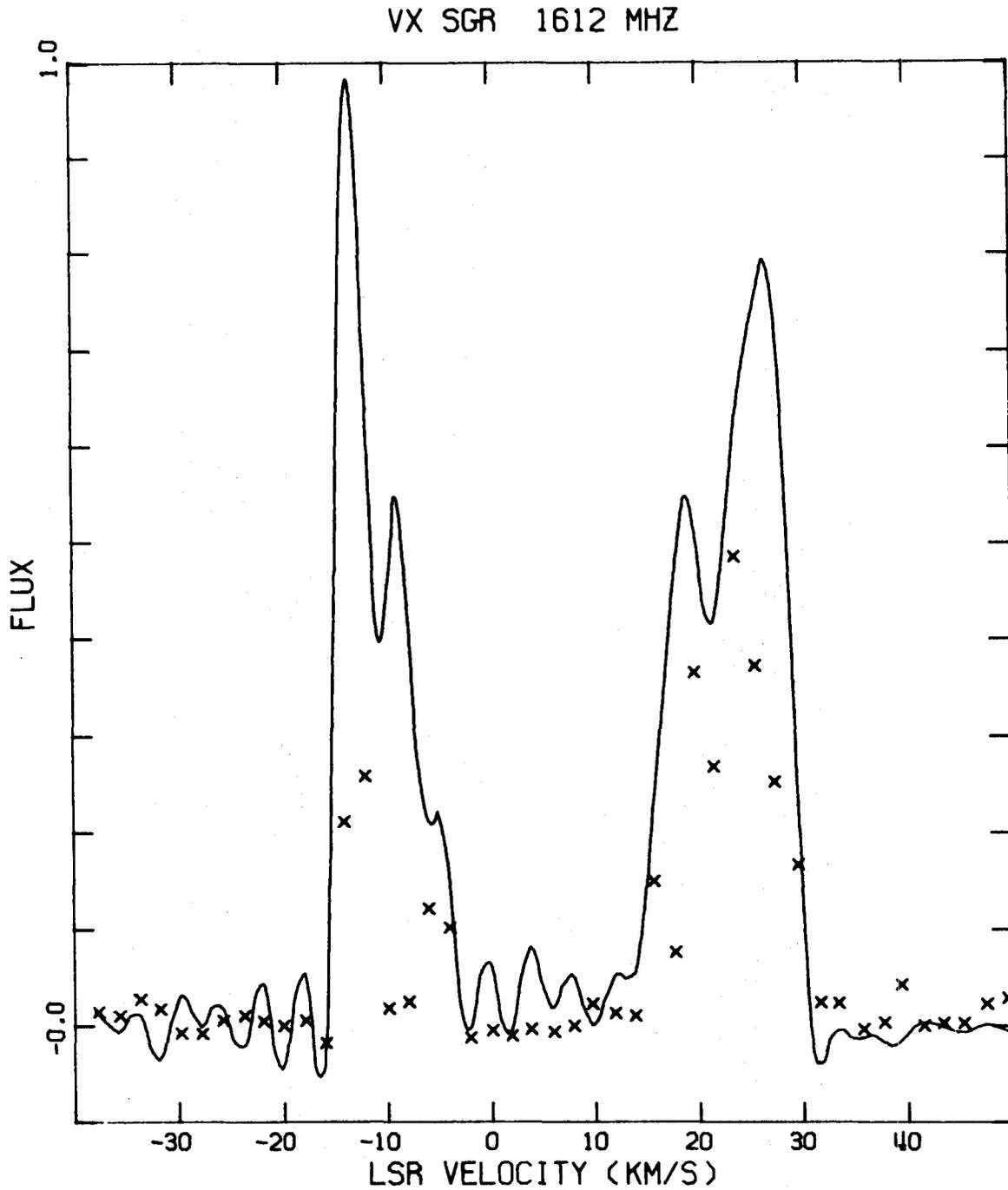
VX SGR   1612 MHZ

Figure 13 - The single telescope spectrum (solid line) and the inter-
ferometer spectrum (crosses) at a fringe spacing of ~0".4 for the
1612 MHz emission of VX Sgr. The spectra are uniformly weighted
and the peak flux of 1.0 corresponds to about 30 Jy.

IV. 1667-MHz OBSERVATIONS

a) R Aquilae, W Hydrae, IRC-10529

Three of the strongest main line emitting Mira type OH masers –

R Aql, W Hya, and IRC-10529 – were observed with the Owens Valley to Hat

Creek baseline. Very weak fringes may have been detected on W Hya with

a fringe spacing of 0".4 which is the shortest projected interferometer

spacing for this source. This should be confirmed by future observation.

No fringes greater than 2 Jy were detected on R Aql or IRC-10529 with

fringe spacings of about 0".1. Lower limits on the apparent size of the

strongest maser component in each of these sources are 0".2 for W Hya,

0".07 for R Aql and 0".05 for IRC-10529.

V. DISCUSSION

One of the most fundamental problems concerning Mira variable stars has been the determination of the dynamical structure of the gas and dust surrounding many of these objects. The recent discovery of OH radio emission from the reddest Mira and infrared variables (Wilson et al., 1970) has enabled the circumstellar material to be examined in great detail through the very intense OH maser radiation at 18 cm. The OH maser radiation from these stars exhibits a characteristic spectrum of two complexes of emission lines each several km/s wide and separated by several tens of km/s. The velocities of these OH emission complexes correlate with the optical emission and absorption line velocities. However, unlike most of the optical lines which are formed in or near the stellar photosphere, the OH radio lines are formed directly in the circumstellar material. Thus, the radio emission is probably one of the best probes of the conditions and structure of the circumstellar material.

Once the characteristic OH emission pattern was established, several models of the dynamical structure of the circumstellar material were proposed. Wilson et al. (1970) originally outlined most of the models being discussed in the present literature. These models require either expanding or contracting spherical shells, rotating disks, or shock waves propagating into the circumstellar material. The primary results of this Chapter and Chapter I lead to the conclusion that the simple expanding models are in best agreement with the available data.

a) Contracting or Rotating Circumstellar Envelopes

Direct observational evidence exists which strongly rules out either contraction or rotation as the primary motion of the circumstellar material. If circumstellar material is contracting onto the central star then the OH emission could originate from the approaching and receding sides of a spherical circumstellar envelope. If the circumstellar material possesses enough angular momentum to have formed a stable rotating disk, the OH emission could originate from the approaching and receding edges of this disk, if its rotation axis is not aligned parallel to our line sight. These models require the velocity separation of the two complexes of emission, Δv, to be supported by gravitational energy alone. In both cases, one-half of the velocity separation must be less than the escape velocity of the material at the distance, r, of the masing molecules from the central star:

$$\frac{\Delta v}{2} < \sqrt{\frac{2GM_*}{r}} \quad . \tag{1}$$

The observations reported in this paper place a lower limit on the apparent angular size, θ_{app}, of the stellar maser components. The distance of the masing molecules from the star must be greater than the apparent linear size, ℓ_{app}, of the maser components. If this were not the case, then velocity differences across the apparent amplifying region would equal half the velocity separation of the two emission complexes in the contracting or rotating models. This would produce two very broad and partially blended emission features instead of the narrow, well separated emission complexes observed. Thus,

$$\ell_{app} = \theta_{app} d < r$$

where d is the distance from the star to the sun. Substituting this relation into equation (1) and solving for the mass of the star yields

$$M_* > \frac{1}{8} \frac{\Delta v^2 \theta_{app} d}{G} \qquad (2)$$

All of the parameters on the right hand side of equation (2) are well determined by observation. For example, parameters for a typical stellar maser (e.g. IRC+100 11) are $\Delta v \sim 35$ km/s, $\theta_{app} \sim 0\overset{''}{.}5$, and d ~ 300 pc. Thus, for contraction or rotation to lead to the observed OH maser spectra, equation (2) requires the mass of the central star to exceed 25 $M_\odot$.

The lower limit to the mass of the central star obtained from equation (2) is likely to be a very conservative one for at least two reasons. First, the linear size of the masing regions was estimated, assuming that it is no larger than the apparent size. It has long been recognized that the apparent size of a maser can be less than its physical size (e.g. Goldreich and Keeley, 1972; Litvak, 1973). Although it is difficult to obtain a reliable estimate of this factor except under highly idealized assumptions of the geometry and state of saturation of the maser, it is likely that the masers are at least a factor of two larger than their apparent size. Second, equation (1) does not include a factor of two required in the rotational case by the Virial theorem, and probably a larger factor for the contracting case if outward pressure due to radiation from the central star and collisions in the circumstellar material are considered. Taking these points into account,

a more realistic lower limit to the mass of the central star required for

for a contracting or rotating circumstellar cloud is 50 to 100 $M_\odot$. It

is unlikely that the masses of the Mira stars with OH maser emission

could be in excess of a few solar masses , and

almost impossible for them to approach $\sim 50 M_\odot$. Thus, neither contrac-

tion nor rotation can be the dominant dynamical condition of the cir-

cumstellar material in the stellar maser sources.

It is important at this point to discuss the possibility of scat-

tering being responsible for the large observed apparent maser sizes.

The preceding discussion required relating the apparent source size with

the physical size of the masing regions. If scattering broadens an

intrinsically small maser feature to the $\sim 1"$ sizes observed in our

experiments, contraction or rotation of the circumstellar material could

not be ruled out. The masers could originate much closer to the central

star where gravitational forces could reasonably support the velocity

separations observed in the OH maser spectra. However, it seems impos-

sible that scattering from electrons in either the interstellar or the

interplanetary medium could enlarge the 18 cm OH maser sizes to $\sim 1".0$.

Both pulsars and quasars are observed through the interstellar and inter-

planetary medium to have apparent sizes more than two orders of magni-

tude smaller than the stellar OH masers.

While material between the earth and the masing stars cannot scat-

ter the OH emission into a 1" beam, one could hypothesize that material

directly associated with the circumstellar cloud could scatter maser

radiation passing through it. This possibility is difficult to rule out.

However, observations of maser components in the same star with widely differing apparent sizes argues against this hypothesis. The 1612-MHz emission from U Ori reported in this paper has components with sizes smaller than 0''.02, and others with sizes larger than 0''.5. Also, in the supergiant source VY CMa, some 1665 and 1667-MHz maser components are smaller than 0''.05 (Chapter III) and one 1612-MHz component is smaller than 0''.02 (Moran et al., 1975), while other 1612-MHz maser components are about 0''.1. Therefore, unless the circumstellar scattering material is exceedingly clumpy on a 1" scale it cannot be invoked to explain the observations. In light of this difficulty and the problem of creating a very strong, wide angle ($\sim 180^{\circ}$ scattering angles) scattering region (presumably by ionizing a large amount of circumstellar material) at distances $\geq 10^{15}$ cm from a cool M-type star, the apparent sizes of at least the largest maser components will be assumed to represent an intrinsic property of the maser.

b) Expanding Circumstellar Envelopes

The large size of the OH masers around Mira variables appears to make expansion of the circumstellar material the most plausible possibility. There is little problem in supplying the energy needed to drive circumstellar material outward by means of radiation pressure, since the luminosities of the central stars are very high: $L \sim 10^4 \, L_{\odot}$. Theoretical arguments (Salpeter, 1974) suggest that the stellar radiation strongly accelerates dust grains which form a few stellar radii from the star. Collisions between grains and gas transfer momentum to the gas, which causes it to flow

outward (a detailed discussion of the physical processes involved in the

circumstellar envelope is given by Goldreich and Scoville, in press).

Two classes of models of expanding circumstellar shells exist which

could lead to the velocity separation observed in the OH maser spectrum.

The one which is the simplest and in best agreement with all of the

observational data is that the OH emission originates in the approaching

and receding sides of expanding thick circumstellar "shell." This model

will be discussed later. The other model is that the OH emission we

observe originates only from the approaching side of an expanding

"shell." A stationary shock front is postulated in which material

inside (i.e. closer to the star) the shock front is expanding at the

velocity of the low velocity OH complex, and the material outside the

shock front is stationary with respect to the star which is moving at

the velocity of the high velocity OH complex. A schematic diagram of

the shock front model is given in Fig. 14. This model has several

serious drawbacks. If the maser is amplifying its own spontaneous emis-

sion, then maser emission from the far side of the star (which would

have a higher velocity than the "high velocity" OH complex) should exist.

The absence of this (third) emission complex requires a strong absorber

between the masing region and the central star. Alternatively, if the

maser is amplifying 18 cm emission originating closer to the central

star than the masing region, then the maser components we observe would

be images of the central source. This could explain the absence of

observed maser emission from the far side of the star. These conditions

could be met by postulating an ionized region which is optically thick
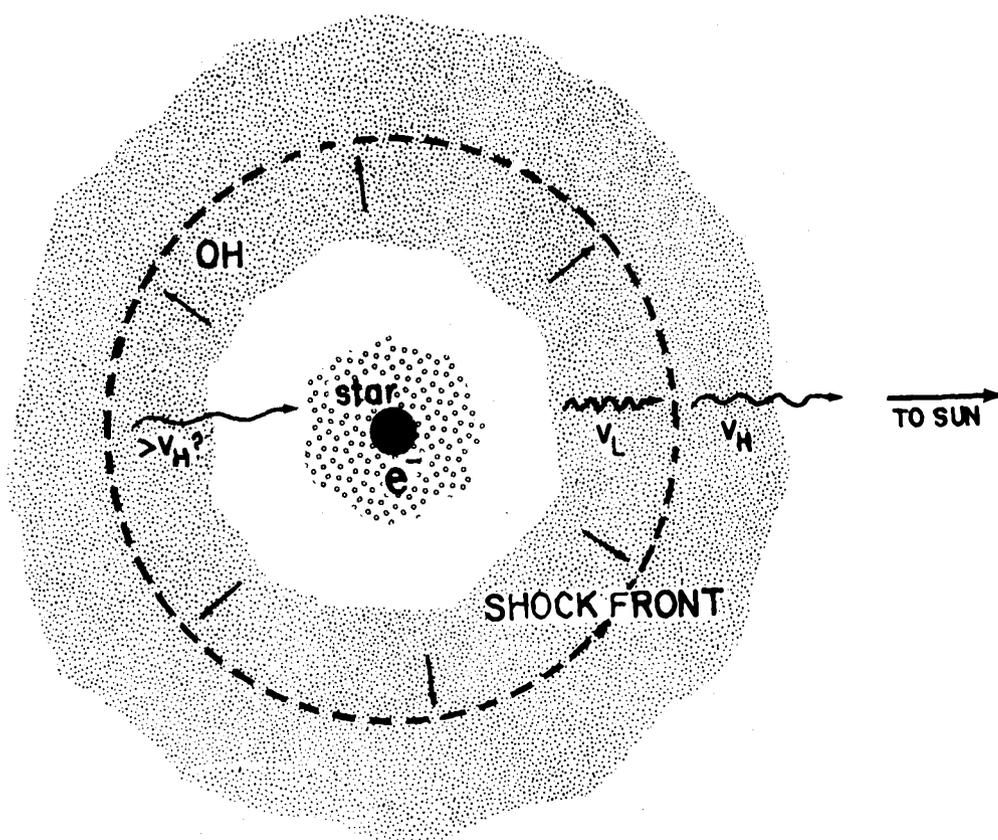
Figure 14.--A schematic diagram of the shock front model for the origin
of stellar OH maser emission. The low velocity emission, v_L,
comes from OH expanding radially outward from the star, and the
high velocity OH emission, v_H, comes from OH at rest with respect
to the star and at a greater distance. The two regions of OH
emission are separated by a shock front (dashed circle). No
emission is observed from the far side of the star. An ionized
region (denoted by e^-) is postulated to absorb far side emission.

at 18 cm. Whether this ionized region is an absorber or a source, it must have a diameter at least as large as the apparent extent of the low velocity OH emission complex to explain the OH spectra observed. Another important constraint on the ionized region is that it must not extend far enough outward from the star to absorb either of the observed OH emission complexes. This excludes the shock front as the ionization source.

Several observations have been reported in the literature which have a bearing on the existence of an ionized region surrounding the central star. Observations by Wilson (1971) place an upper limit of 0.27 Jy for the 3.5 mm flux density of NML Cyg, a very strong 1612-MHz OH emitting "supergiant." Observations by Harris (1974) with the Cambridge 5 km telescope place an upper limit on the 6 cm flux density of NML Cyg at 0.010 Jy. Recently, Goss et al. (1974) reported a detection of NML Cyg at 2.8 cm of 0.012 ± 0.002 Jy. Goss et al. argue that these observations can be explained by a small HII region of $0\overset{..}{.}07$ (5×10^{14} cm at 500 pc) with an electron density of 10^6 e/cm^3 and an electron temperature, T_e, of 10^4°K. While such an ionized region can (and perhaps does) explain the 2.8 cm emission, it cannot be invoked to totally absorb 18 cm OH maser emission from the back side of the circumstellar cloud surrounding NML Cyg. In order to absorb back side emission, the ionized region must be assumed to be at least as large as the total angular extent of the front side emission which Davies et al. (1972) report as greater than 1 arc second (7.5×10^{15} cm at 500 pc). Using this size for the ionized region, the observations reported by

Goss et al. (1974) require that

$$T_b \Omega_s \; = \; 2.8 \times 10^{-9} \; \text{°K st.rad} \quad ,$$

where T_b is the brightness temperature and Ω_s is the solid angle of the HII region. This in turn implies $T_b \sim 40°\text{K}$. If the cloud is optically thick at 2.8 cm, this implies an electron temperature of close to the brightness temperature of $40°\text{K}$ which is improbable in a region containing dust at about $600°\text{K}$ (Hyland et al., 1972).

It is nearly impossible to construct a set of parameters for an ionized region which can satisfy all of the observational constraints and still absorb the OH emission from the back side of the star. The best one can do is make the optical depth as low as possible at 2.8 cm, in order to obtain a more reasonable electron temperature, but not so low that the HII region becomes optically thin at 18 cm and cannot absorb the back side maser radiation. If $\tau_{2.8 \text{ cm}} = 0.1$, then $\tau_{6.0 \text{ cm}} = 0.5$ and $\tau_{18 \text{ cm}} = 4$. In this case, the 2.8 and 6.0 cm emissions are from optically thin regions where the spectral index of an HII region is nearly zero. Thus, the expected 6.0 cm flux density could only be slightly lower than the 2.8 cm flux density of 0.012 Jy; it is possible, then, that it could have barely escaped detection at the (6.0 cm) 0.010 Jy limit of Harris (1974). An optical depth of four at 18 cm is just barely large enough to absorb most of the 200 Jy which would come from the back side if it is assumed to be similar to the front side (low velocity emission complex) and escape detection at about a 3 Jy limit (Wilson et al., 1970). With a 2.8 cm optical depth of 0.1, the brightness temperature of $40°\text{K}$ calculated above implies an electron

58

temperature of 400°K which is closer to, but still below, the dust

temperature. These parameters define an emission measure of 2.5×10^8

cm^{-6} pc, a rms electron density of 3×10^5 cm^{-3}, and a total mass of

ionized hydrogen of $\sim 10^{-4}$ $M_\odot$. This requires a significant fraction

($\sim 10\%$) of the hydrogen be ionized assuming the model calculations of

Goldreich and Scoville (in press) of molecular abundances.

While it is highly improbable, given the NML Cyg observational

limits, to have a large obscuring HII region close to the star as

required by the shock front model, there is another argument against

this model. The shock front model requires the high velocity OH emis-

sion complex to have the same radial velocity as the star and the low

velocity OH emission complex to be systematically approaching us.

Historically, the only reason that such a model was postulated was to

make the OH velocity structure agree with the belief that the optical

absorption line velocities (which lie close to the high velocity OH

complex) represented the stellar radial velocity. The analysis pre-

sented in chapter I casts strong doubt on this long-standing view, at

least for stars with molecular masers. It appears that the high velo-

city OH complexes and the optical absorption lines are redshifted with

respect to the stellar radial velocity. The existence of OH maser lines

which are redshifted with respect to the central star in an expanding

system requires emission from the back side. This cannot be accounted

for by the shock wave models.

The only simple model which explains the large sizes of the stellar

OH masers and the existence of OH emission, which is redshifted with

respect to the stellar radial velocity, is the expanding thick shell

model. The OH emission is observed from the front (approaching) and back

(receding) of the expanding "shell" and corresponds to the low and high

velocity emission complexes. Since the paths of maximum maser gain are

those in which the line-of-sight velocity to the observer are nearly

constant, the emission ideally should be confined to a small "cap"

directly in front of and in back of the star. Maser emission from the

limbs is not expected because the line-of-sight velocity gradient is

steepest there. This model requires the stellar radial velocity to lie

halfway between the velocities of the two maser complexes. This pre-

dicted stellar velocity is about one standard deviation away from the

nominal value determined by statistical methods (Chapter I).

From this qualitative model, one would expect that the low and high

velocity OH emission complexes would, in general, overlap on the sky.

The present observations are not adequate to critically evaluate this

prediction, since most of the objects were overresolved with the

baselines used. However, in the cases of IRC+10011 and U Ori the

qualitative predictions of the simple expanding model are not violated.

The features detected in each emission complex in IRC+10011 appear to

be spatially separated by a distance which is less than the apparent

sizes of the dominant (overresolved) components, and it is possible that

no measurable offset exists. The various maser components mapped in the

low velocity complex of U Ori (the high velocity complex has not been

observed to be stronger than ~ 1 Jy at 1612-MHz) are all confined within

0."2 and thus, are extended over a region which is smaller than the apparent sizes of the heavily resolved components ($\gtrsim$ 0."5).

c) Structure of the Circumstellar Material

The optical, infrared, and microwave observational data available on stellar maser sources together define a fairly complete, semiquantitative description of the entire object. However, many details remain uncertain; for example, at this point we cannot distinguish between an expanding spherical or disk-like structure of the circumstellar material. More detailed and quantitative theoretical descriptions of the physical properties of the circumstellar material are given by Litvak (1973) and Goldreich and Scoville (in press).

The central star is a late M-type Mira variable with photospheric temperatures which average about 2000°K and have peak-to-peak fluctuations of about 500°K (Merrill, 1940). The central star is quite large; Nather and Wild (1973) have measured the size of R Leo to be $\approx 2.5 \times 10^{14}$ cm at visible to near-IR wavelengths by a lunar occultation. The optical emission lines of hydrogen and absorption lines of excited metallic oxides are formed near the turbulent photosphere, and the neutral metallic absorption lines are formed further from the star in the cooler circumstellar material. An extensive layer of dust, probably containing silicates, envelopes the star and in many cases totally obscures the optical emission from the stellar photosphere. The dust distribution has been determined for IRC+10011 by a lunar occultation (Zappala et al., 1974) to be densest near 5×10^{14} cm and to extend outward to at least

5×10^{15} cm. The temperature in the dust appears to range from about 600°K to about 200°K at these two distances.

The radial velocity structure of the circumstellar material is probably dominated by radiation pressure (Salpeter, 1974). Since radiation pressure and gravitational force decrease as the inverse square of the distance from the central star, the expansion velocity should monotonically increase with distance and rapidly approach a terminal velocity (cf. Kowk, 1975). Since the 1612-MHz OH emission occurs at such large distances from the star, the terminal expansion velocity should be given by half the velocity separation of the outer edges of the low and high velocity 1612-MHz emission complexes. In general, the 1665 and 1667-MHz complexes have a slightly smaller velocity separation than the 1612-MHz complexes. Thus, theoretically, the main line emission should occur closer to the central star than the 1612-MHz emission. This seems to be observationally confirmed for the supergiant source VY CMa, in which the main line emission complexes have smaller apparent sizes (Chapter III) and are confined to a smaller region than the 1612-MHz complexes (Masheder et al., 1974; Harvey et al., 1974; and Chapter III). The three strongest main line emitting Mira variables, W Hya, R Aql, and IRC-10529, exhibited sizes greater than ~ 0''06. This implies, at least for R Aql where the stronger 1612-MHz emission was detected, that while the main line emission sizes may still be smaller than the 1612-MHz sizes, the difference is probably not very great. If this tentative

conclusion is substantiated by further observations, it means that the main line emission occurs under similar conditions as the 1612-MHz emission.

While the simple expanding shell model seems qualitatively correct, it cannot account for the detailed structure often observed in each OH emission complex. In many masers, each emission complex is composed of many maser components separated by $\lesssim$ 5 km/s. The simple shell model requires maser components at the outer edge of each emission complex to originate from points directly in front or back of the shell. Maser components near the inner regions of each emission complex should originate away from our line-of-sight to the star. The line-of-sight velocity, v_{los}, of material observed across the shell surface is given by

$$v_{los}(\theta) = V\cos(\theta) + v_s \qquad (3)$$

where V is the expansion velocity, v_s is the stellar radial velocity, and θ is the angle at the star between the sun and a point on the shell surface (see Fig. 15a). The path length through the shell over which maser amplification seems favorable (i.e. over which v_{los} does not change by more than the width of the maser line) can be estimated assuming a constant expansion velocity at large distances from the star and maser path lengths less than the "shell" thickness. If a maser component originates at a distance, a, from our line-of-sight to the star and a distance, r, (see Fig. 15b) from the center of the star, then its velocity is given by equation (3) with $\theta = \sin^{-1}(a/r)$. The path length
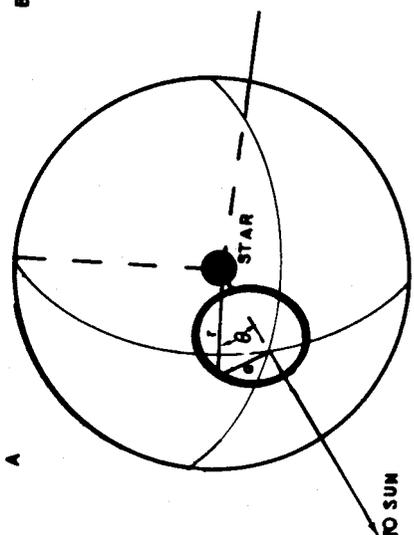
Figure 15a - A schematic diagram of an expanding circumstellar shell.
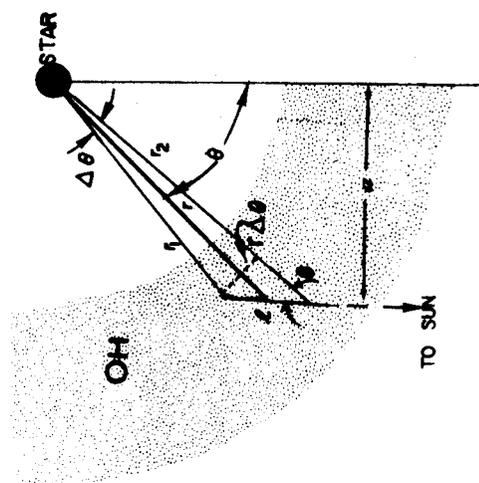
Figure 15b - The path length of maser amplification, ℓ, as a function of distance from the central star r, and position on the shell θ or a.
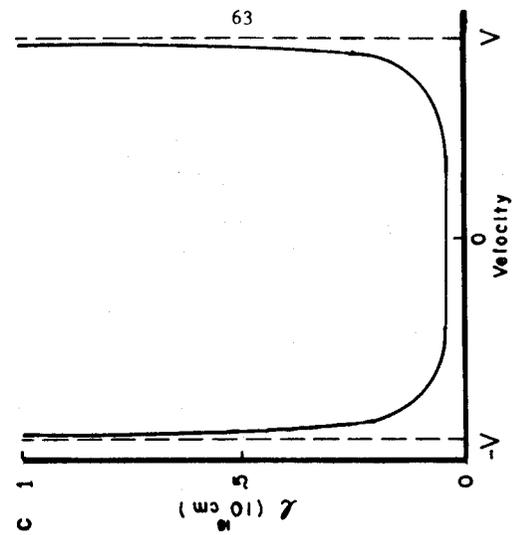
15c - ℓ as a function of the line-of-sight velocity as observed at the sun for representation parameters of stellar masers. The shell is assumed to be uniformly expanding with velocity, V, from the central star.

of maser amplification, ℓ, is given by

$$\ell(r,\theta) \approx r\Delta\theta/\sin(\theta) \qquad (4)$$

where $\Delta\theta$ is the angular extent of the maser amplification path length

seen by the star and is determined by requiring that

$$|\Delta v| = |v(\theta + \Delta\theta) - v(\theta)| \approx V\Delta\theta\sin\theta \qquad (5)$$

be less than a typical Doppler width of the maser line. Combining

equation (3), (4), and (5), the maser path length as a function of velo-

city is determined

$$\ell(r,v) \approx r(\Delta v/V) \frac{1}{1 - (v/V)^2} \qquad (6)$$

As an example, equation (6) is plotted in Fig. 15c for values of

$\Delta v = 0.5$ km/s, $r = 10^{16}$ cm and $V = 15$ km/s. The rapid decrease in the

maser path length as the line-of-sight velocity departs from the expan-

sion velocity explains the existence of the two emission complexes, but

it predicts velocity widths of the complexes which are narrower than

observed in many sources.

The simple expanding shell model predicts that maser components

with velocities relative to the central star, which are geometrically

less than the expansion velocity, should be located on a ring (see

Fig. 15a) centered on the star's position and with a radius, a, given

by

$$a(v) = r \sin\theta = r\sqrt{1 - (v/V)^2} \qquad . \qquad (7)$$

Very few high quality maps of the OH emission exist which can be used to test this prediction. The low velocity emission complex of NML Cyg mapped by Masheder et al. (1974) has components whose offsets from the center of the shell, presumably given by the -26 km/s components, roughly follow the simple shell model (i.e. eq. (7)). However, the spatial distribution of the components in the high velocity complex do not. It is possible that the severe blending of the maser components, evident in the high velocity emission complex of NML Cyg, could lead to systematic errors in the spatial mapping; this problem can only be attacked with multi-baseline observations (see discussion in Chapter III). The map of the low velocity complex of U Ori (Fig. 6) does show the higher velocity maser components falling further from the presumed center of the shell, represented by the -46.2 km/s position, but is not in good quantitative agreement with the simple shell model. Accurate maps of VY CMa (Chapter III) at 1612 and 1665-MHz show no strong agreement with the simple shell model. Therefore, the observational indications are that there is less coherent and more random structure than predicted by the simple shell model.

There are two reasons why the structure of each OH emission complex could depart from that predicted by the simple shell model. First, there is little reason to believe that the circumstellar shell should be homogeneous. The shell may fragment into many small clumps as it expands due to irregularities present during its formation and initial acceleration near the stellar surface. This may explain the apparent preference for

the individual maser components to appear as localized emission spots, rather than rings predicted by the simple shell model.

The second possible reason for departures from the simple shell model is that it seems more likely that the many concentric circumstellar shells should be formed around the central star. The formation and acceleration of the dust grains a few stellar radii from the central star are probably critically tied to the temperature cycle of these highly variable stars (cf. Salpeter, 1974). Temperatures near the photosphere of the star typically change by 25% over one period (Merrill, 1940). It is possible that during a particularly bright cycle, or every cycle, a dust shell may condense and be accelerated away from the star. Since the temperature and hence the radiation pressure of the star at different maxima can vary considerably, each shell probably receives a different initial acceleration, and therefore, a different terminal velocity. Thus, the collection of maser components in the OH emission complexes could originate in several circumstellar shells with slightly different expansion velocities. These shells may not be homogeneous and emissions within any one may be somewhat offset from our line-of-sight to the star. In this case, maps of the spatial structure of the maser components in each OH emission complex would still be confined to a region much less than the diameter of the shell, but would be somewhat randomly distributed within this region.

The 1612-MHz OH emission from Mira variables is characterized by (at least) three levels of structure. The grossest structure is the existence of two well defined narrow emission complexes separated by

tens of km/s. This can be explained by a simple expanding shell model. The next level of structure lies in the organization of many masing components in each emission complex. This structure seems to indicate the existence of more than one inhomogeneous circumstellar shell. Finally, the observations reported in this chapter imply that individual maser components have complex structures. This structure is interesting because it probably represents an intrinsic property of the maser process.

The fringe visibility amplitudes of maser components in IRC+10011, R Aql, and U Ori are characterized by low values which are nearly constant over fringe spacings of $\leqslant$ 0".08 to $\sim$ 0".4. For R Aql, the fringe amplitude is observed to turn up (see Fig. 10) only at fringe spacings $\geqslant$ 0".5. This type of a fringe visibility function can be explained in several ways. Each maser component may be composed of two components: one very small $\leqslant$ 0".05 and weak (excluding U Ori)$\sim$ 2 Jy component, and one very large $\geqslant$ 0".5 strong component which comprises most of the flux observed in total power spectra. Alternatively, a single uniformly bright "disk" $\sim$ 0".8 could mimic the present data. In order to distinguish between these possibilities further high sensitivity observations with 0".5 to 1" fringe spacings are crucial.

REFERENCES

Cimerman, M. 1975, personal communication.

Davies, R.D., Masheder, M.R.W., and Booth, R.S. 1972, Nature (Phys. Sc.), 237, 21.

Evans, N. 1975, personal communication.

Goldreich, P., and Keeley, D.A. 1972, Ap.J., 174, 517.

Goldreich, P., and Scoville, N. in press.

Goss, W.M., Winnberg, A., and Habing, H.J. 1974, Astron. & Astrophys., 30, 349.

Harris, S. 1974, M.N.R.A.S., 166, 29P.

Harvey, P.J., Booth R.S., Davies, R.D., Whittet, D.C.B., and McLaughlin, W. 1974, M.N.R.A.S., 169, 545.

Hyland, A.R., Becklin, E.E., Frogel, J.A. and Neugebauer, G. 1972, Astron. & Astrophys., 16, 204.

Kaifu, N., Buhl, D., and Snyder, L.E. 1975, Ap.J., 195, 359.

Kwok, S. 1975, Ap.J., 198, 583.

Litvak, M.M. 1973, Ap.J., 182, 711.

Masheder, M.R.W., Booth, R.S., and Davies, R.D. 1974, M.N.R.A.S., 166, 561.

Merrill, P.W. 1940, Spectra of Long Period Variables (Chicago: University of Chicago Press).

Moran, J.M., Ball, J.A., Yen, J.L., Johnston, K.J., Schwartz, P.R., and Knowles, S.H. 1975, (in preparation).

Mutel, R. 1974, personal communication.

Nather, R.E. and Wild, P.A.T. 1973, Ast. J., 78, 628.

Pataki, L., and Kolena, J. 1974, IAU Cir., No. 2680.

Reid, M.J., and Muhleman, D.O. 1975, Ap.J. (Letters), 196, L35.

Salpeter, E.E. 1974, Ap.J., 193, 579.

Wilson, R.E. and Merrill, P.W. 1942, Ap.J., 95, 248.

Wilson, W.J., Barrett, A.H., and Moran, J.M. 1970, Ap.J., 160, 545.

Wilson, W.J. 1971, Ap.J. (Letters), 166, L13.

Wilson, W.J., Schwartz, P.R., Neugebauer, G., Harvey, P.M., and
    Becklin, E.E. 1972, Ap.J., 177, 523.

Zappala, R.R., Becklin, E.E., Matthews, K. and Neugebauer, G. 1974,
    Ap.J., 192, 109.

Chapter III

VY CMa

I. INTRODUCTION

VY CMa is a strong stellar OH maser source. Its peculiar optical properties were noted by Herbig (1970), and its OH maser emission was first detected by Eliasson and Bartlett (1969). Masheder et al. (1974) attempted to map the OH emission from observations with a 20 km microwave linked interferometer. These workers are cautious about the reliability of their map because of the limited U-V coverage available to them and the severe blending or overlapping of individual spectral components in the source. Recently, VY CMa was observed on two spectral line VLB experiments: a 550 km baseline from Owens Valley to Hat Creek and a 225 km baseline from Greenbank to Maryland Point. This paper presents the results of a direct Fourier inversion analysis of the 1612-MHz emission from VY CMa. This is the first attempt at such an analysis procedure for VLB data, and it has proven to be of great value in determining the structure of this complex source. In addition, a fringe phase map of the 1665-MHz emission of VY CMa is presented.

II.1612-MHZ

a) Analysis Procedures

Most attempts to map spectral line sources from interferometric observations have relied upon modeling the differences in fringe phase between one reference component and the remaining spectral components (i.e., relative phases). In some cases this procedure has been highly successful (e. g., Moran et al., 1968; Johnston et al., 1971; Knowles et al., 1974; and Chapter II of this work). However, there are two major problems in modeling relative fringe phase data which have limited the analyses of many sources. First, the fringe phase, ϕ, is usually determined by

$$\phi = \arctan\ [Im(\gamma)/Re(\gamma)] \tag{1}$$

where γ is the complex fringe visibility. This causes fringe phases to have integral turn (1 turn = 360 degrees of phase) ambiguities which are often difficult to resolve. In order to avoid phase ambiguities, many investigators have chosen to analyze the fringe rate, f (i.e., the time derivative of ϕ), rather than the phase itself, at the cost of at least an order of magnitude loss of sensitivity to relative positions of spectral components. The second problem in modeling fringe phase data is that, in many sources, individual spectral components are blended (i.e., overlap in frequency). This may be a result of inadequate instrumental spectral resolution or it may be an intrinsic property of the source. When several spectral components are blended, the fringe power in a spectral channel can have very complex time variations. For example, in stellar maser sources, such as VY CMa, many components are widely spaced on the sky and are severely spectrally blended.

In spectral channels where this occurs, the fringe amplitudes exhibit

deep nulls, and, correspondingly, the phase exhibits large jumps.

The problem of mapping sources with blended spectral components

is very difficult. In a few cases, the data in a spectral channel

containing blended components can be modeled by a single component

with a position offset with the hope of determining the position of the

centroid of the fringe power. However, if large phase jumps exist in

the data, the centroid position estimate can easily be in error by an

interferometer fringe spacing. Further, in cases where the phase data,

rather than the fringe rate data, are used, the phase jumps due to

blending of components can seriously hinder the resolution of fringe

phase ambiguities.

There are two methods of analyzing interferometer data which can

handle problems of spectral blending. First, one can attempt to model

the fringe power in a spectral channel as the sum of several components

and simultaneously fit a set of parameters representing each component's

strength, shape and position to both the fringe amplitudes and phases.

This process requires resolution (i.e., phase connection) of the phase

turn ambiguities. Two component models have been fit to spectral inter-

ferometer data by Harvey et al. (1975). Multiple component modeling

generally is a difficult task. It is a non-linear process and requires

significant a priori knowledge of the source structure to assure a

reasonable expectation of success.

An alternative procedure for analyzing interferometer data is to

Fourier invert the complex fringe visibility, and then to deconvolve

the response of the transform to a point source from the inverted map.

The major advantages of this procedure are that it is linear (and thus requires no a priori knowledge of the source structure), and that it does not require resolution of the fringe phase ambiguities. However, without a large array of well spaced baselines the deconvolution or "cleaning" (Högbom, 1974) process can create spurious features in the maps.

VY CMa was observed in two experiments approximately one month apart. Since VY CMa is only slowly variable on a scale of $\sim$ 1000 days (Herbig, 1970), the data obtained from the two experiments can be combined confidently. The U-V coverage obtained by observing the source when it was more than 10° above the horizon is shown in Fig. 1. A zero spacing point is included since the total power spectrum of the source is known from auto-correlating the Mk II VLB tapes for each telescope. Although the absolute zero spacing flux is poorly known, it is known with great precision relative to the interferometer fluxes. The scaling of the interferometer flux to the total flux of the source can be done with a precision of better than 5% for strong signal to noise sources.

The response of the Fourier inversion of data representing a point source is called the "dirty beam" and is shown in Fig. 2. The major structure of the dirty beam can be represented by a main lobe with a full width at half maximum of roughly 0".2 by 0".1 extended along the north-east, south-west direction. Dominant sidelobes peaking to -50% and +50% of the main lobe strength are approximately 0".25 and 0".45 from the center of the main lobe, respectively. Secondary side lobes of up to ± 20% extend over a second of arc in the north-south direction.
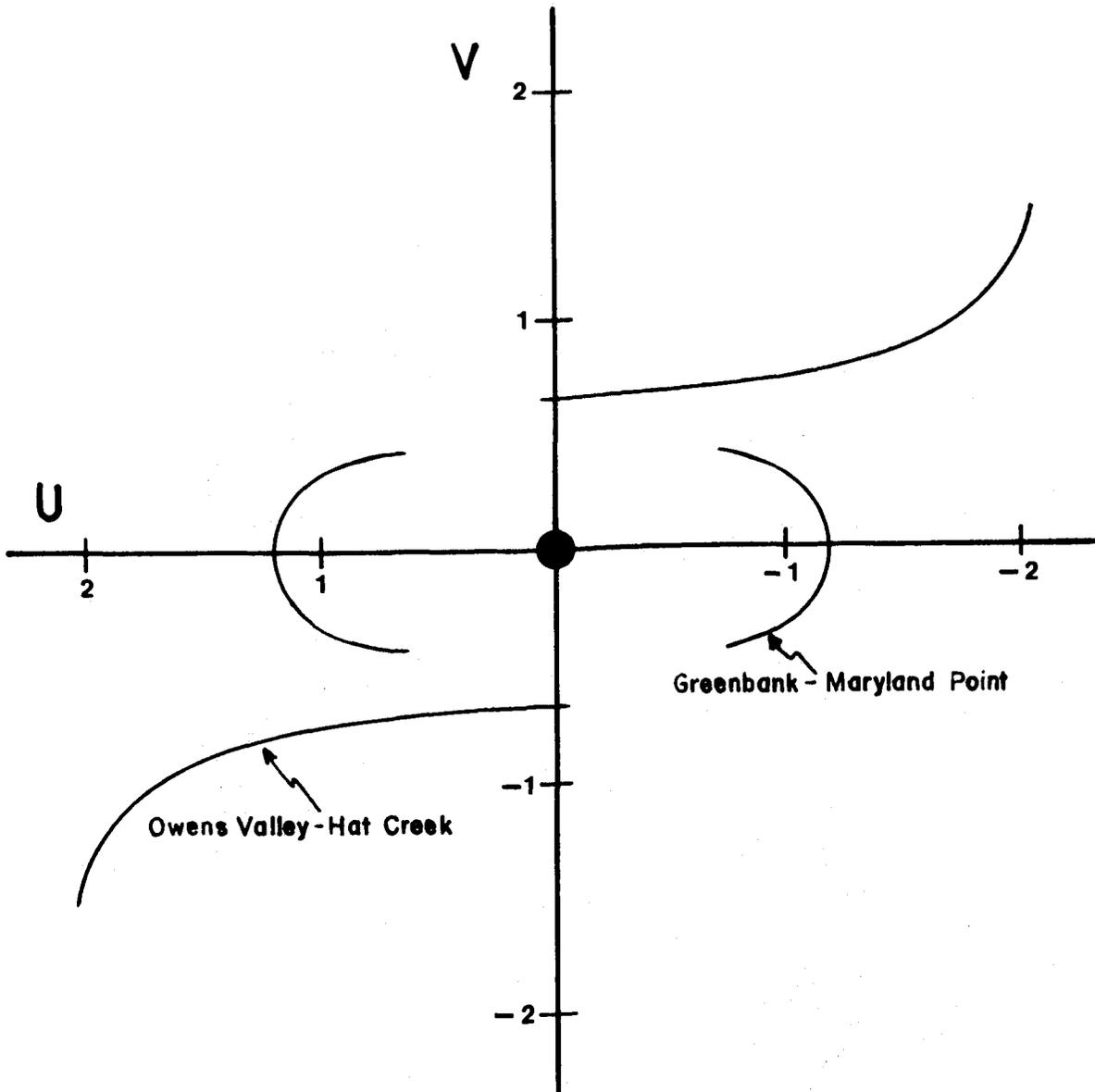
Figure 1 - Projected interferometer spacings in the east (U) and north (V) directions for the two baseline observations of VY CMa at 1612 MHz. U and V are in units of 10^6 wavelengths. A second U-V track is indicated for each baseline since the complex fringe visibility, γ, satisfies the relation $\gamma(U,V) = \gamma*(-U,-V)$. A zero spacing point is also indicated.
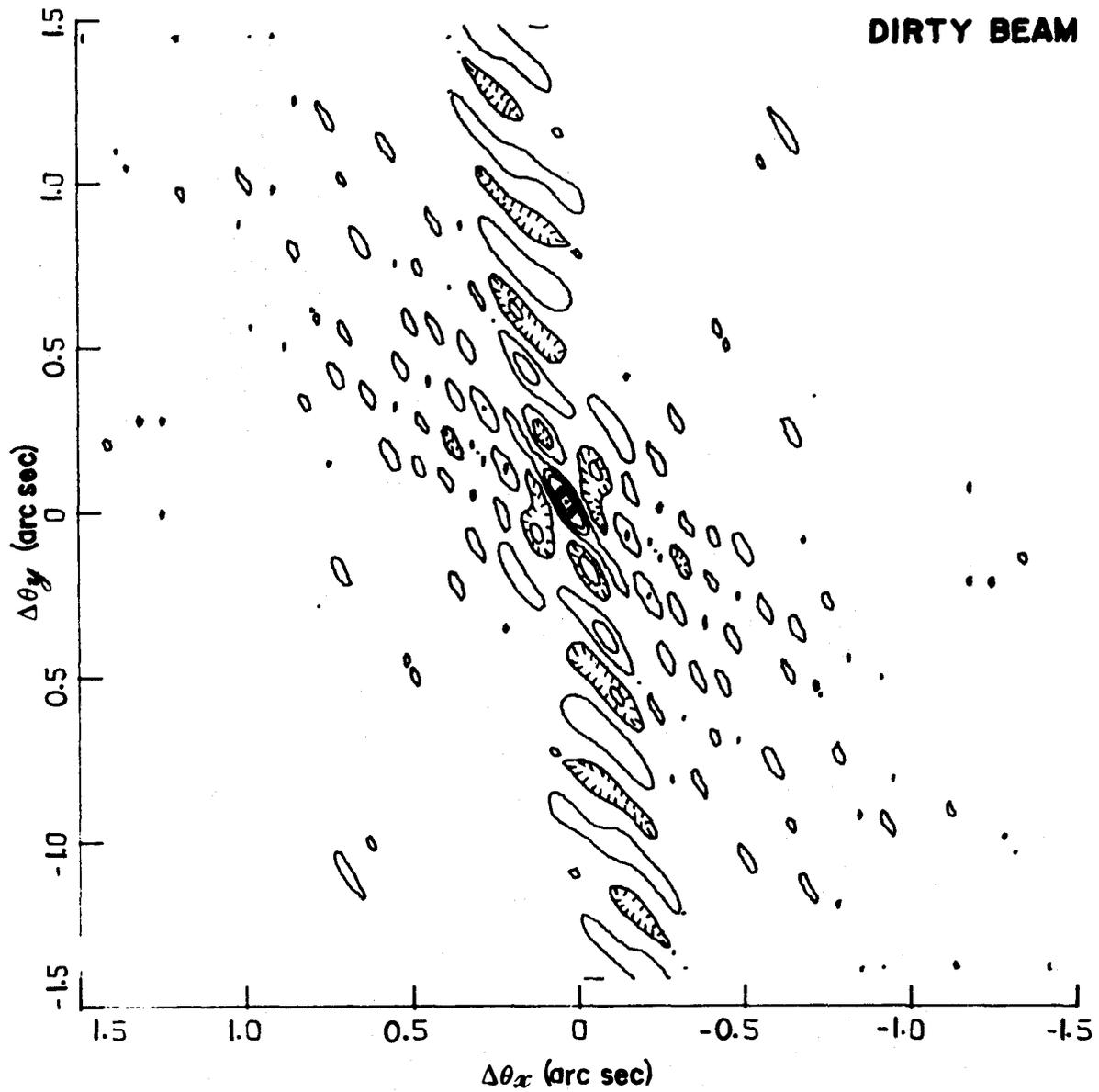
Figure 2 - The response of the Fourier inversion of a point source, the dirty beam, with the U-V coverage of Fig. 1. Contour levels are at intervals of 20% of the peak response and the zero level contour has been suppressed.

The large side lobes of the dirty beam preclude a very accurate cleaned map. For example, a source composed of very many closely spaced components is difficult to analyze with such a beam. In order to test the limitations of the Fourier inversion and cleaning process with the U-V coverage obtained for VY CMa, many source structures were tested by calculating their fringe visibilities and adding noise to simulate real data. The simulated data were then Fourier transformed and cleaned. The results of some of these simulations are presented in Figs. 3a and 3b, where the correct source structure is compared with the full simulated, transformed, and cleaned maps. The quality of these and other simulated maps indicates that source structures such as a single large elongated component or a double source are easily identified and mapped with the U-V coverage available for VY CMa. The simulations also indicated that with the very limited U-V coverage available in this set of experiments, that data noise had far less effect upon the final cleaned maps than the cleaning process itself. Fig. 3c shows the limitations of the cleaning process on a very complex (five component) source when no noise is added to the simulated data.

Two procedures were used to approach the cleaning process. First, a direct cleaning of the entire region to be mapped was done for most spectral channels. Examples of some of these cleaned maps are shown in Figs. 4, 5, 6 and 7. The total power spectrum of the 1612-MHz emission of VY CMa is presented in Fig. 8. It became clear from the simulations that such a cleaning process introduced errors of roughly 25% of the peak brightness of the map (see Fig. 3c). Thus, spurious
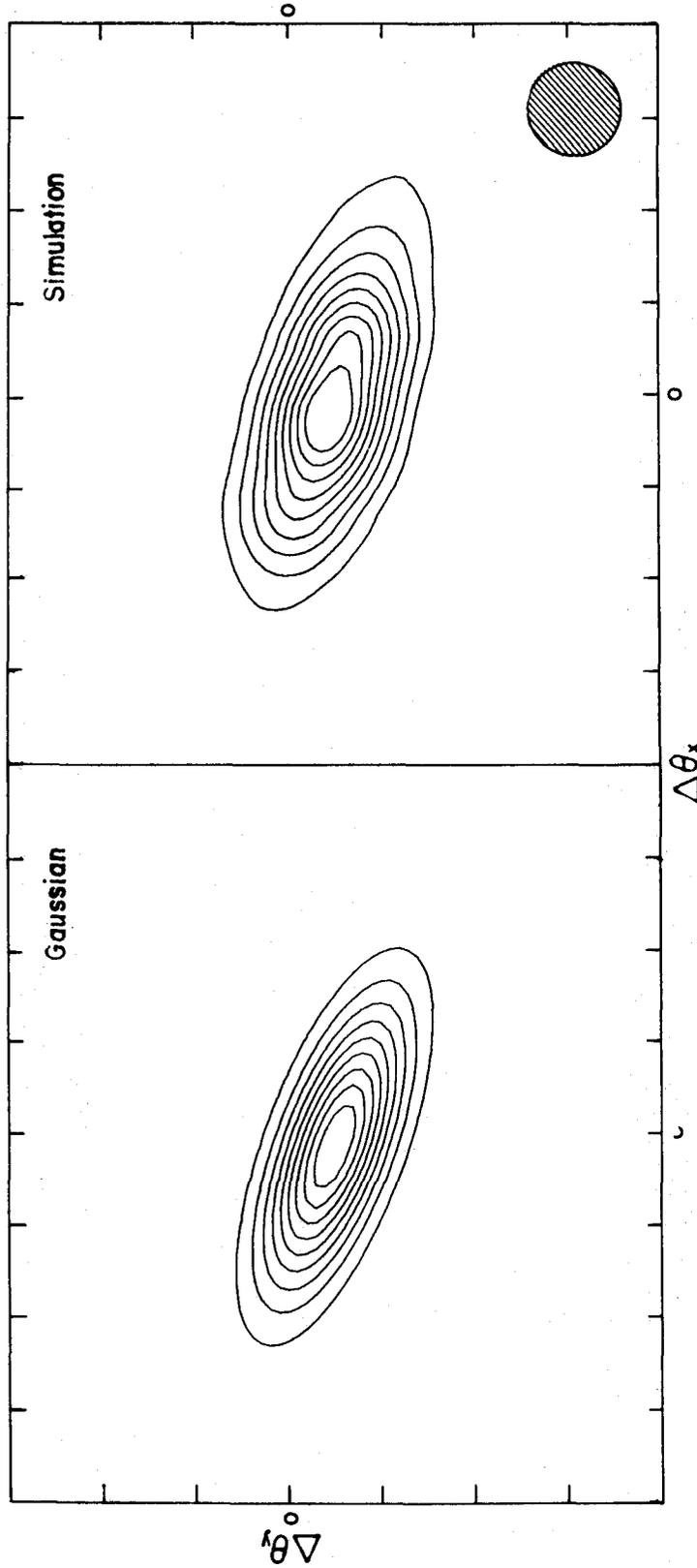
Figure 3a.--Simulation of the Fourier inversion and cleaning process for the U-V coverage of
Fig. 1. Fringe visibilities for a gaussian ellipse (left) were calculated and noise added.
The resultant simulated data were inverted and cleaned. The cleaned map (right) is shown
with contour intervals of 10% of the peak with the zero level suppressed. Tick marks on
both axes are 0.1 arc sec.
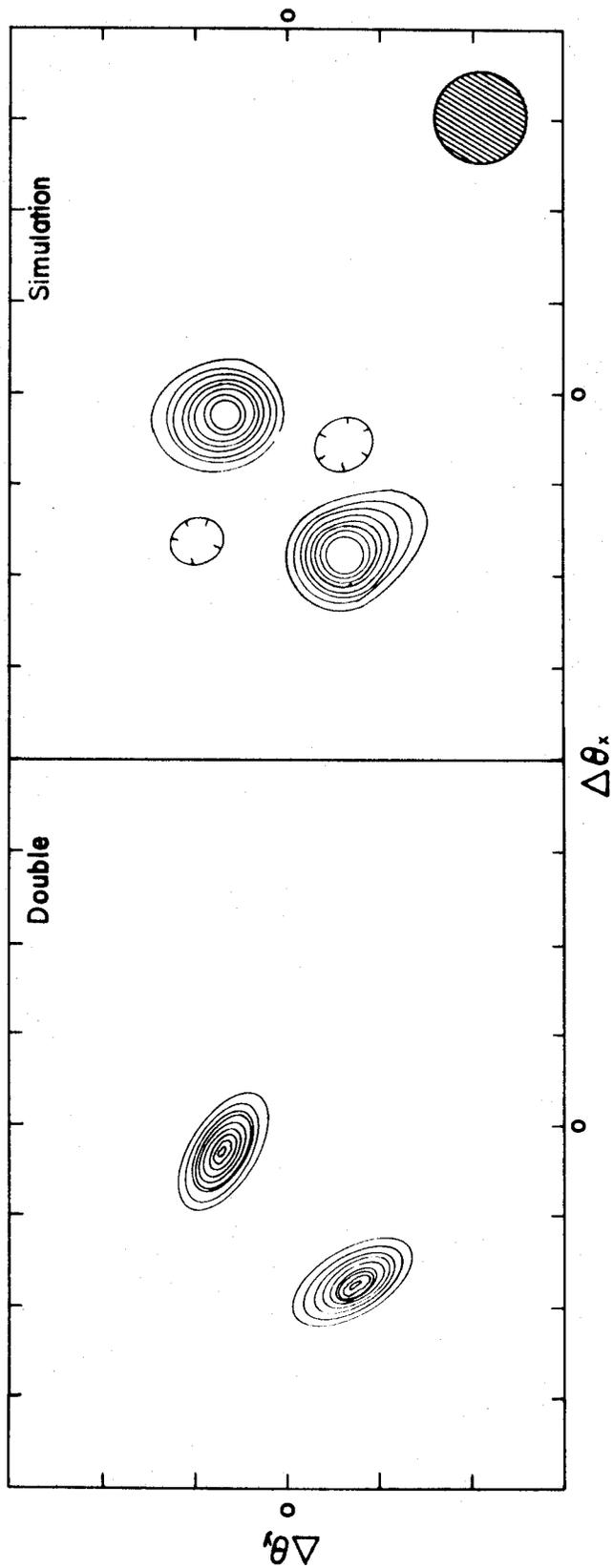
Figure 3b.--Simulation of the Fourier inversion and cleaning process for the U-V coverage of Fig. 1. Fringe visibilities for a double source (left) were calculated and noise added. The resultant simulated data were inverted and cleaned. The cleaned map (right) is shown with contour intervals of 10% of the peak with the zero level suppressed. Tick marks on both axes are 0.1 arc sec.
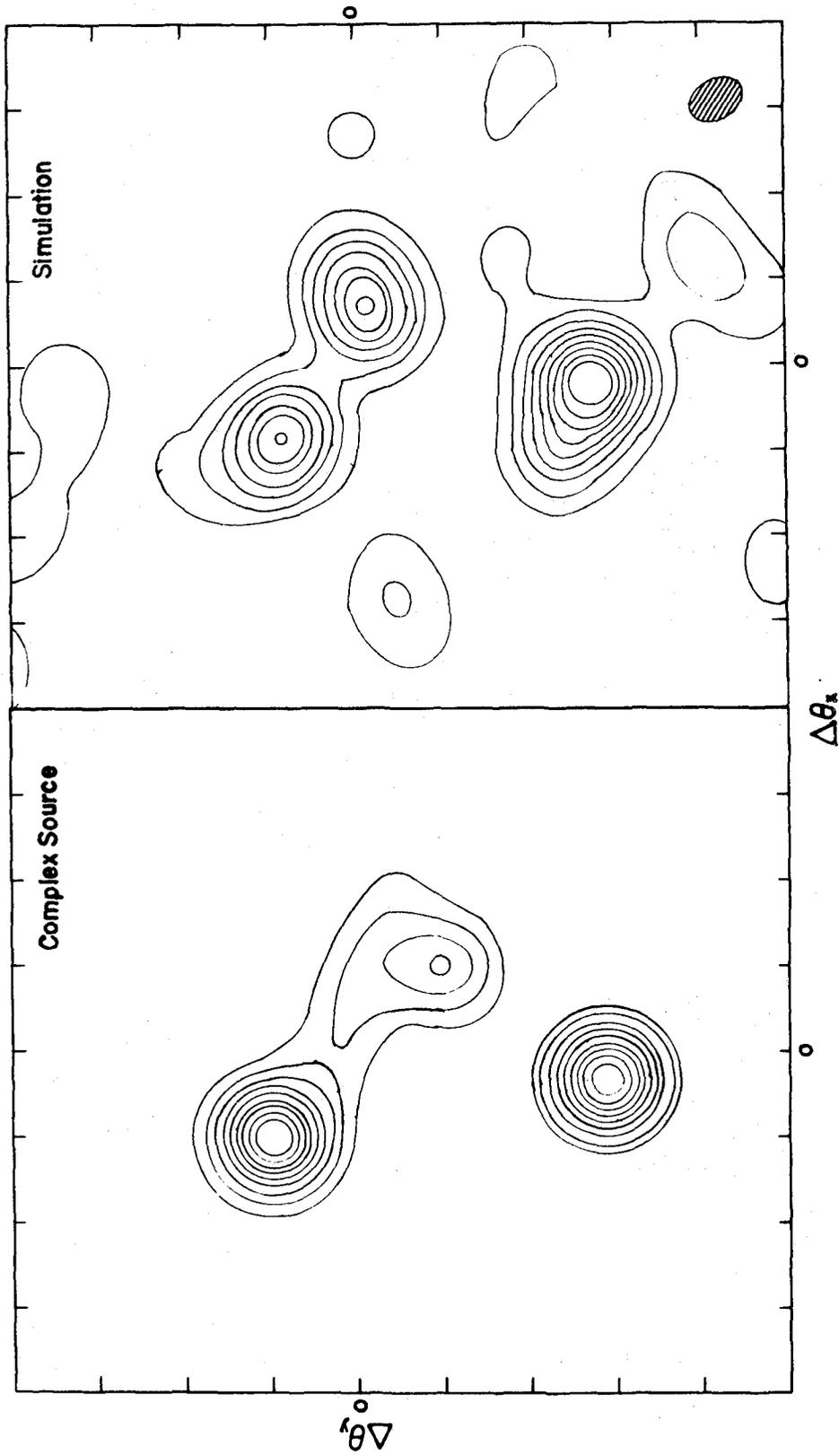
Figure 3c.--Test of the cleaning process. Noise free data for the complex source (left) was Fourier inverted and cleaned. The cleaned map (right) is shown with contour intervals of 10% of the peak with the zero level suppressed. The dimensions of the map (1".5 x 1".5) and the clean beam, indicated at the lower right of the map, are identical to those used in the actual maps presented in Figs. 4-7.
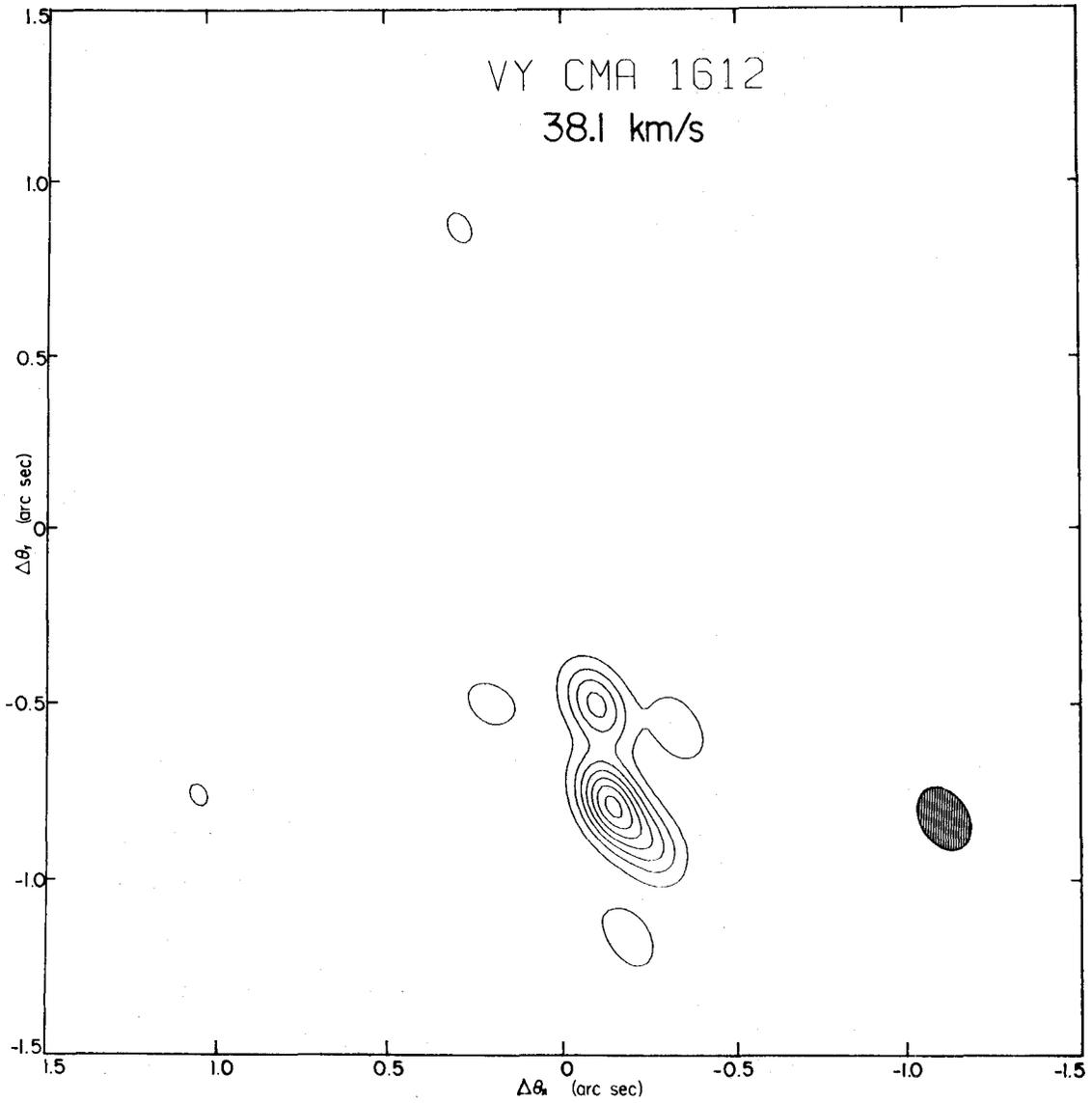
Figure 4.--Map of the 1612-MHz emission of VY CMa at 38.1 km/s. Contour levels are 10°K with the zero level contour suppressed. 1 Jy corresponds to 18°K for a point source with the 0.20 by 0.13 arc sec clean beam indicated in the lower right of the map.
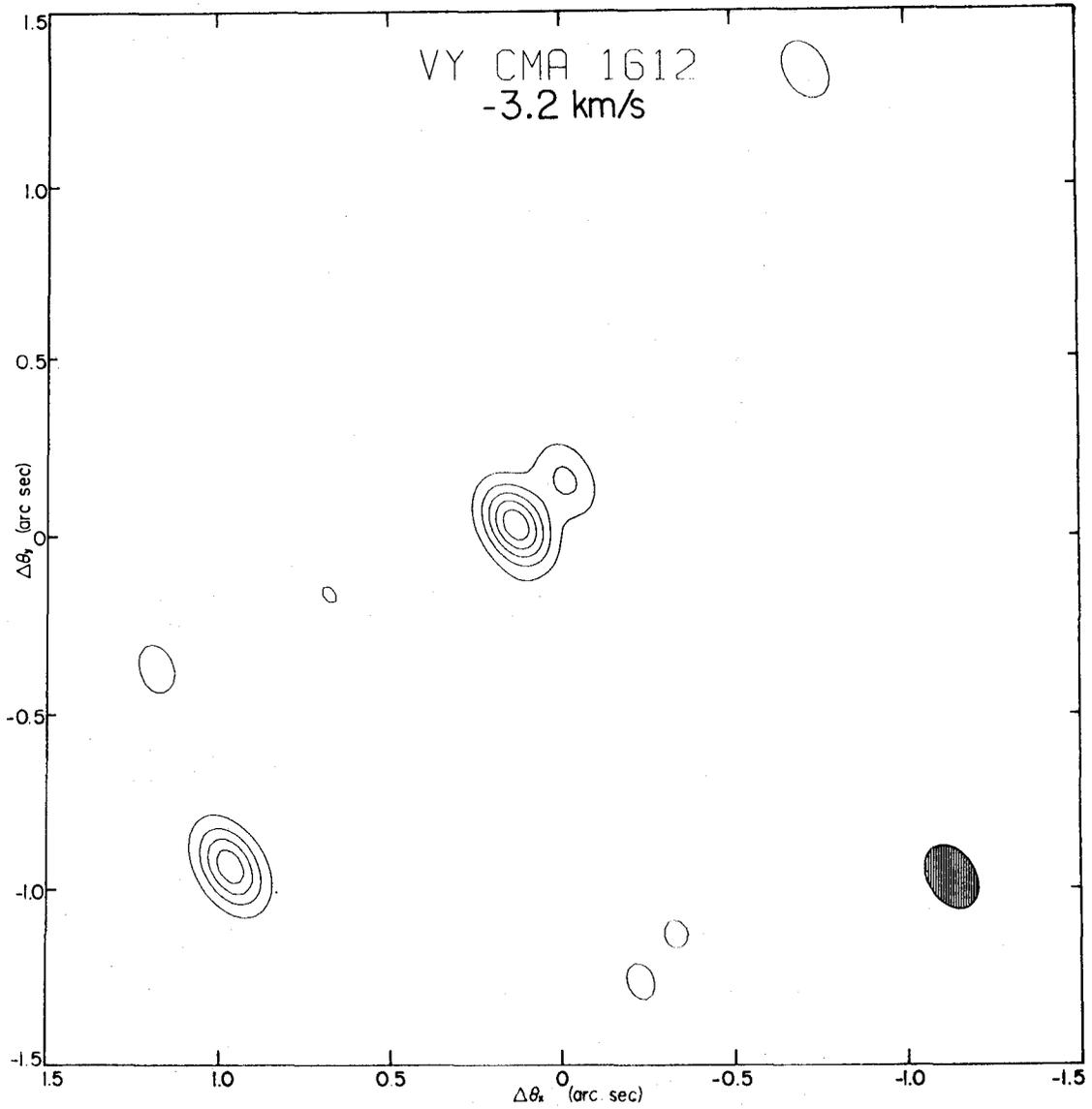
Figure 5.--Map of the 1612-MHz emission of VY CMa at -3.8 km/s. Contour levels are 10°K with the zero level contour suppressed. 1 Jy corresponds to 18°K for a point source with the 0.20 by 0.13 arc sec clean beam indicated in the lower right of the map.
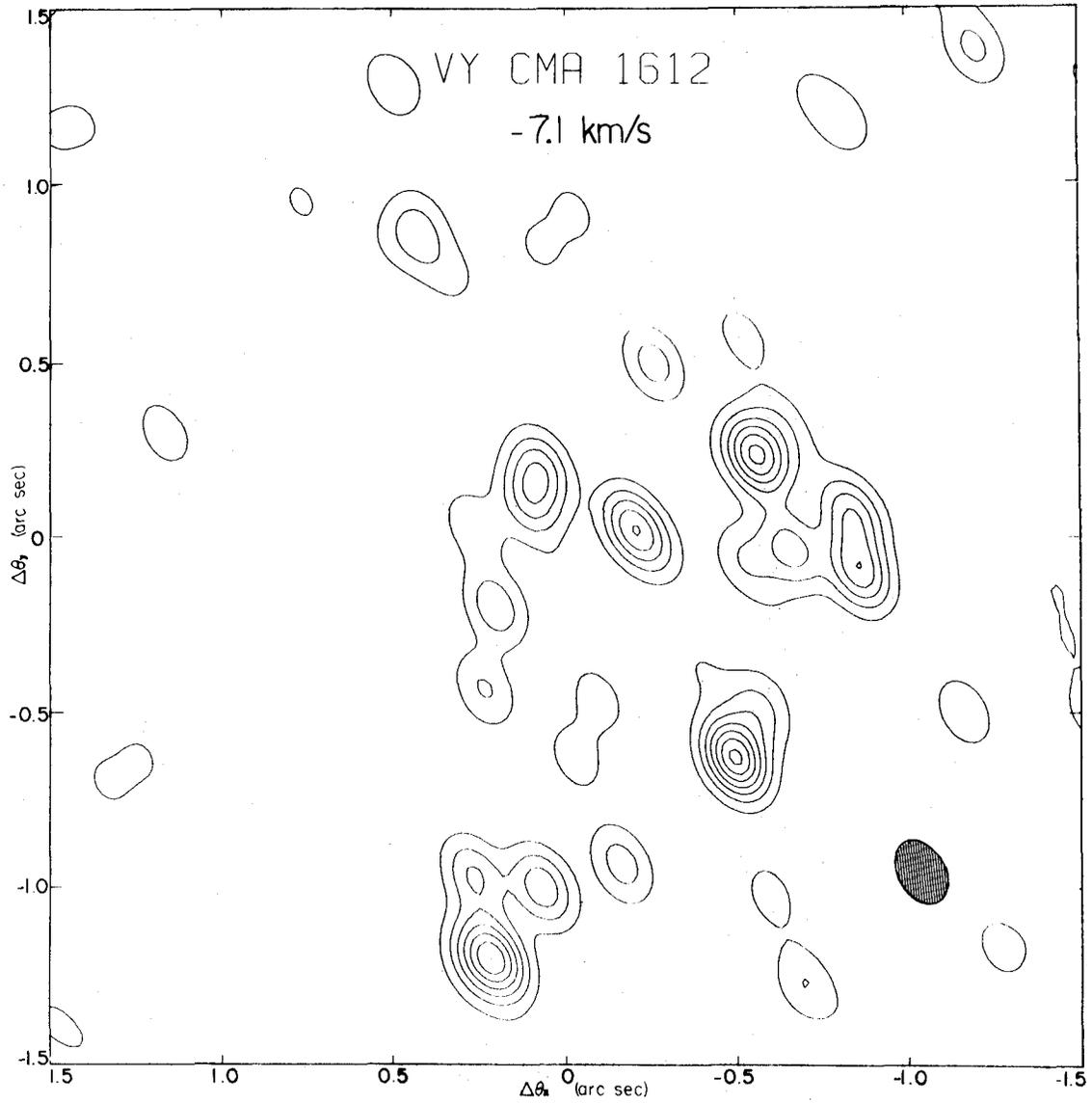
Figure 6.--Map of the 1612-MHz emission of VY CMa at -7.1 km/s. Contour levels are 10°K with the zero level contour suppressed. 1 Jy corresponds to 18°K for a point source with the 0.20 by 0.13 arc sec clean beam indicated in the lower right of the map.
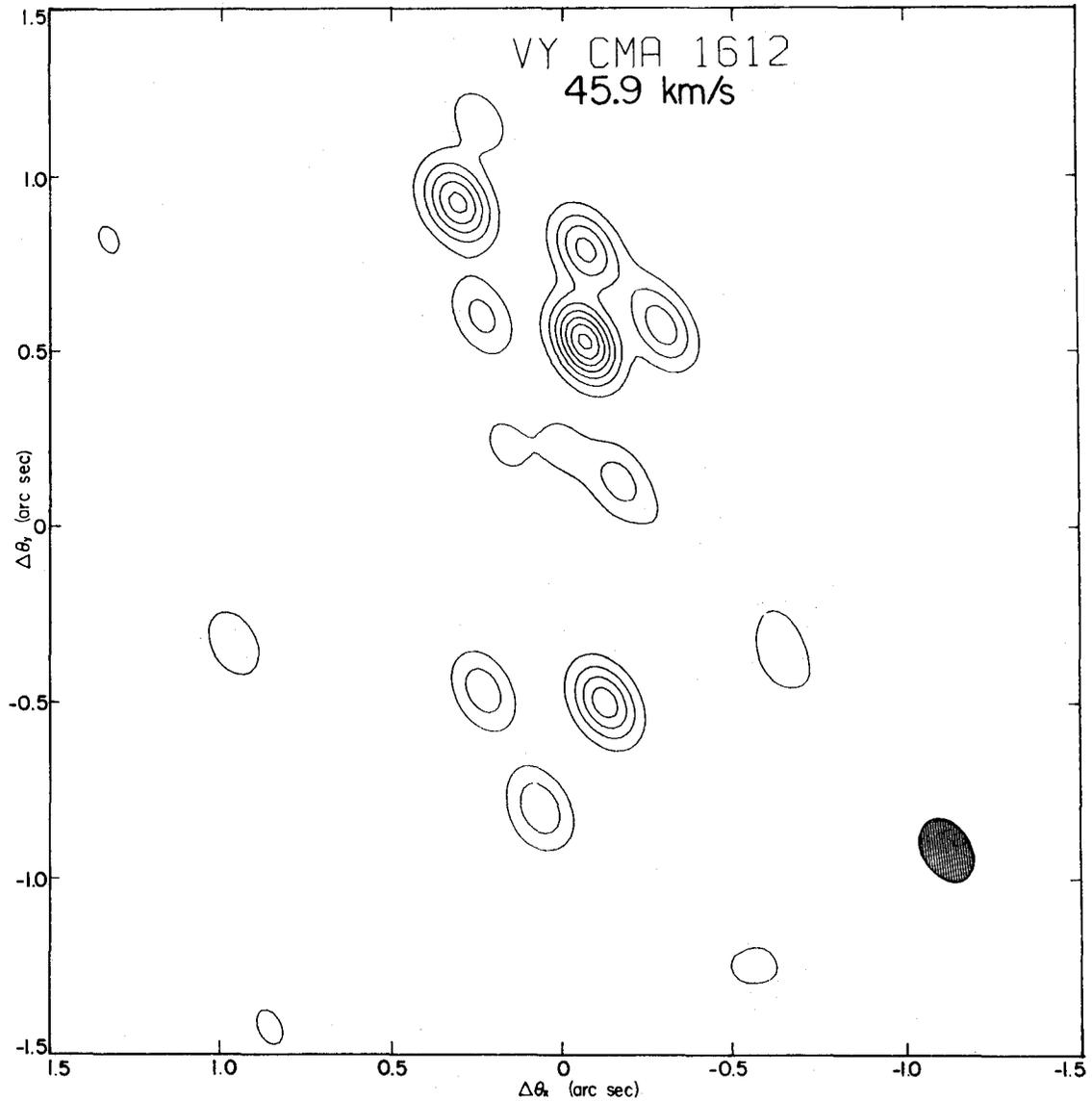
Figure 7 - Map of the 1612 MHz emission of VY CMa at 45.9 km/s. Contour levels are 10°K with the zero level contour suppressed. 1 Jy corresponds to 18°K for a point source with the 0.20 by 0.13 arc/sec clean beam indicated in the lower right of the map.
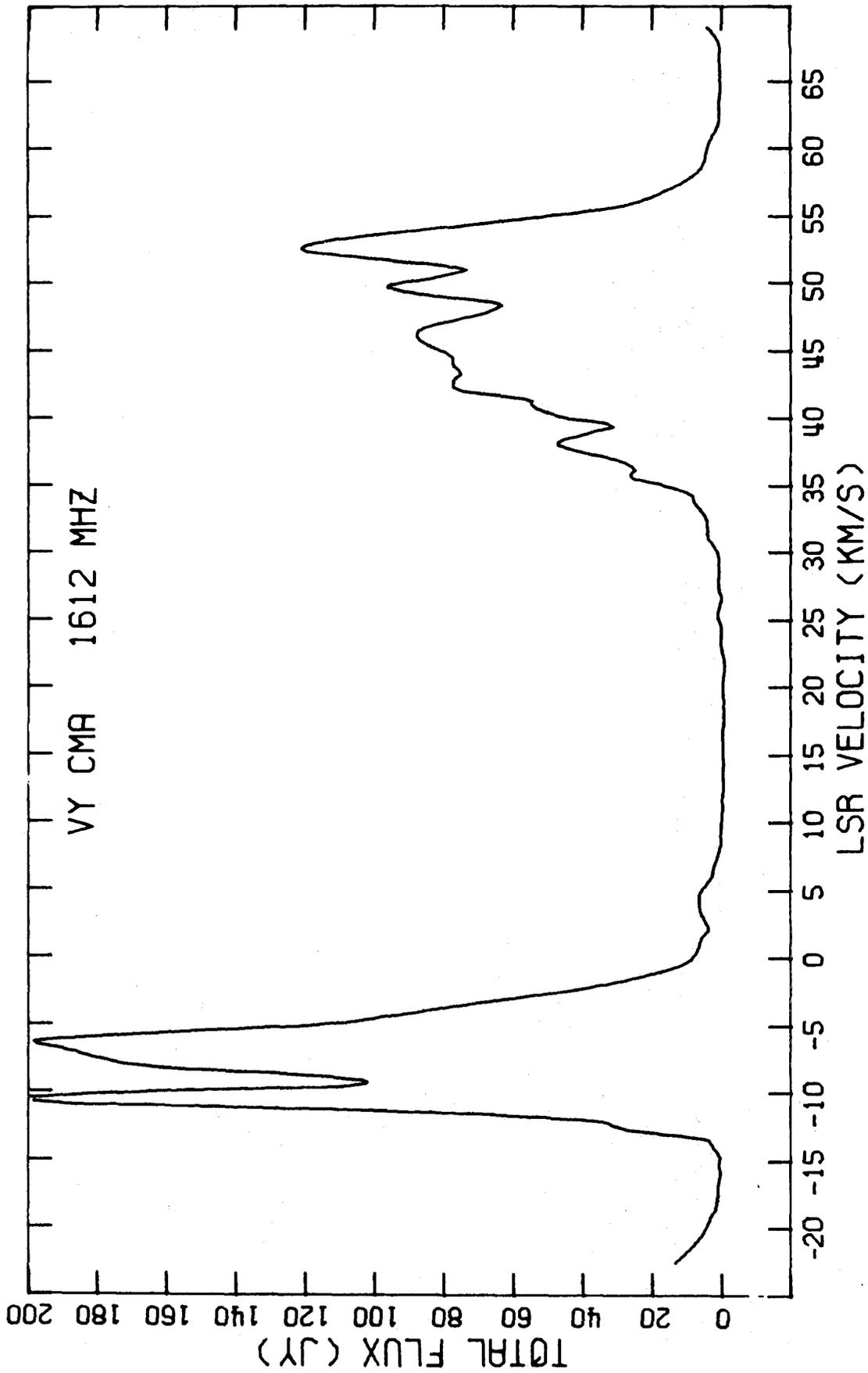
Figure 8.--Total power spectrum of the 1612-MHz emission of VY CMa. The spectrum is from a 144 point hanning weighted autocorrelation function and has a resolution of 1.3 km/s.

components of this strength probably appear in these maps and the

strengths of real components probably have similar uncertainties.

The direct cleaning process indicated that most of the spectral

channels of VY CMa contained from two to four components of sufficient

strength to be real. These components seemed generally to be small

and isolated. Since it would be very difficult and confusing to

present the structure of VY CMa by displaying maps of all the spectral

channels analyzed, an alternative, and very conservative, cleaning

process was employed. In each spectral channel, the strongest feature

in the dirty map (i. e., the uncleaned or raw inverted map) was

identified and the component subtraction of the cleaning process

(Högbom, 1972) limited to a square area or box $0\overset{..}{}35 \times 0\overset{..}{}35$ centered

upon the strongest feature. After the effect of the strongest component

was removed from the dirty map, the residual map was examined to find

the strongest remaining component. The original dirty map was then

reanalyzed by simultaneously cleaning in two boxes centered upon the

two strongest components. This procedure was continued in an iterative

fashion until the noise level left in the residual dirty map did not

significantly decrease with the addition of a new box. In most cases

this required about three cleaning boxes. By comparing maps of

adjacent spectral channels it appeared that the positions of the

strongest two components in each map often repeated.

A composite map of the two strongest components in each spectral

channel of VY CMa is presented in Fig. 9a. In many cases, components

in different spectral channels overlap. Accuracies of the positions

of such components should be better than $0\overset{..}{}1$. When components of a
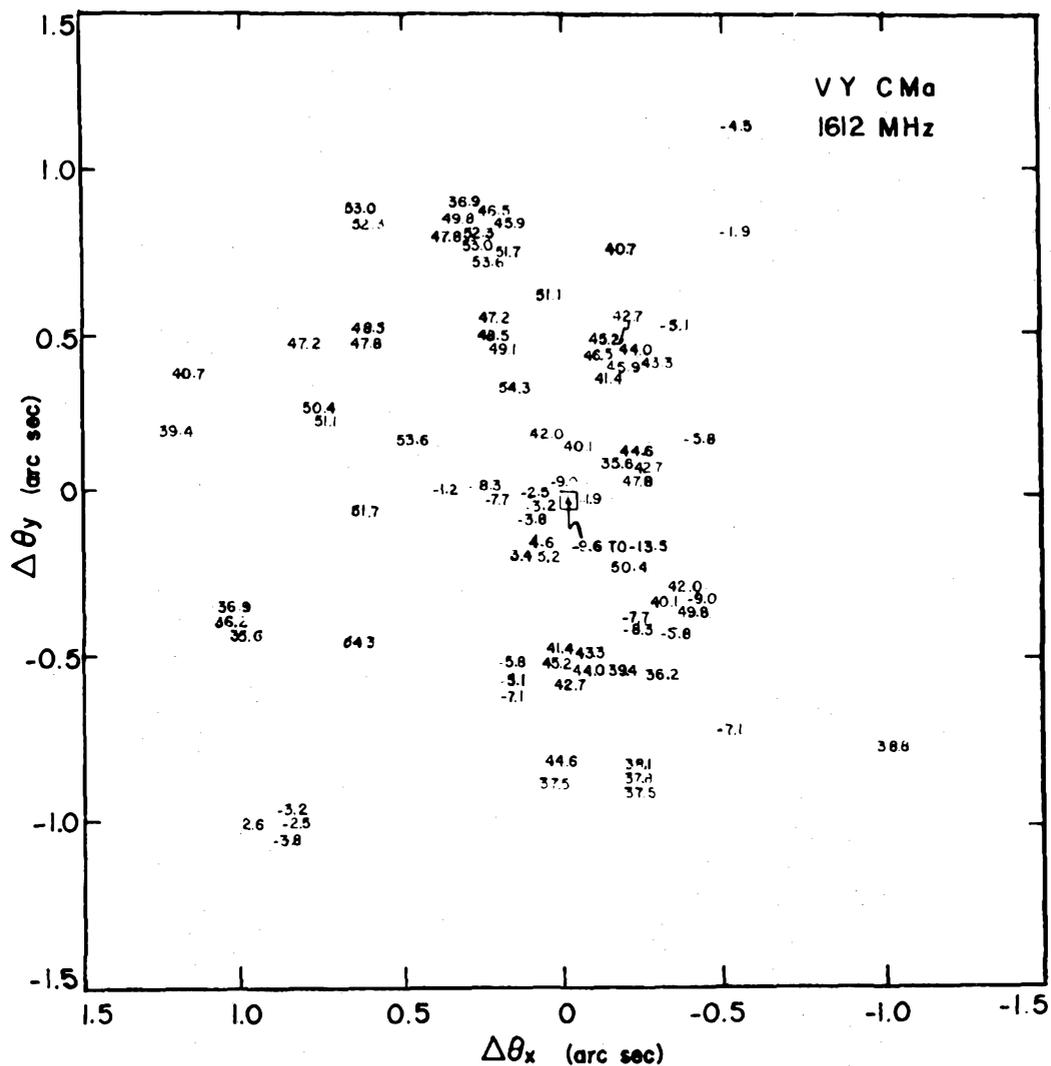
Figure 9a.--Raw composite map of the 1612-MHz emission of VY CMa indicating the positions
of the two strongest components in each spectral channel determined by Fourier inver-
sion and cleaning. Position uncertainties are discussed in the text.
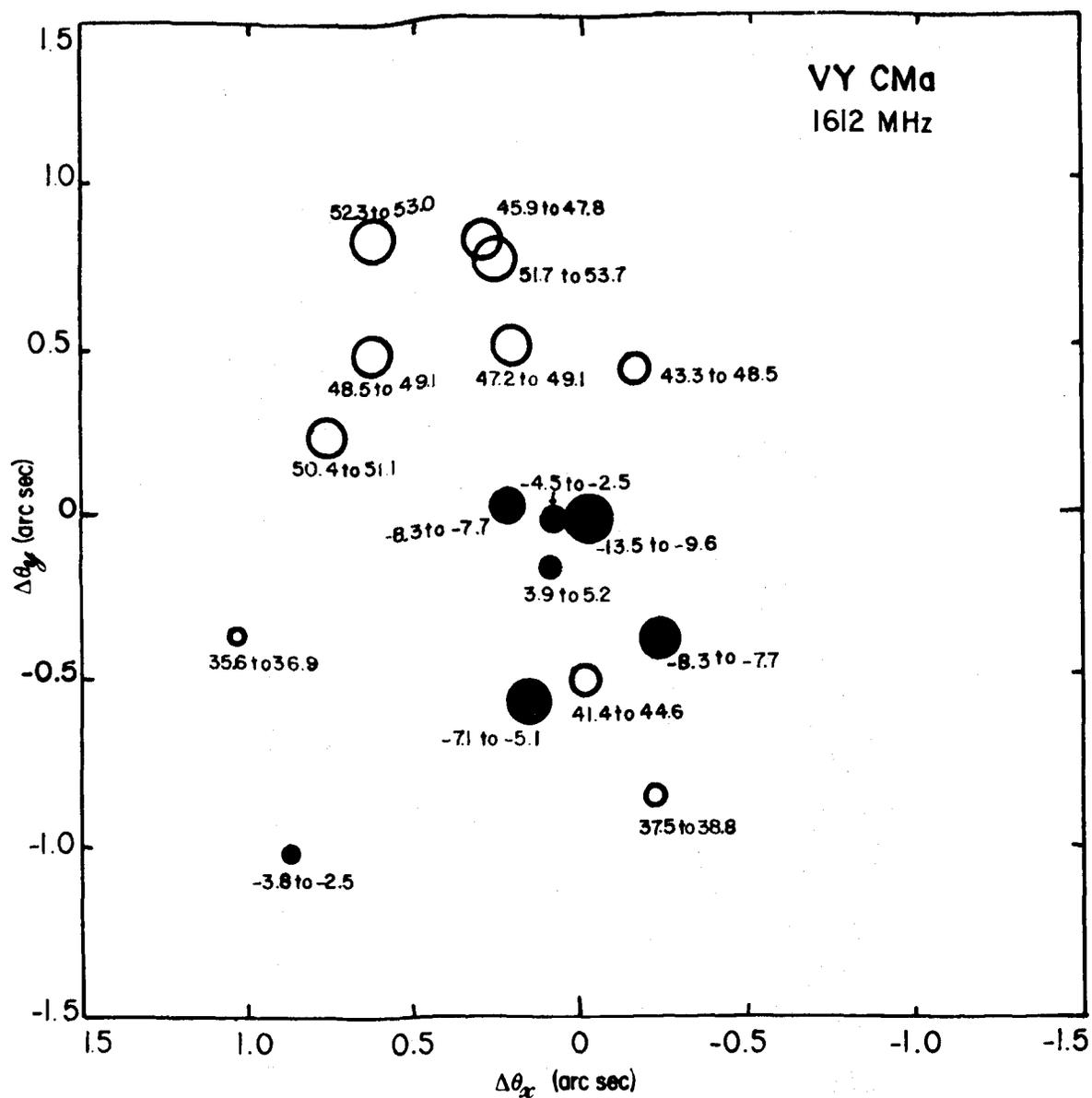
Figure 9b.--Composite map of the 1612 MHz emission of VY CMa adopted
in this work. Only maser components whose positions in adjacent
spectral channels agreed to within 0."1 are included. This is a
very conservative map, since all of the maser components included
have well determined positions. Many maser components have not
been plotted due to large uncertainties in their positions. The
size of the circles is proportional to the strength of the total
power at that velocity.
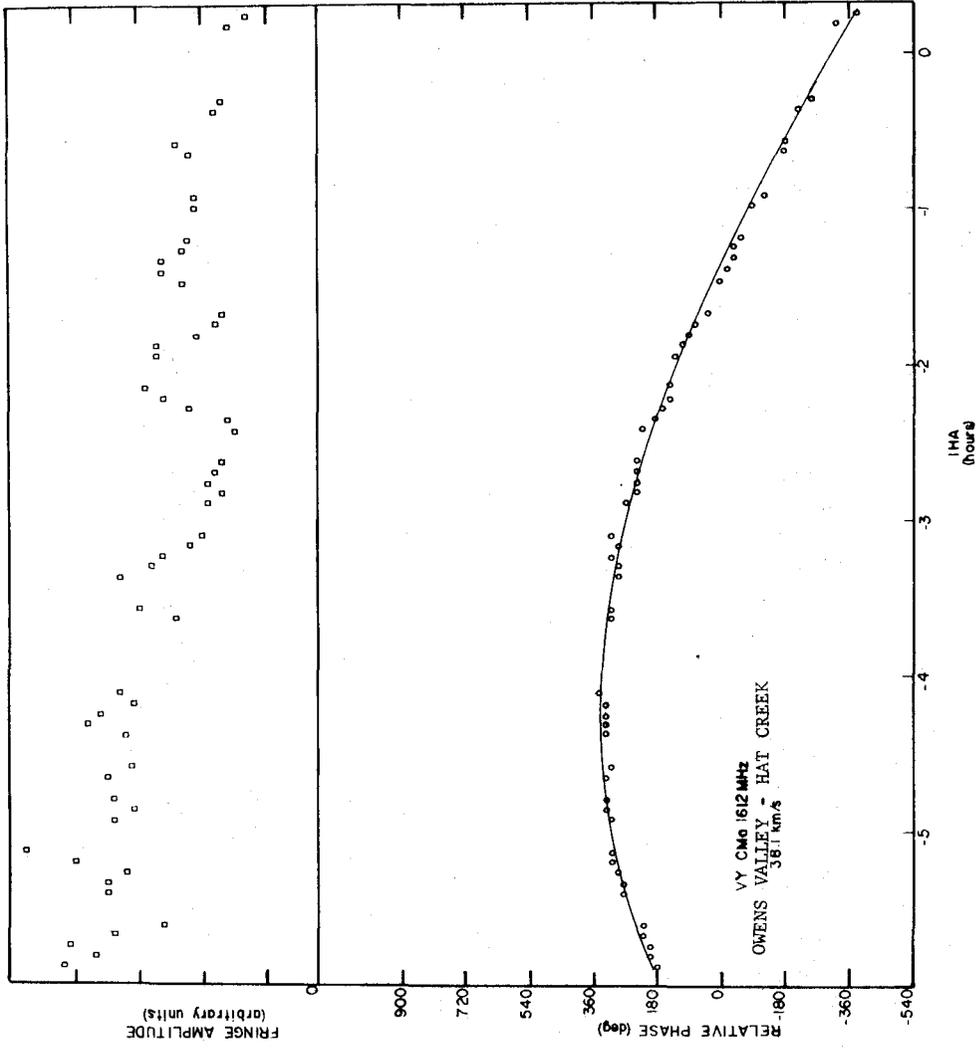
Figure 10a.--Fringe amplitudes and phases in the 38.1 km/s spectral channel for the Owens Valley to Hat Creek baseline. The solid line through the phase points is the expected phase variation for a position offset of -0.180 in α and -0.733 in δ determined by least-squares fitting of the data from both baselines (see Fig. 10b) simultaneously. The cleaned map (Fig. 4) indicates that fringe power in this channel is dominated by one component at the determined position.

VY CMa 1612MHz
GREENBANK – MARYLAND POINT
38.1 km/s

IHA
(hours)

FRINGE AMPLITUDE
(arbitrary units)

RELATIVE PHASE (deg)
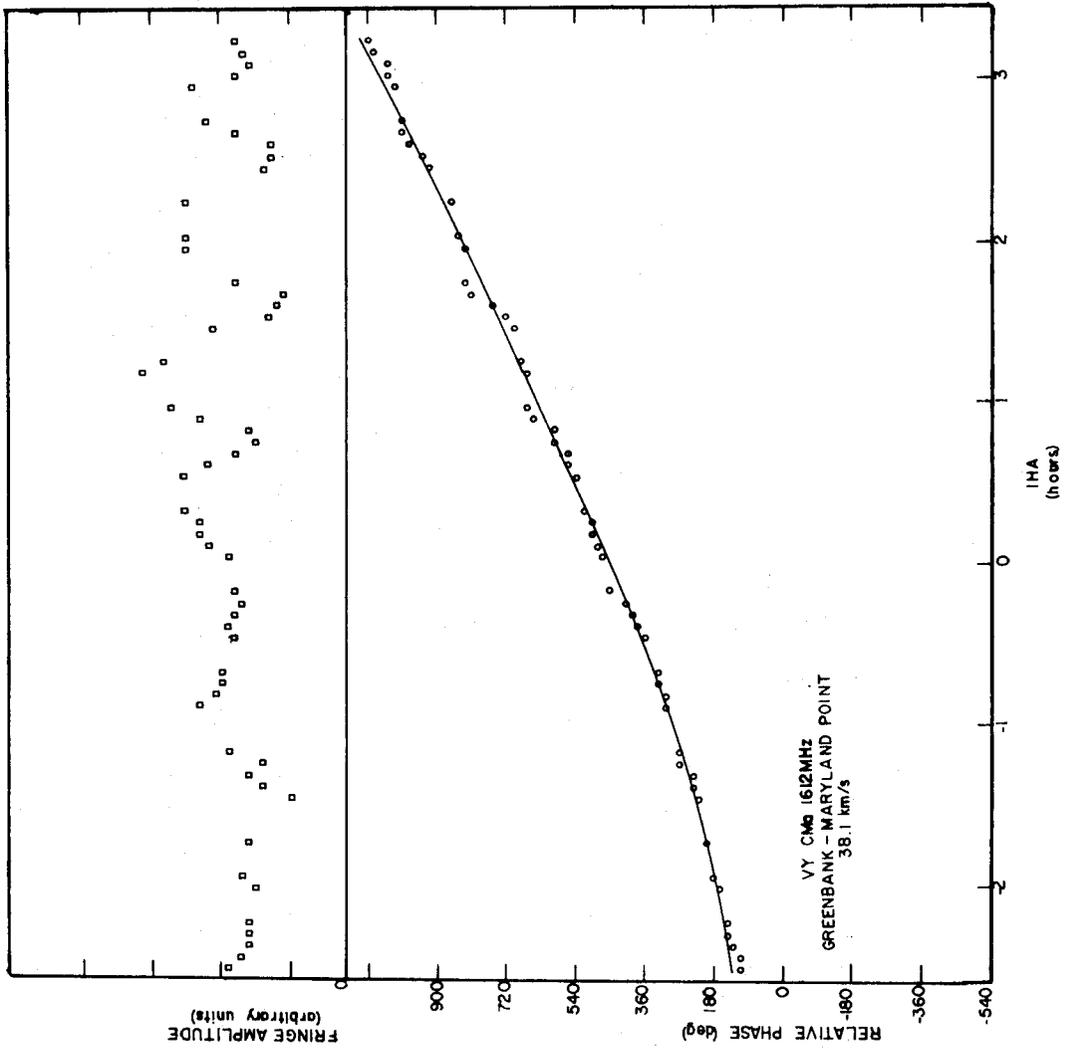
Figure 10b.—Fringe amplitudes and phases in the 38.1 km/s spectral channel for the Greenbank to Maryland Point baseline. See the caption for Fig. 10a for more details.

spectral channel appear isolated, the possibility of side lobe confusion cannot be ignored. An estimate of the position uncertainties of isolated components should include the possibility of mistaking the first large positive sidelobe of the dirty beam. Thus, an uncertainty of $\approx 0\overset{..}{.}5$ is adopted for these components. With these uncertainties in mind, a very conservative composite map of the 1612-MHz emission of VY CMa is presented in Fig. 9b. While many weak maser components have been left out of the map, the positions of the stronger components depicted are almost certainly correct.

In several spectral channels the fringe power is dominated by a single component and a position offset can be determined for this component by modeling the fringe phases. The agreement between the inverted map and phase modeled positions of these components is excellent. The relative fringe phases from the two baselines are shown in Fig. 10 for the 38.1 km/s component. The phases predicted by the position offset of this component are given by the solid line. This channel appeared to be dominated by a single component in the Fourier inverted and cleaned map (see Fig. 4). In most spectral channels, however, the blending of several spectral components precludes a very meaningful mapping by modeling the fringe phases. For example, the relative fringe phases from the two baselines are shown in Fig. 11 for the -3.8 km/s channel. This channel appeared to contain a widely spaced double in the Fourier inverted and cleaned map (see Fig. 5). The oscillation of the fringe amplitudes (also shown in Fig. 11) and the abrupt phase jumps at amplitude nulls clearly demonstrate that the fringe power at -3.8 km/s is not from one component. The position
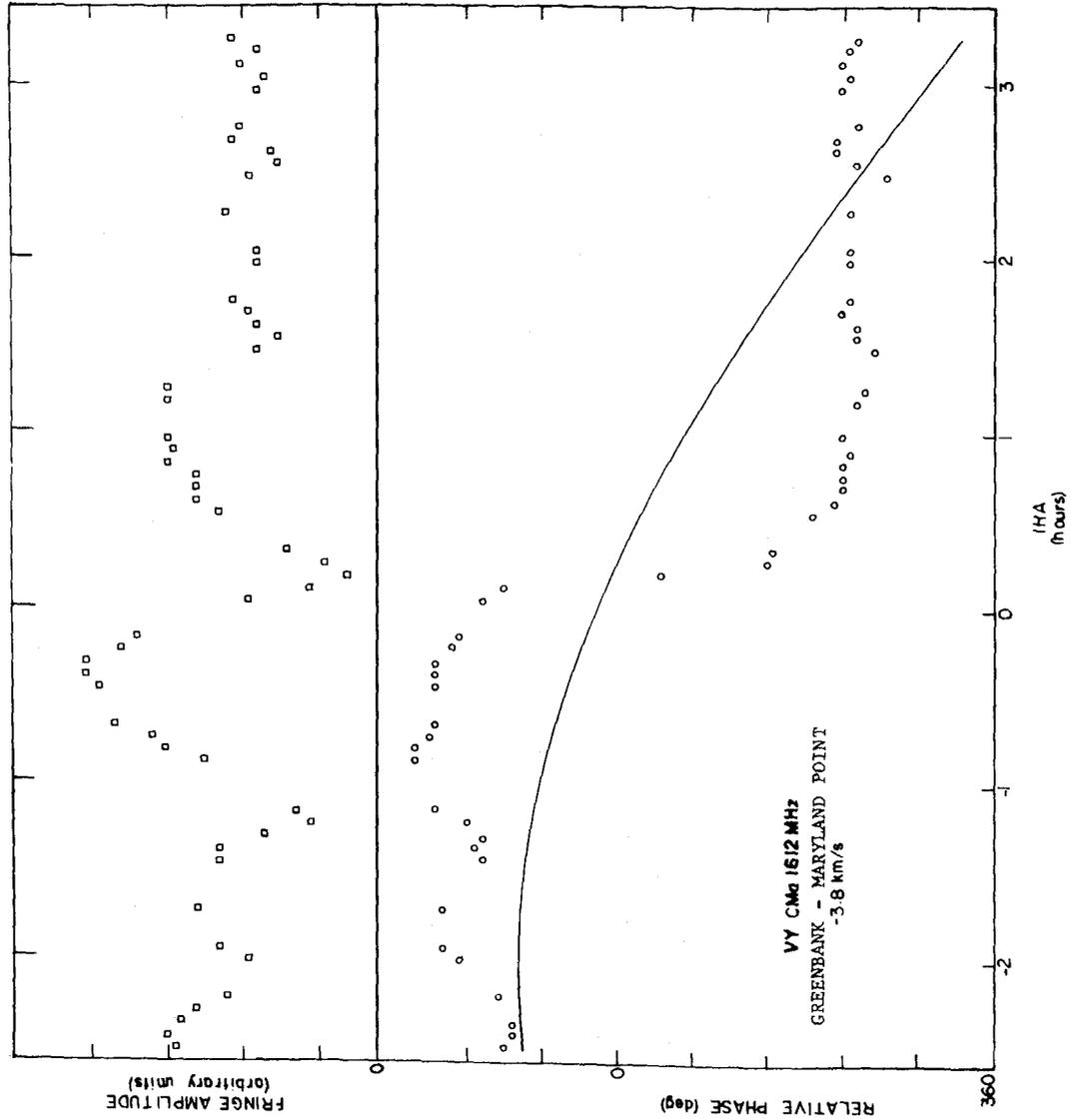
Figure 11a - Fringe amplitudes and phases in the -3.8 km/s spectral channel for the Greenbank to Maryland Point baseline. The solid line through the phase points is the expected phase variation for a position offset of 0.19 in α and 0.29 in δ determined by least-squares fitting of the data from both baselines (see Fig. 11b) simultaneously. The cleaned map for the adjacent spectral channel indicates a widely spaced double. This example indicates the inadequacy of a single component phase model for such source structures.
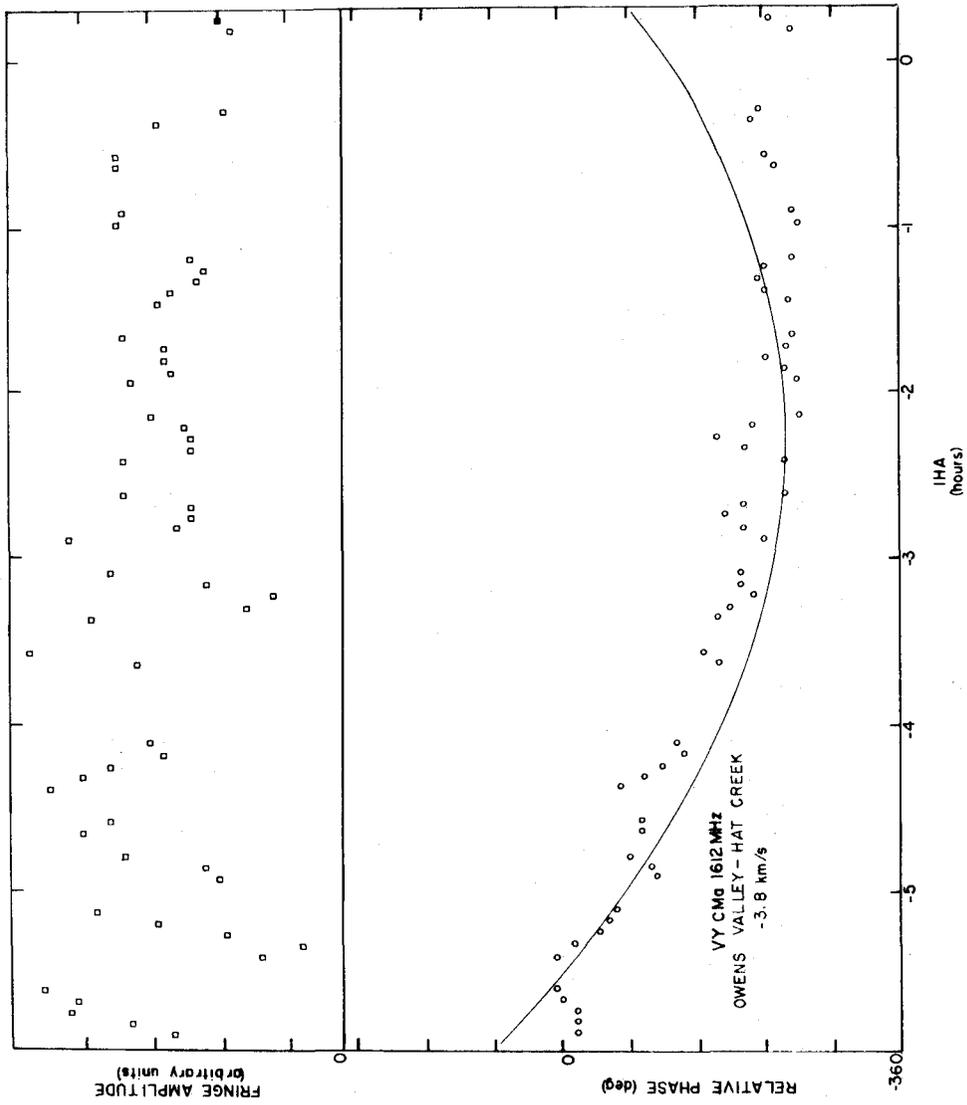
Figure 11b - Fringe amplitudes and phases for the -3.8 km/s spectral channel for the Owens Valley to Hat Creek baseline. See the caption for Fig. 11a for more details.

offset determined by modeling the fringe phases as one component is

0''2 in right ascention and 0''3 in declination which does not give the

position of either of the two strongest components or the centroid

of emission at -3.8 km/s. This example demonstrates the superiority

of the Fourier inversion and cleaning process over simple phase modeling

when the source structure is complex.

b) Discussion

The composite map (Fig. 9b) of the 1612-MHz emission of VY CMa

presented in this chapter can be compared with two other maps.

Masheder et al. (1974) attempted to map VY CMa with the 24 km micro-

wave linked Jodrell Bank interferometer. Their approach was to model

the fringe phases as one component and to determine a position offset

in each spectral channel. Although very high spectral resolution was

used, most of the channels analyzed showed signs of a complex structure

probably due to a blending of lines in the source. For this reason

Masheder et al. suggest caution in adopting the position offsets of

most of the components mapped.

The map of VY CMa by Masheder et al. (1974), henceforth called

the Jodrell map, qualitatively agrees with the map presented in this

chapter only if both the right ascension and declination axes are

inverted. Position offsets of the correct magnitude but oriented 180°

from the correct direction can be caused by inconsistencies in the

sign conventions used to calculate U and V and the residual fringe

phase. The possibility that the position offsets in the Jodrell map

have the wrong sign (or orientation) was first noticed by Moran

(1975). Moran compared single telescope maps of H_2O maser sources

with those generated by spectral line

VBL observations and data reduction programs in order to determine the correct sign conventions. Additional checks of the spectral line programs have been performed by myself in which baseline corrections are determined from residual fringe rates of sources with well known positions. When the corrected baselines are used to reprocess the same Mk II video tapes, zero residual fringe rates are obtained. If inconsistent sign conventions existed in the reduction programs, the residual fringe rates of these sources would have been increased rather than reduced toward zero.

The Jodrell map with the axes inverted is given in Fig. 12. The Fourier inverted and cleaned composite map, henceforth called the inverted map, presented in this chapter (Fig. 9a, and 9b) can be compared with the Jodrell map as presented in Fig. 12. Approximately $\frac{1}{4}$ of the "components" in the Jodrell map agree within about $0\rlap{.}{''}2$ of the positions given in the inverted map. About $\frac{1}{2}$ of the "components" in the Jodrell map fall near the centroid of a complex structure (e. g., a double) in the inverted map. The remaining "components" are more than $0\rlap{.}{''}5$ away from the strongest components in the inverted map. These cases of strong disagreement usually involve spectral components which do not cluster close together in the inverted map and/or appear highly spectrally blended in the Jodrell spectrum of VY CMa. In these cases it is difficult to evaluate which map is better. However, over most of the 1612-MHz emission spectrum of VY CMa, the Fourier inverted and cleaned map indicates the presence of two or more components and should be more accurate and meaningful than the Jodrell map.

The other existing map of the 1612-MHz emission of VY CMa is by Moran et al. (1975). This map was constructed by modeling relative
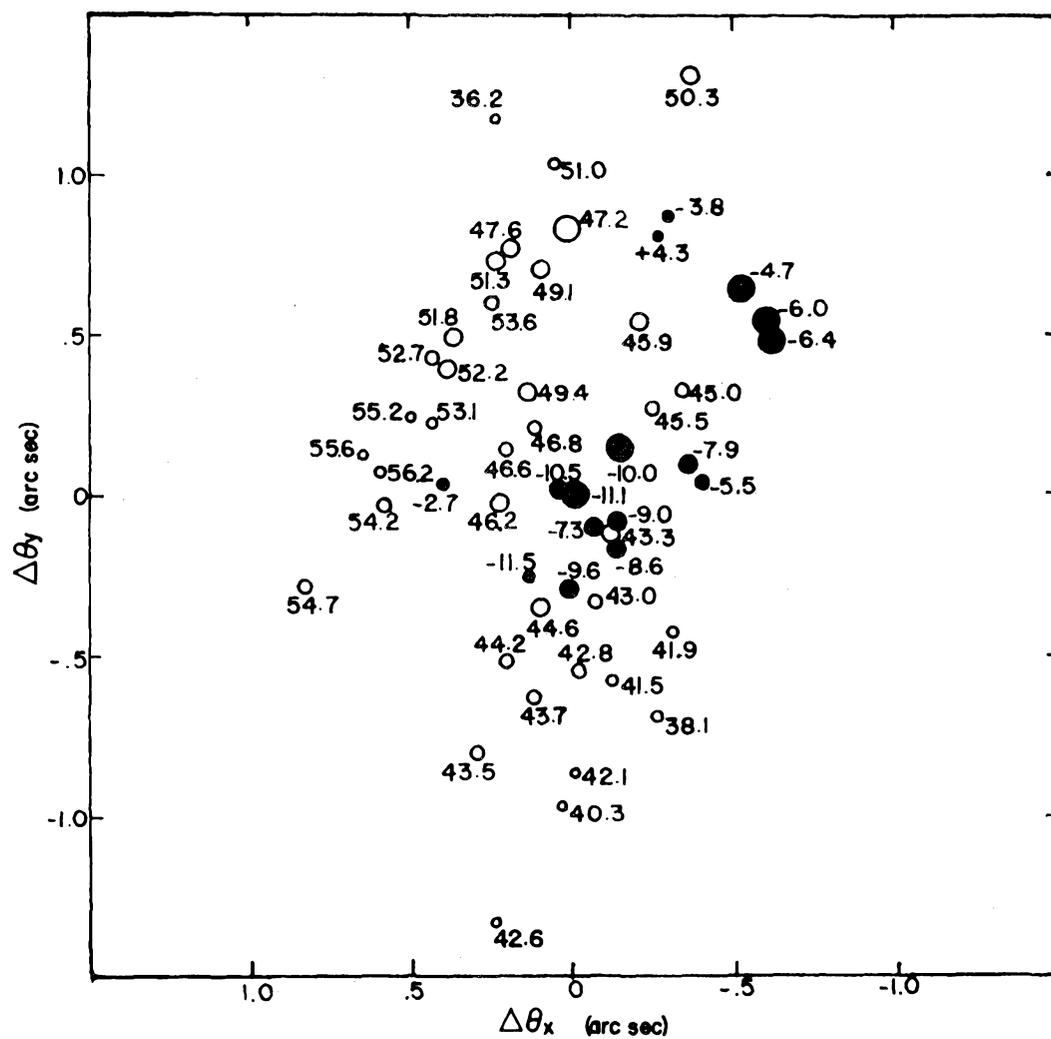
Figure 12.--Map of VY CMa at 1612-MHz adapted from Masheder et al. (1974). The sign on the two axes has been reversed, the origin shifted to the center of the -10.5 and -11.1 km/s components, and $\theta_x = \Delta\alpha\cos\delta$ plotted instead of $\Delta\alpha$ to facilitate comparison with the Fourier inverted and cleaned composite maps in Figs. 9a and 9b.

fringe rates from an 850 km baseline. This long baseline heavily resolves most of the maser components in VY CMa and makes it difficult to map them. Of the components mapped by Moran et al., about $\frac{1}{2}$ agree, within their uncertainties of about 0".2, with the inverted map presented in this chapter. Approximately $\frac{1}{4}$ of the components mapped by Moran et al. fall near the centroid of the more complex spectral channels, and the remaining $\frac{1}{4}$ of the components do not agree very well (e. g., differences $\gtrsim$ 0".5).

The 1612-MHz map of VY CMa presented in this chapter should be the most accurate presently available. It is based upon data from two baselines whose interferometer fringe spacings range from about 0".1 to 0".5. These spacings are suitably sensitive for mapping this source but do not heavily resolve isolated maser components. Further, the Fourier inversion process simultaneously uses both the fringe amplitude and the fringe phase data in each spectral channel to obtain relative position information. This allows complex structures, such as double or triple sources in one spectral channel, to be mapped, whereas the techniques employed by Masheder et al. (1974) and Moran et al. (1975) can yield misleading results. At this time, a conservative approach should be adopted toward accepting maps of very complex spectral sources such as VY CMa. Only those spectral components whose positions systematically agree over two or more spectral channels as in the composite Fourier inverted map (Fig. 9b) can be confidently accepted. One can be reasonably sure that their true uncertainties are only $\approx$ 0".1.

c) Interpretations

The composite map of the 1612-MHz emission of VY CMa obtained by
Fourier inversion and cleaning (Fig. 9b) does not lend itself to any
obvious, simple interpretations. However, several general points
can be made. As noted by Masheder et al. (1974) and Moran et al. (1975),
the blue-shifted emission complex components tend to be somewhat
enclosed by red-shifted emission complex components. However, some of
the emission near -3 km/s fall outside all of the red-shifted emission
components. In addition, there appears to be a tendency for the
velocity of the red-shifted emission complex components to decrease in
velocity as one goes clockwise around the map (about a central
position of 0".3 in right ascention and 0".0 in declination). In some
instances, several spectral components which are widely separated in
velocity appear to overlap on the map.

On the whole, the distribution of maser components as a function
of velocity does not appear very uniform or consistent. As mentioned
in Chapter II, inhomogeneities in the circumstellar envelope may
dominate the locations of the maser emission. In that chapter I
speculate that in addition to a single circumstellar envelope the
structure of the maser emission may be due to multiple concentric
expanding shells. This could explain the spatial overlapping of
maser components with greatly differing velocities in either the low
or high velocity emission complex.

III. 1665 MHz

a) Analysis Procedures

The 1665-MHz emission of VY CMa is weaker than the 1612-MHz emission. It is highly circularly polarized and the velocity difference between its two emission complexes is slightly less than for the 1612-MHz emission. The low velocity emission complex of VY CMa at 1665-MHz has been mapped with the Jodrell Bank interferometer (Harvey et al, 1975). The short, 23 km baseline used was adequate to discover that the 1665 (and 1667) MHz emission is confined to a much smaller region than the 1612-MHz emission. However, the small spatial separations of the maser components in the low velocity emission complex were barely resolved.

The 1665-MHz emission of VY CMa was observed with the Owens Valley to Hat Creek baseline (see Table 1 of Chapter II for a description of the telescope and equipment employed). The observations were conducted with a linearly polarized feed (PA = 90°) which should accept both the left and right circularly polarized maser components (with a loss of a factor of two in signal to noise). This baseline is 550 km and affords adequate spatial resolution to map the very compact structure of VY CMa at this frequency. Since data from only one baseline was available, no attempt was made to Fourier invert the data. Instead, a map of most of the maser components was determined by modeling the fringe phase data in each spectral channel. A discussion of this technique is in Chapter II and in the appendix.

b) Discussion

The cross-correlated spectrum of VY CMa at 1665-MHz is shown in Fig. 13. Most of the spectral features were either unresolved or only

VY CMA   1665 MHZ
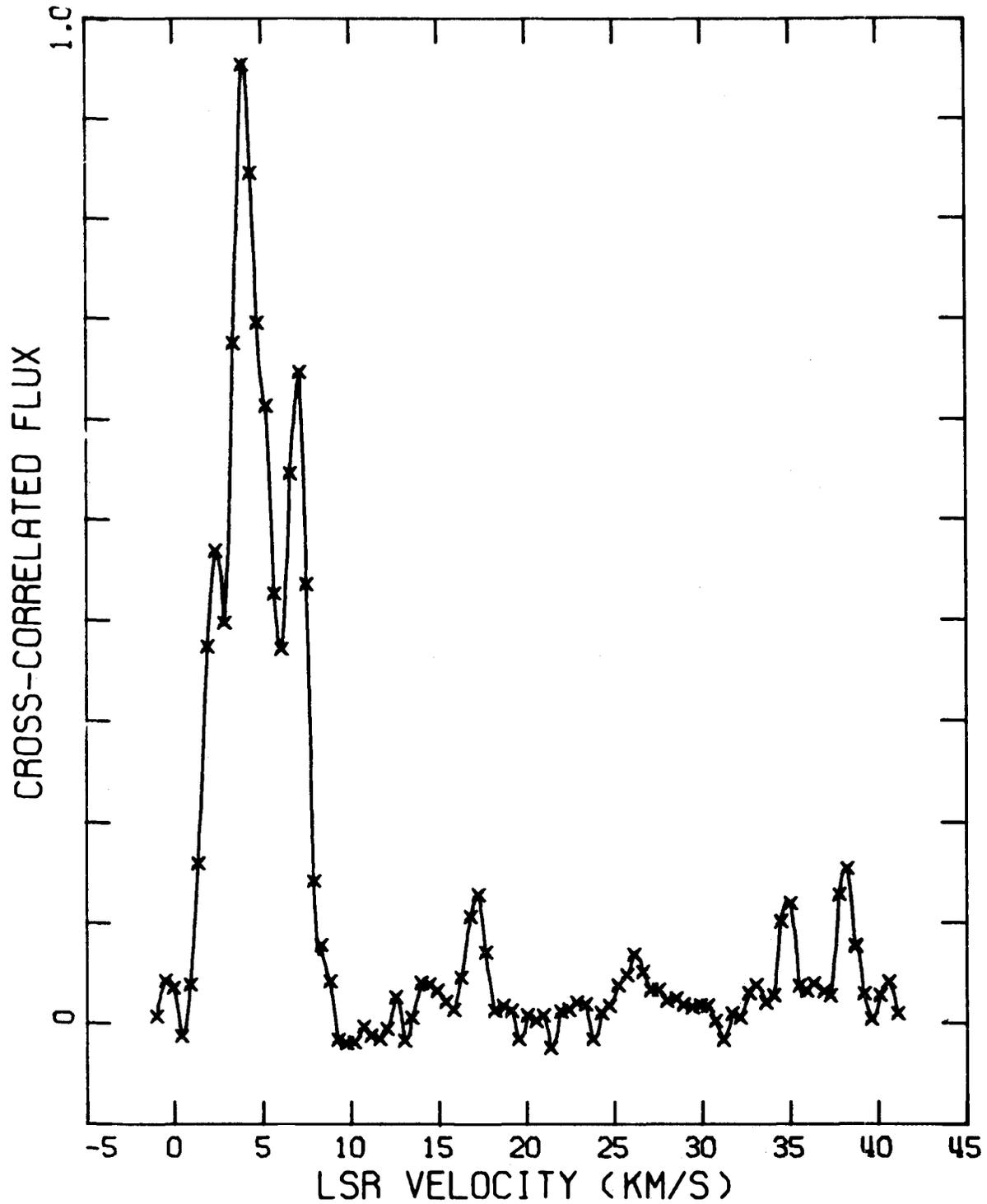
Figure 13.—Interferometer spectrum of the 1665-MHz emission of VY CMa. The data is an average of all scans from the Owens Valley to Hat Creek baseline with linearly polarized feeds at 90° PA. The spectrum is from a 96 point hanning weighted cross-correlation function with a resolution of 1.0 km/s.

partially resolved with this baseline. This indicates that the

apparent sizes of these maser components are less than approximately

0".05. Thus, the apparent sizes of maser components of VY CMa at

the main lines (i. e., 1665 and 1667-MHz) are somewhat smaller than

the typical 0".1 sizes at 1612-MHz.

The spatial orientation of most of the spectral components of the

1665-MHz emission of VY CMa is shown in Fig. 14. All of the maser

components with strong signals in the cross-correlated spectrum are

mapped with uncertainties of less than 0".02. The strongest compo-

nents in the low velocity emission complex cluster near the reference

component, 6.8 km/s, with the exception of a component (or components)

at about 3 to 4 km/s. This component lies about 0".3 south-west of the

reference component. There are two very weak centers of emission

in the high velocity complex at 37 and 35 km/s. These components

could not be reliably mapped because of their low signal to noise

ratios and the existence of a gap in the data caused by a "bug" in the

new 288 channel VLB processor. It may be possible to reprocess this

data in the future and accurately map these two components. With the

present data, the 37 km/s component could either lie about 0".05 south

or roughly 0".5 south-south-west of the reference position. In the

latter case, this would place the 37 km/s component close to the 4 km/s

component. The fringe phase of the 35 km/s component appeared to

change rapidly between scans which would indicate a position offset

$\gtrsim$ 0".5. This must, however, be confirmed by future observations.

One of the most interesting discoveries in the 1665-MHz

emission of VY CMa is the weak features detected at 17 and 26 km/s.
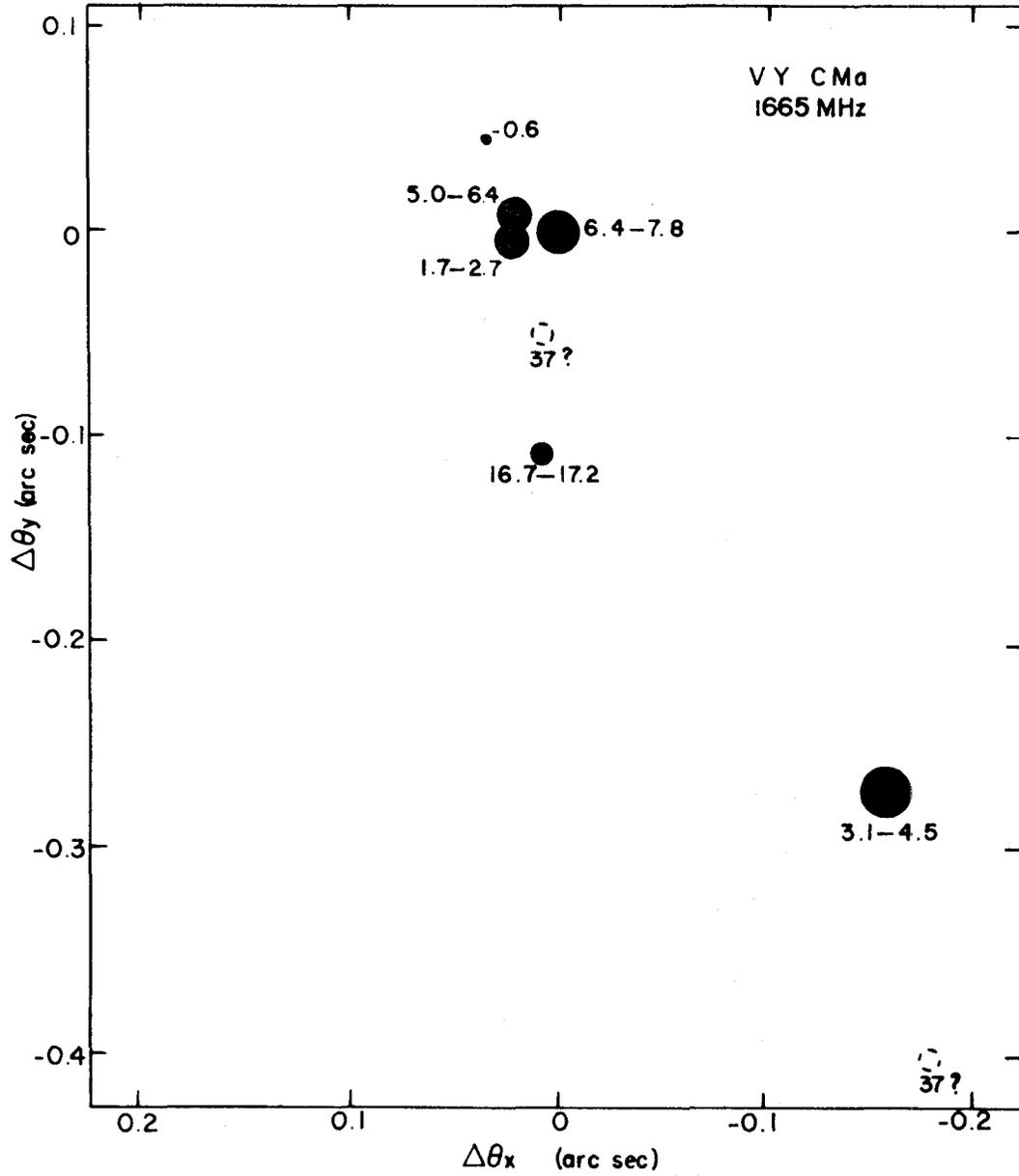
Figure 14.--Map of the 1665-MHz emission of VY CMa. Position offsets were determined by modeling
the phase variation in each spectral channel by a single point component. Uncertainties
of the relative positions are less than about 0.02 arc sec. The size of the circles is pro-
portional to the strength of the emission at that velocity.

Neither of these components can be associated with either the high
or low velocity emission complexes. Instead, they seem to form a
second set of emission complexes whose center is close to that of the
dominant emission complexes, but which has a very small velocity
separation. The 26 km/s component is too weak to map reliably.
However, the 17 km/s component position offset is 0."11 ± 0."02 south
of the reference component.

As in the case of the 1612-MHz emission of VY CMa, there is no
obvious spatial structure to the emission at 1665-MHz. In the low
velocity emission complex, maser components separated by 4 km/s are
spatially coincident while a component with a velocity intermediate
between these is offset by several tenths of an arc second. This
again argues for non-homogeneous models for the circumstellar envelopes.
Even more striking is the small offset of the 17 km/s component from
the reference position. This offset is less than the offset of the
4 km/s component which is in the same emission complex as the refer-
ence feature. The simple expanding shell model would require the
17 km/s component to be located near the limb of the shell and hence
far away from other components. It is impossible to explain simulta-
neously the proximity of the 17 km/s component and the relatively
large offset of the 4 km/s component from the reference feature with
a simple expanding shell model.

The "second set of emission complexes" at 17 and 26 km/s can be
explained by the existence of multiple circumstellar shells. This
emission could originatefrom a newly formed shell, closer to the
central star than the older shell which gives rise to the dominant

emission features. The newly formed shell may not yet have been

accelerated to the full expansion velocity of the older circumstellar

material, and thus, it has a velocity closer to the presumed stellar

velocity near 20 km/s (LSR). Further, the two (or more) shells of

circumstellar material ejected from the central star would be expected

to be concentric and the emission from each should occur from the

centers of the front and back of each shell. Thus, no

position offset would be expected between components in either shell.

The small position offset of the 17 km/s component from the reference

component is a natural consequence of the multiple shell model.

REFERENCES

Eliasson, B. and Bartlett, J.F. 1969, Ap.J. (Letters), 155, L79.

Harvey, P.J., Booth, R.S., Davies, R.D., Whittet, D.C.B., and

McLaughlin, W. 1974, M.N.R.A.S., 169, 545.

Herbig, G.H. 1970, Cont. Lick Obs., No. 302.

Högbom, J.A. 1974, Astron. Astrophys. Suppl., 15, 417.

Johnston, K.J., Knowles, S.H., Sullivan, W.T. III, Moran, J.M., Burke,

B.F., Lo, K.Y., Papa, D.C., Papadopoulos, G.D., Schwartz, P.R.,

Knight, C.A., Shapiro, I.I., and Welch, W.J. 1971, Ap.J. (Letters),

166, L21.

Knowles, S.H., Johnston, K.J., Moran, J.M., Burke, B.F., Lo, K.Y.,

and Papadopoulos, G.D. 1974, Astron. Journ., 79, 925.

Masheder, M.R.W., Booth, R.S., and Davies, R.D. 1974, M.N.E.A.S., 166,

561.

Moran, J.M., Burke, B.F., Barrett, A.H., Rogers, A.E.E., Ball, J.A.,

Carter, J.C., and Cudaback, D.D. 1968, Ap. J. (Letters), 152, L97.

Moran, J.M., Ball, J.A., Yen, J.L., Schwartz, P.R., Johnston, K.J.,

and Knowles S.H. 1975, in preparation.

Appendix

I. DATA ANALYSIS

The analysis of spectral line VLB data is exceedingly lengthy and complex. This stems from the tremendous quantity of data collected ($\sim 10^{12}$ bits per experiment) and the introduction of the spectral dimension. Essentially all of the problems encountered in spectrometry, enhanced by the complexities of cross-correlation analyses, are combined with the difficulties of "normal" VLBI operations. A successful spectral line VLB experiment consists of three stages: 1) the acquisition of raw data; 2) the reduction of this data to obtain estimates of the normalized complex fringe visibility as a function of U and V (the east and north projected interferometer spacings in wavelengths, respectively), and v (the velocity or Doppler shifted frequency); and 3) the transformation of the fringe visibilities into "maps" of brightness temperature as a function of position on the sky and velocity.

The operation of a microwave interferometer is discussed in many texts (e.g., Christiansen and Högbom, 1969), and the theory of spectral line VLB data analysis is discussed by Moran (1975). In this section, I will concentrate on the practical details of converting raw Mk II video tapes into the "finished product." Because of the complexity of spectral line VLB analysis, it is essentially useless to present a rigorous discussion of all of the factors which can enter into the observations and data reduction. In order to make this section complementary to existing discussions of some of these subjects, and a

useful guide to those who wish to conduct spectral line VLB experi-

ments, I will use many conceptual simplifications commonly employed

to approach the data.   Throughout this discussion examples will be drawn

from the most simple, realistic type of source which consists of

spatially and spectrally isolated emission line sources with no con-

tinuum emission (e.g., stellar masers).   More complex cases, which

can include a mixture of structured absorption lines and continua

(e.g., radio absorption line quasars) will be discussed by Romney

(in preparation).

II. DATA ACQUISITION

Figure A.1 is a block diagram which traces the signal of a radio source through a typical Mk II VLB terminal. Figure A.2 then traces the recorded signals through the Mk II VLB processor. In these diagrams, all of the frequency dependent terms have been retained in order to bring out the important ones for spectral line observations. There are several simplifications which have been made without loss of generality. First, the source is assumed to be a monochromatic, point source; any arbitrary source brightness distribution can, in principal, be constructed by a superposition of such signals. Secondly, ideal electronic operation is assumed; most electronic phase terms can be lumped in with the oscillator phase, β. Most frequency dependent amplitude terms can be treated as part of the fringe amplitude, $\langle A_1(\omega)A_2(\omega)\rangle$. Finally, the analysis is performed for a complete analog case; the effects of clipping the recorded signals and of the digital (three level) fringe rotator in the processor are discussed later.

The phase difference of the signal incident upon the two telescopes is represented by $\omega\tau$ where ω is the frequency of the source at telescope 1, and τ is the time difference between the arrival of the wavefront at telescopes 1 and 2. τ can be thought of as the geometric delay, given the source position vector, $\hat{S}$, and the baseline vector, $\vec{B}$, defined at the instant the wavefront arrives at telescope 1, plus a correction due to the change in the baseline during the transit of the signal from telescope 1 to 2.

$$\tau = \tau_g + \tau_{atm} + \tau_{ion} + \dots$$

$$\tau_g \equiv \frac{1}{c}\,\vec{B}\cdot\hat{S}\left(1 + \frac{\frac{d\vec{B}}{dt}\cdot\hat{S}}{c}\right)$$

$A(\omega)\cos\left[\omega t + \zeta(\omega)\right]$

$A_1(\omega)\cos\left[\omega t + \zeta(\omega)\right]$

$A_2(\omega)\cos\left[\omega(t-\tau) + \zeta(\omega)\right]$

$\cos\left[\omega_1 t + \beta_1(t)\right]$

$\cos\left[\omega_2 t + \beta_2(t)\right]$

$\frac{1}{2}A_1\cos\left[(\omega \mp \omega_1) t + \zeta \mp \beta_1\right]$

$\frac{1}{2}A_2\cos\left[(\omega \mp \omega_2) t - \omega\tau + \zeta \mp \beta_2\right]$

$\frac{1}{2}A_1\cos\left[(\omega - \omega_1) t + \zeta - \beta_1\right]$

$\frac{1}{2}A_2\cos\left[(\omega - \omega_2) t - \omega\tau + \zeta - \beta_2\right]$

$t_2 = t_1 + \Delta t_c$

$\frac{1}{2}A_1\cos\left[(\omega - \omega_1) t_1 + \zeta - \beta_1\right]$

$\frac{1}{2}A_2\cos\left[(\omega - \omega_2)(t_1 + \Delta t_c) - \omega\tau + \zeta - \beta_2\right]$

LO$_1$, LO$_2$, LPF, freq. standard$_1$, freq. standard$_2$, Clock$_1$, Clock$_2$, Video tape$_1$, Video tape$_2$

Mk II VLB terminal #1

Mk II VLB terminal #2
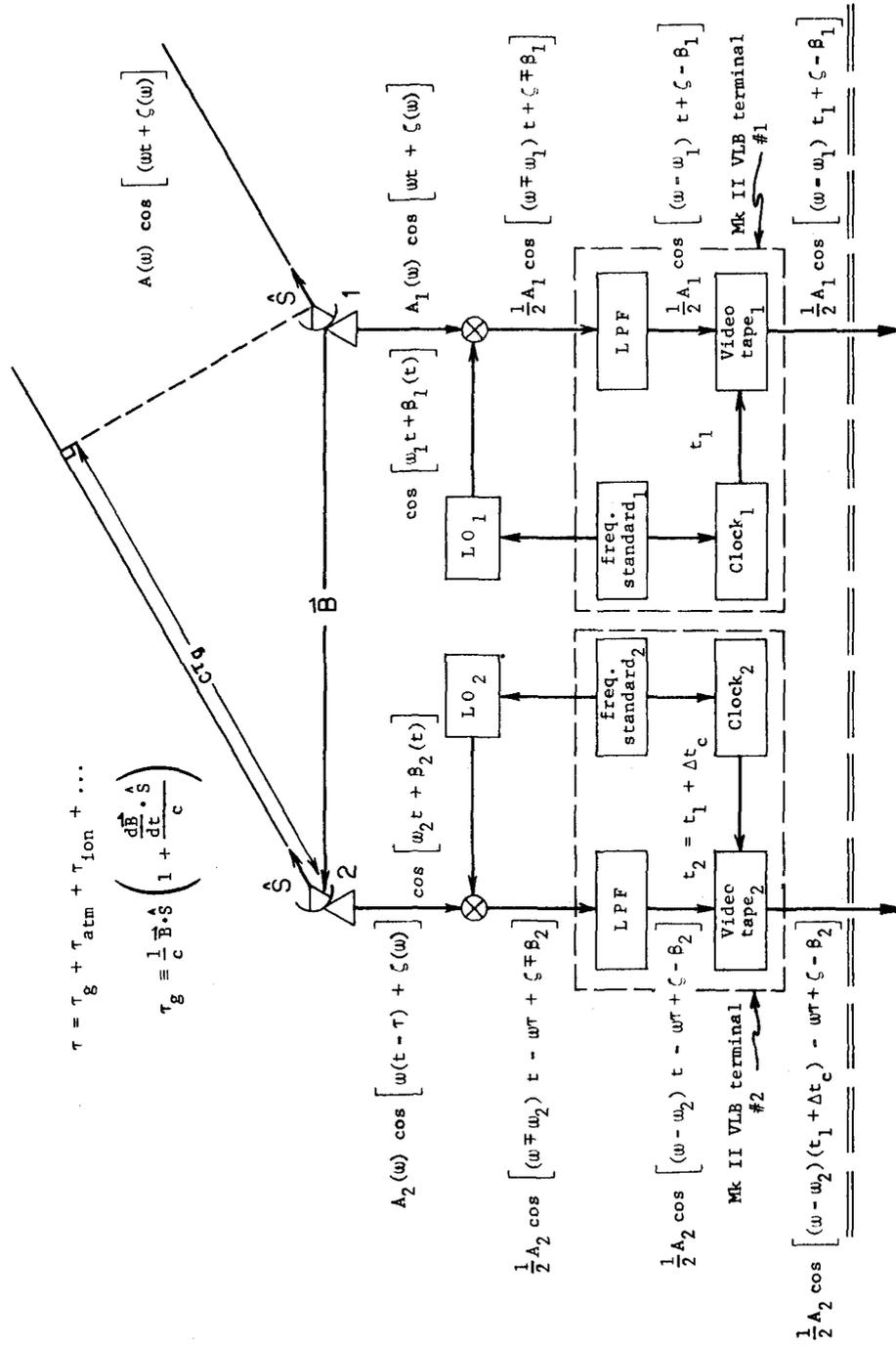
Figure A.1 - Schematic representation of the Mk II VLB data acquisition. A monochromatic signal, $A(\omega)\cos[\omega t + \zeta]$, from a radio source, $\hat{S}$, is detected by an interferometer with an instantaneous baseline vector, $\vec{B}$. The time delay, τ, between the arrival of the signal at telescopes 1 and 2 is given by the sum of a geometric (τ_g), atmospheric (τ_{atm}), ionospheric (τ_{ion}), and possibly other terms. The detected signal at the first telescope is mixed with a local oscillator (LO$_1$) of frequency ω_1 and phase β_1; passed through a low pass filter (LPF$_1$); and recorded on video tape. A time difference of Δt_c is assumed between clocks 1 and 2. A shorthand notation is used for the output of each mixer ($\otimes$) in which $\cos(y) \otimes \cos(z) = \frac{1}{2}[\cos(y+z) + \cos(y-z)]$ is represented as $\frac{1}{2}\cos(y\pm z)$.

111

Mk II VLB Processor

sine fringe rotator

$\cos\left[\omega^* t_1 + \delta - \frac{\pi}{2}\right]$

cosine fringe rotator

$\cos\left[\omega^* t_1 + \delta\right]$

j'th digital delay $t_2 \to t_2 - \tau_j$

$\frac{1}{4}A_1\cos\left[(\omega-\omega_1)t_1 \mp \omega^* t_1 + \zeta - \beta_1 \mp \delta\right]$

$\frac{1}{4}A_1\cos\left[(\omega-\omega_1)t_1 \mp \omega^* t_1 + \zeta - \beta_1 \mp \delta \pm \frac{\pi}{2}\right]$

$\frac{1}{16}A_1A_2\cos\left[(\omega-\omega_2)(t_1+\Delta t_c+\tau_j) - \omega\tau + \zeta + \delta\right.$
$\left. -\beta_2 \mp (\omega-\omega_1)t_1 \pm \omega^* t_1 \mp \zeta \pm \beta_1 \mp \frac{\pi}{2}\right]$

$S_j = \frac{1}{16}\langle A_1(\omega)A_2(\omega)\rangle \sin\left[(\omega-\omega_2)(t_1+\Delta t_c+\tau_j) - \omega\tau + (\beta_1-\beta_2) - \omega^* t_1 - (\omega-\omega_1)t_1 + \delta\right]$

0.2 integ.

j'th **sine channel**

$\frac{1}{16}A_1A_2\cos\left[(\omega-\omega_2)(t_1+\Delta t_c+\tau_j) - \omega\tau + \zeta + \delta\right.$
$\left. -\beta_2 \mp (\omega-\omega_1)t_1 \mp \omega^* t_1 \mp \zeta \pm \beta_1\right]$

$C_j = \frac{1}{16}\langle A_1(\omega)A_2(\omega)\rangle \cos\left[\cdots\right]$

0.2 integ.

j'th **cosine channel**

computer tape
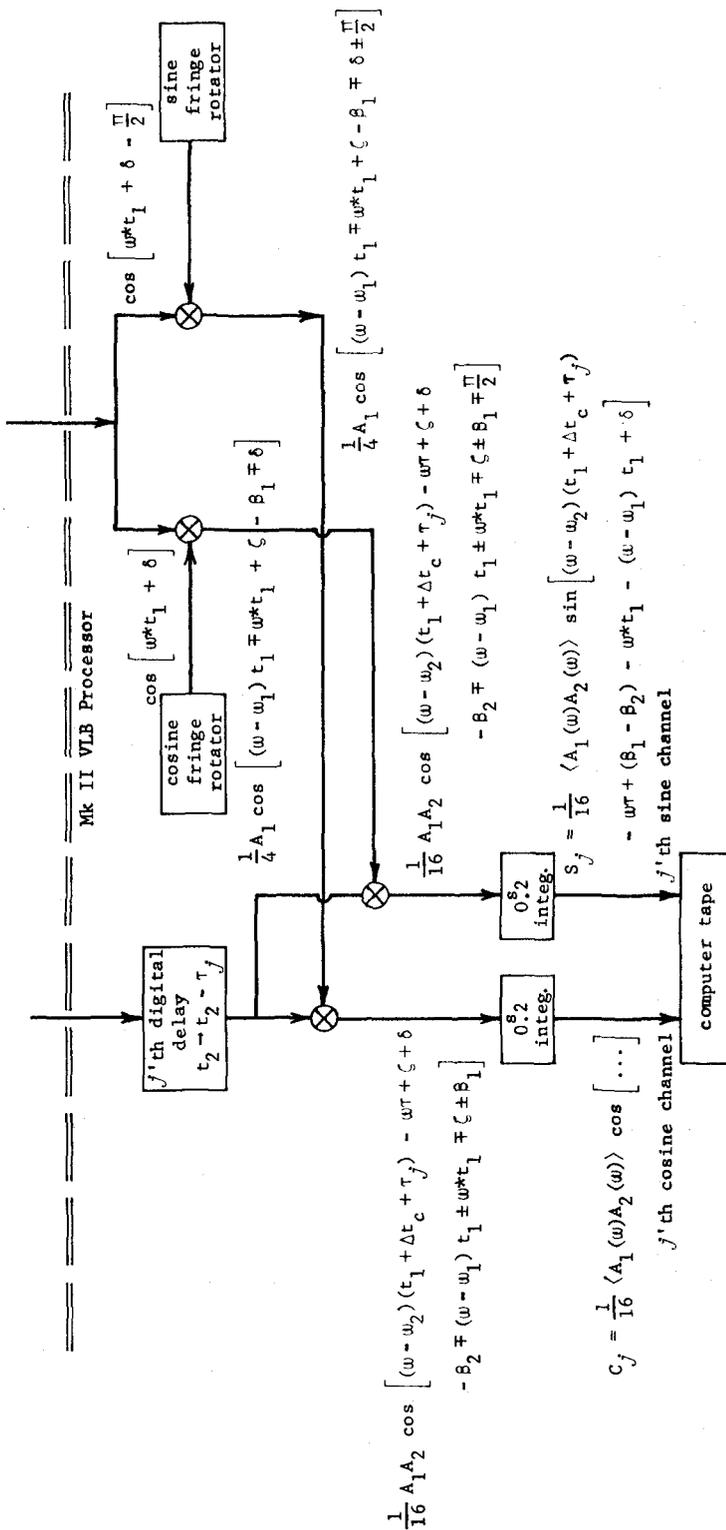
Figure A.2 - Schematic representation of the Mk II VLB data processor. The data recorded on the first video tape are mixed with two sinusodial fringe rotators each at a frequency ω* and a phase δ; an additional phase of π/2 is inserted in the "sine" fringe rotator. The data recorded on the second video tape are passed through a series of M digital delay channels and then each channel is mixed with the output of the two fringe rotators. This produces an M channel complex cross-correlation function. Only the j'th delay channel is shown in this figure. The data in each channel is integrated for 0.2 seconds and placed on computer tape.

The output of the Mk II VLB processor presently consists of up to 288 complex cross-correlation channels, or up to 576 auto-correlation channels. Both cross- and auto-correlation are used in spectral line experiments. The cross-correlation channels are used to obtain an interferometer (cross-correlated) spectrum. The auto-correlation channels are used to determine a "mean" total power spectrum of the telescopes which is used to calibrate the cross-correlated spectrum. The normalized fringe visibility spectrum for each spectral channel is the cross-correlated spectrum (multiplied by a correction factor) divided by the mean total power spectrum. Thus, the spectral fringe visibility can be obtained without measurement of receiver or source temperature. Details of this calibration procedure are discussed later.

The real and imaginary (cosine and sine) parts of the cross-correlation function sampled by the j'th delay channel represented by C_j and S_j, respectively, are the (unnormalized) fringe amplitude times a cosine and sine of the residual fringe phase, Φ. If we make several substitutions, Φ can be expressed in a more simple and useful form. The j'th channel time delay, τ_j, can be represented as

$$\tau_j = \tau + \Delta\tau_j \quad , \tag{A-1}$$

where $\Delta\tau_j$ is the difference between τ_j and the true delay, τ. Define

$$\Delta\beta(t) = \beta_2(t) - \beta_1(t) \tag{A-2}$$

which represents the phase difference between the two local oscillators. Finally, the fringe rotator frequency is set to

$$\omega^* = (\omega_1 - \omega_2) - \omega_2 \frac{d\tau}{dt} + \omega_{res} \qquad \text{(A-3)}$$

where the residual fringe frequency, ω_{res}, represents the difference between fringe rotation frequency determined with perfect values of ω_1, ω_2, and $\tau(t)$ and the value derived from the actual (imperfect) values adopted for these parameters. The residual fringe frequency, ω_{res}, is used to determine physical parameters such as baseline and source position corrections.

Employing eqs. (A-1) to (A-3), the spectral interferometer phase can be written as

$$\Phi_i = \omega_{res} t_1 - \Delta\beta + (\omega - \omega_2) \Delta t_c + (\omega - \omega_2) \Delta\tau_i \qquad \text{(A-4)}$$

This formulation of the residual fringe phase is used since the four terms in eq. (A-4) are usually treated separately in the data analysis.

The first term of eq. (A-4) is the "geometrical" part of the residual fringe phase. It is non-zero if there are errors in the model in calculating τ_g in the VLB processor. $\tau_g(t)$ is usually in error because of inaccurate baselines, source positions, and UT1 (i.e., t_1) values. Other modeling errors, such as ignoring relativistic effects or atmospheric and ionospheric propagation effects, can be thought of as contributing to the "geometric" residual fringe phase, $\omega_{res} t_1$, or treated as noise (and included in $\Delta\beta(t)$) depending upon the nature of a particular experiment. Source structure can be determined by analyzing the time variations of $\omega_{res} t_1$ after the contributions to the fringe phase of the other terms in eq. (A-4) have been accounted for. A good example of this is the map of the spatial

distribution of many spectrally isolated components in U Ori given in Chapter II (Fig. 6).

The second term in eq. (A-4) is the "noise" in the residual fringe phase. Although it has been schematically represented in Fig. A.1 as solely due to local oscillator instabilities, it can represent other forms of unmodeled time dependent phase fluctuations such as can be caused by electronic time delay drifts in the receivers, mixers, and filters. In general, the time over which $\Delta\beta$ varies by ≈ 2 radians is the coherence time, T_c. Integrating the cross-correlation function $(C_j + iS_j)$ over a time, T_c, decreases the amplitude to one-half its original value.

The third term in eq. (A-4) describes the effects of clock differences, Δt_c, which are not accounted for in the processor delay offset (i.e., the CLOCK parameter in the REDPREP program). A small error of Δt_c (e.g., $\sim 1\mu s$) will not be noticeable in the geometric part of the residual fringe phase, $\omega_{res} t_i$; but will cause the peak of the cross-correlation function to shift several delay channels from the center channel in the processor. When a small clock error is examined in the frequency domain (by Fourier transforming the cross-correlation function), it results in a phase slope across the video bandpass given by $(\omega - \omega_2)\Delta t_c$. This effect can seriously affect the analysis of the spatial structure of spectral sources. For example, if the relative positions of two spectral components (within the same bandpass) are determined from the phase difference between the two components, the contribution of the phase difference from a clock

error (usually a constant) must be known. Since typical VLB clocks drift by less than $\sim 1\mu s/day$, the clock offsets can be calibrated out by observing strong continuum sources with well known positions throughout the experiment. If this calibration of the clock offsets is not done with high accuracy (e.g., $\sim 1/10B$, where B is the observing bandwidth), the relative positions of spectral components can be systematically in error, since phase differences due to clock errors can mimic (or correlate with) certain combinations of right ascension and declination offsets.

The fourth term in eq. (A-4) is used to describe the effects of the digital delay tracking in the VLB processor. The third and fourth terms in eq. (A-4) have the same functional form, a video frequency $(\omega - \omega_2)$ multiplied by a small time "error," and could have been combined into one term. They are separated because the effects of slowly drifting clock offsets, Δt_c, and the rapidly changing digital delay tracking error, $\Delta \tau_i$, are treated differently. Even if the baseline, source positions, and clock parameters are perfectly known, the delay of the center channel of the VLB processor will oscillate about the ideal continuous function, $\tau(t)$, by $\pm 1/8$ μs as it approximates $\tau(t)$ in discrete $1/4$ μs steps. The delay steps typically occur on time scales of from 0.05 to 10 seconds. If these jumps occur more rapidly than the fundamental integration time $(0\overset{s}{.}2)$ of the VLB processor, the phase errors given by $(\omega - \omega_2)\Delta \tau_i$ are averaged and the signal is partially decorrelated. In many spectral line observations, narrow bandwidths are used so that $\Delta \tau_i = 1/8$ μs

does not cause an appreciable phase error $(\omega - \omega_2)\Delta\tau_i$, and this effect can often be ignored. However, if spectral lines appear with video frequencies $\omega - \omega_2 \gtrsim 1$ MHz, sizable phase errors can result.

III. DATA REDUCTION

The MkII VLB processor produces complex cross-correlation coefficients as a function of time and cross-correlation delay. Thus, the data can be graphically represented in three cartesian coordinates where the x-axis is time, the y-axis is cross-correlation delay, and the z-axis is the real or imaginary part of the fringe visibility (Fig. A.3a). The time sequence of data can be Fourier transformed into fringe frequency estimates as is often done in a continuum analysis (Fig. A.3b); the cross-correlation function can be Fourier transformed into a cross-correlated spectrum (Fig. A.3c); or both transformations can be done (Fig. A.3d). The Fourier transforms are linear and the time and cross-correlation delay axes are orthogonal. Thus, the order in which the transforms are done is unimportant.

Fringe visibilities can be estimated in any of the four "spaces" shown in Fig. A.3. However, the maximum signal to noise on these estimates is obtained in the space in which all of the signal power is condensed into the narrowest region. For example, in continuum observations, the cross-correlation delay function is more sharply peaked than the spectrum. Thus, continuum fringe visibilities are estimated in the fringe frequency-cross-correlation space (Fig. A.3b). In spectral observations, the cross-correlation delay function of a spectral line is usually very broad, and cross-correlated spectrum is sharply peaked at the line frequency. Thus, in spectral analyses, the fringe visibilities are usually estimated in the fringe frequency-spectral frequency space (Fig. A.3d). The computer programs which
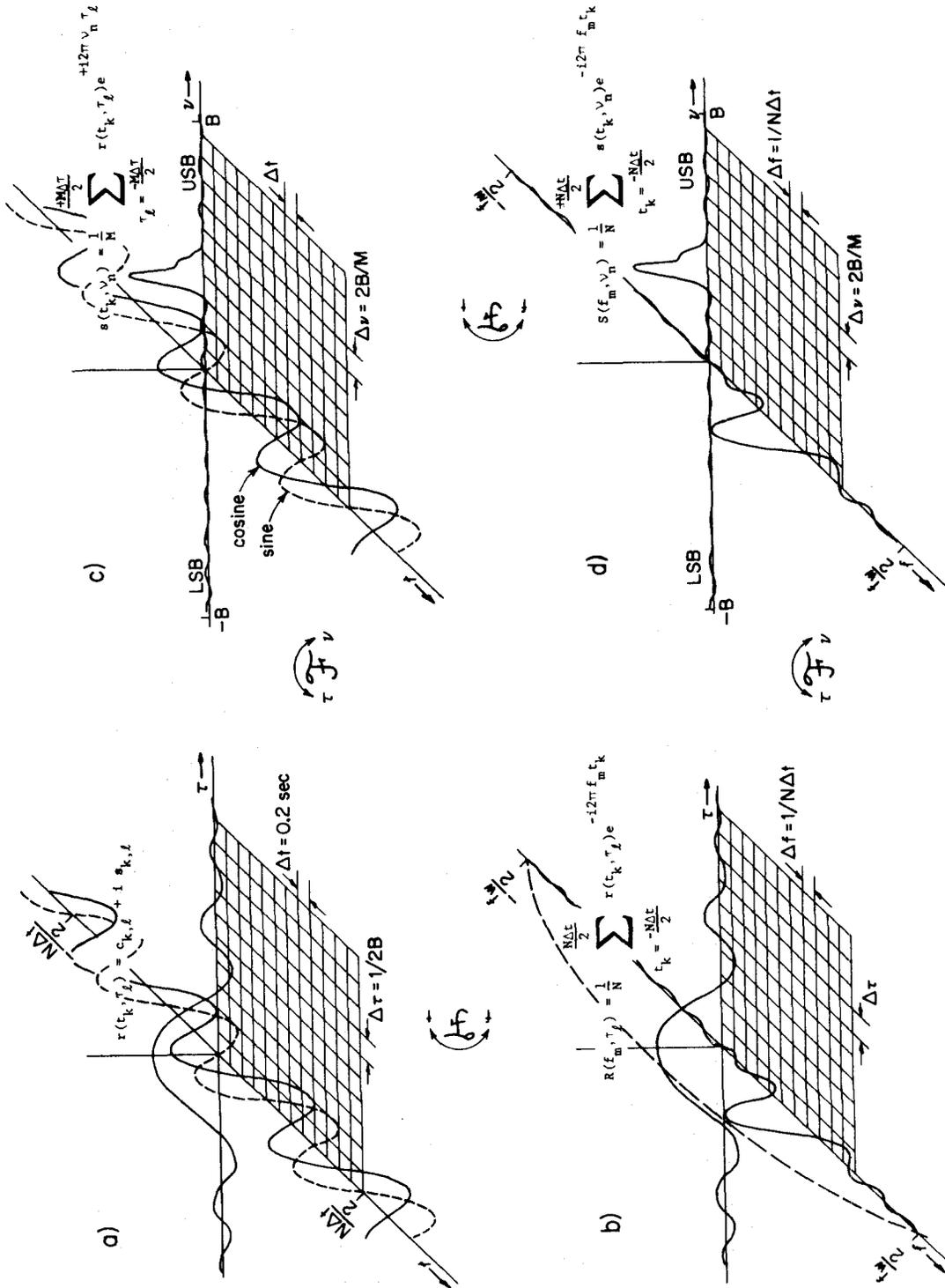
Figure A.3 - Schematic representation of the four "spaces" in which VLB data can be analyzed. The time sequence (t-axis) of the cross-correlation functions (τ-axis) stored on computer tape in the VLB processor (see Fig. A.2) are illustrated in part (a). The time sequence of N data points can be transformed into fringe frequency (f-axis) extimates and/or the M point cross-correlation function can be transformed into spectral frequency (ν-axis) estimates as illustrated in parts (b), (c), and (d).

carry out these Fourier transforms and estimate spectral fringe visibilities will now be described.

In order to determine the spatial structure of a spectral line source, the Mk II processor output tapes (Varian tapes) are analyzed by a series of computer programs designed to compress and calibrate the data. The first program (DECODE) examines the processor data for problems in decoding the data and time codes on the raw video tapes. This program rewrites a new tape which is a condensed version (by increasing the tape density and data blocking factor) of the Varian tape with any questionable data "flagged."

The second reduction program (SPECTRUM) edits our questionable data and separates auto-correlation from cross-correlation data for subsequent analysis. The auto-correlated data is accumulated for output at the end of each observing scan. The cross-correlated data is handled as follows. First, a 0.2 second set of cross-correlation coefficients

$$r(t_k, t_\ell) = C(t_k, t_\ell) + iS(t_k, t_\ell) \qquad (A-5)$$

may have its phase ($\tan^{-1}(S/C)$) changed or "rotated" to remove a small residual fringe rate, $\dot{\Psi}$, (e.g., ≈ 0.2 Hz) purposely left in the data at processing time to separate actual fringes from spurious d.c. (zero frequency) fringes often introduced by recorder or processor hardware problems. A phase, Ψ, is removed from the data as follows:

$$C \longleftarrow C \cos \Psi - S \sin \Psi$$

$$S \longleftarrow C \sin \Psi + S \cos \Psi$$

The data are then averaged for several seconds, in order to reduce the number of spectral Fourier transforms required. This averaging is coherent integration (or vector averaging), because the cosine and sine channels are summed independently. Averaging the data for a time, T, reduces (decorrelates) the fringe amplitude by a factor of $\sin(\pi \Delta fT)/(\pi \Delta fT)$, where Δf is the fringe rate remaining in the data after the initial phase rotation. The averaging time, T, should be as long as possible to reduce subsequent computer time requirements, but short enough so as not to decorrelate the data. If, for example, residual fringe rates could be confidently reduced to less than .02 Hz after removal of the intentional 0.2 Hz processing offset, then the data could be averaged to 5 seconds with less than 1% decorrelation of the fringe signal and any hardware d.c. biases would be 100% "decorrelated" and thus entirely removed from the data.

The averaged cross-correlated data are then Fourier transformed (spectral transform) to yield a cross-correlated spectrum. If the time between the 1/4 μs delay jumps in the processor is greater than the processor integration time ($0.^{s}2$), the spectral transform can be done prior to time averaging and the frequency dependent "digital delay tracking" phase term $(\omega - \omega_2) \Delta \tau$ can be removed from the spectrum. Finally, a cross-correlated spectrum is written on an output tape (SPECTRUM tape) for every averaged set of data.

The auto-correlated data for each telescope (i.e., video tape) are accumulated for the entire scan (typically $\sim$ 15 min.). The data are corrected for the effects of clipping by the Van Vleck relation:

$$\rho(\tau) = \sin\left(\frac{\pi}{2} \rho_c(\tau)\right) \qquad \text{(A-6)}$$

where ρ_c is the clipped auto-correlation coefficient, and ρ the

corrected value which would be obtained in the absence of clipping.

The corrected auto-correlation coefficients are Fourier transformed

to yield total power (single dish) spectra from each telescope. These

spectra are stored either on cards or disks for later use.

The subsequent analyses of the cross-correlated spectra are

designed to yield fringe amplitude, residual fringe rate, and residual

fringe phase estimates. Two procedures can be used to obtain these

estimates. The fringe amplitude, rate, and phase can be estimated

either in the fringe frequency-spectral frequency plane (see Fig. A.3d)

by Fourier transforming the time sequence of data (from the SPECTRUM

tape) into fringe frequency estimates, or they can be estimated in

the time-spectral frequency plane (see Fig. A.3c) by a least-squares

fitting of sinusoids to the time sequence of data.

The fringe frequency Fourier transform analysis of cross-corre-

lated spectra is done by program FRINGE. One major function of

FRINGE is to search for fringes which may be weak and have unknown

residual fringe rates. The major difficulty encountered in most

reduction programs is the limit upon the number of data points which

can be transformed into fringe frequency channels. For weak signals,

it may be necessary to coherently integrate up to the coherence

time of the interferometer to maximize the signal to noise ratio.

Integrating to the coherence time, T_c, by Fourier transforming in the

time domain requires preaveraging the data to T_c/N seconds if N points

are transformed. This requires the residual fringe rate to be within

a "window" of $\pm \frac{1}{2}[1/(T_c/N)]$ Hz in order to avoid serious decorrelation

during preaveraging. If coherence times are long and/or if source

positions, baselines, and oscillator frequencies are not known well

enough to keep the residual fringe rates low, the number of points

transformed can become quite large (e.g., several hundred). The only

limitation in fringe searching (beyond core storage and computation

time) is the $\pm$ 2.5 Hz fringe rate "window" imposed by the 0.2 second

averaging in the MkII processor.

In the normal mode of operation, FRINGE is designed to yield

estimates of fringe amplitude, and the residual fringe rate and phase

in the fringe frequency-spectral frequency plane. First, the data

are averaged for a time period, T. At this point the fringe amplitude,

$A(\nu_n)$, and residual fringe phase, $\Phi(\nu_n)$, could be directly estimated

(in the time-spectral frequency plane) prior to performing the fringe

frequency transform as follows:

$$A(\nu_n) = \sqrt{C(\nu_n)^2 + S(\nu_n)^2} \qquad \text{(A-7)}$$

$$\phi(\nu_n) = \arctan [S(\nu_n)/C(\nu_n)] \qquad \text{(A-8)}$$

where $C(\nu_n)$ and $S(\nu_n)$ are the n'th cosine and sine channels associated

with spectral frequency ν_n. However, the residual fringe rate, f,

must be very close to zero to avoid significant decorrelation of the

signal. The decorrelation produced by time averaging the data is

given by a factor $\sin(\pi f T)/(\pi f T)$. If the residual fringe frequency,

f, can be reduced such that for a coherence time, T_c

$$f \lesssim 1/4(1/T_c) \qquad , \qquad (A-9)$$

the averaging time can approach the coherence time which optimizes

the signal to noise ratio. However, in most cases, very accurate

source positions and baselines are needed to reduce sufficiently the

residual fringe rates.

A generally more convenient analysis procedure involves averaging

the data over a short period of time, $T \ll T_c$, and Fourier transforming

many averaged groups. In this case, a less stringent requirement than

equation (A-3) is placed upon the magnitude of the residual fringe

rate:

$$f \lesssim 1/4(1/T),$$

in order to avoid serious decorrelation during the preaveraging.

Fourier transforming the data yields estimates of the real and imagi-

nary parts of the complex fringe visibility at many fringe frequencies.

The fringe frequency Fourier transform can be thought of rigorously as

attempting to "stop the fringes" by rotating the phase at many "trial"

fringe rates and averaging the data over the total transform time.

The response of the Fourier transformation of N (averaged) data points

is

$$\begin{Bmatrix} C(\nu_n) \\ \\ S(\nu_n) \end{Bmatrix} = \frac{\sin[\pi(f-f_m)NT]}{\pi(f-f_m)NT} \, A(\nu_n) \begin{Bmatrix} \cos \phi \\ \\ \sin \phi \end{Bmatrix} \qquad (A-10)$$

in the m'th fringe frequency channel associated with fringe frequency,

f_m. If $f_m \approx f$, the sin x/x factor in equation (A-10) is close to one

and the estimates of amplitude and phase can be obtained from equations
(A-7) and (A-8). If $f_m \not= f$, the sin X/X factor diminishes toward
zero so the residual fringe rate can be estimated easily. It is
important to note that if the transform time interval, NT, exceeds
the coherence time, T_c, the residual fringe frequency is not constant
during the "integration." This will cause a spreading of the fringe
amplitude over several fringe frequency channels, and the estimates
of A and Φ from equations (A-7) and (A-8) will be systematically low.

In some cases where the coherence time is not well known and
sources are weak, it may be advisable to have the transform time
interval exceed the coherence time (as in the Mk II continuum reduc-
tion program BCA). In this case, the fringe power is spread over
several fringe frequency channels, and equations (A-7) and (A-8) can-
not be directly employed to estimate fringe amplitude and phase.

An alternative method of estimating the fringe amplitude and
residual fringe rate and phase is to do a non-linear least-squares
fit to the data in the time-spectral frequency plane. The data in
the cosine and sine channels are modeled in program FIT as

and
$$A \cos[2\pi f(t - \bar{t}) + \phi]$$
$$A \sin[2\pi f(t - \bar{t}) + \phi]$$
(A-11)

respectively, in each spectral channel where $\bar{t}$ is the weighted mean
time of the entire set of data points which is to be analyzed.
In this case, the data should be preaveraged as in the Fourier trans-
form method, and the parameters A, f, and Φ estimated by minimizing
the sum of the squares of the differences between the data and the
model. It is important that the time sequence of data in each spectral

channel is modeled as in equation (A-11), so that the estimates of the parameters A, f, and Φ are entirely uncorrelated. If, for example, $\bar{t}$ is chosen to be the beginning (or end) of the data set, then estimates of the phase, Φ, are highly correlated to estimates of the fringe rate, f. This doubles the noise (or uncertainty) of the phase estimates.

Analysis of the cross-correlated data by the fringe frequency transform technique (FRINGE) and the least-squares technique (FIT) leads to the same parameter estimates if the signal to noise ratio is high. Each technique has its own advantages and disadvantages. The major advantage of the fringe frequency analysis is that the Fourier transform is linear, and large ranges of fringe frequencies can be searched with moderate amounts of computer time because of the efficiency of fast Fourier transform algorithms. Its main disadvantages are that accurate parameter estimates require interpolation between fringe frequency channels, and fringe amplitude estimates are biased toward larger values and require a complex non-linear correction which is a function of the signal-to-noise ratio (Purcell, 1973). The major advantages of the least-squares technique is that it directly yields unbiased parameter estimates. Formal uncertainties are of great value if the subsequent data analysis involves model fitting, since the data can be properly weighted. The main disadvantage of the least-squares analysis is that it is a non-linear process; therefore, good initial estimates of all parameters are required.

The auto-correlation spectra from each telescope are used to calibrate the cross-correlation amplitudes estimated in FRINGE or FIT. The fringe visibility amplitude, γ, as a function of spectral frequency, ν_n, (cf. Moran, 1974) is given by

$$\gamma(\nu_n) = \frac{b \, S_{12}(\nu_n)}{\sqrt{S_1^{on-off}(\nu_n) \, S_2^{on-off}(\nu_n)}} \qquad \text{(A-12)}$$

where S_{12} is the cross-correlation fringe amplitude, S_j^{on-off} is auto-correlation spectrum obtained by differencing on (S_j^{on}) and off (S_j^{off}) source spectra for the jth telescope, and b is a correction factor for the cross-correlated data (cf. Cohen et al., 1975).

Equation (A-12) is not of great practical value for two reasons. First, a correction must be applied to the auto-correlation spectra, since they are calculated from a normalized auto-correlation function $[\rho(\tau)]$. The total power integrated over either an on or off source spectrum is, by Parseval's Theorem,

$$\int_{Bandpass} S_j^{on \ or \ off}(\nu) \, d\nu = \rho_j(0) \equiv 1$$

This causes the baseline (or background noise) level to be lower (for an emission line source) in an on-source spectrum than in an off-source spectrum. This effect cannot easily be removed from the differenced spectra, $S_j^{on-off}(\nu_n) = S_j^{on}(\nu_n) - S_j^{off}(\nu_n)$, since, in general, its baseline level varies as the observing filter's frequency response. The "differenced-ratioed" spectrum,

$$\frac{\overline{\text{on-off}}}{S_j^{\text{off}}(\nu_n)} \equiv \frac{S_j^{\text{on-off}}(\nu_n)}{S_j^{\text{off}}(\nu_n)} \qquad ,$$

however, has a flat baseline and can be corrected for a non-zero baseline level.

The second reason equation (A-12) is not directly employed is that the signal to noise ratio may be very low on the auto-correlation spectra for one or both telescopes. Thus, a direct application of equation (A-12) may lead to very noisy estimates of $\gamma(\nu_n)$. However, the auto-correlation spectra required to calibrate the cross-correlation fringe amplitudes can be represented as follows:

$$\frac{\overline{\text{on-off}}}{S_j^{\text{off}}(\nu_n)} = Y_j(t)S^{\text{total}}(\nu_n) \qquad (A-13)$$

where $S^{\text{total}}(\nu_n)$ is the source total power spectrum and $Y_j(t)$ a time varying, but frequency independent, scale factor which reflects the receiver temperature and gain of the jth telescope. $S^{\text{total}}(\nu_k)$ is a fundamental property of the source and does not usually vary over the course of an experiment. It can be determined by summing all of the "differenced-ratioed" spectra obtained from each telescope for the entire experiment. The scale factor, $Y_j(t)$, can then be determined for each on-off scan pair by fitting $S^{\text{total}}(\nu_n)$ to $S_j^{\overline{\text{on-off}}/\text{off}}(\nu_n)$. This effectively uses the entire spectrum to obtain an estimate of the telescope sensitivity, and thus using $Y_j(t) S^{\text{total}}(\nu_n)$ instead of $S_j^{\overline{\text{on-off}}/\text{off}}(\nu_n)$ in the calibration procedure usually greatly increases its accuracy.

Equation (A-12) can be rewritten in a more useful form as

$$\gamma(\nu_n) = \frac{\dfrac{b\, S_{12}(\nu_n)}{\sqrt{Y_1 Y_2 S_1^{off}(\nu_n) S_2^{off}(\nu_n)}}}{S^{total}(\nu_n)} \quad .$$ (A-14)

Program DIFSPEC is designed to determine $S^{total}(\nu_n)$ and the scale

factors Y_1 and Y_2. On- and off-source scan pairs for each telescope

are used to compute $S_j \frac{on-off}{off} (\nu_n)$ with a constant baseline offset removed.

All of these spectra are averaged together to define $S^{total}(\nu_k)$.

$S^{total}(\nu_k)$ is then re-entered into the program and the scale factors

Y_j obtained for each scan by least-squares fitting.

The final parameter in the fringe visibility calibration to be

discussed is the cross-correlation correction factor, b. This factor

takes into account several effects which reduce the cross-correlation

amplitude, but which may not reduce the auto-correlation amplitudes.

The two most important contributions to b result from clipping the

video data and employing a digital fringe rotator in the processor.

The correction factor for clipping losses obtained from the low ampli-

tude limit of the Van Vleck relation (eq. A-6) is $\pi/2$.

The correction factor for the three level fringe rotator can be

understood as follows: Assume that the original signal (e.g., tele-

scope 1 video signal of Fig. A.2) is given by $A\cos(\omega t)$. The digital

(cosine channel) fringe rotator, $F(\omega^* t)$, is a "three level" approxima-

tion to $\cos(\omega^* t)$ (Clark et al., 1972). The ori-

ginal signal and the digital fringe rotator are multiplied together

and integrated for a time T (i.e., $0^s.2$). Thus, the fringe rotator

output, A', can be approximated by

$$A' = \int_{-T/2}^{T/2} A\cos(\omega t) F(\omega^* t) dt \qquad . \qquad \text{(A-15)}$$

The phase of the incoming signal (with respect to that of the fringe rotator) is assumed to be zero. In reality the phase of the incoming signal is handled by analyzing the sine as well as the cosine fringe rotator output. The incoming signal is assumed to be analog, since the effects of clipping the data should not affect the correction for the digital fringe rotator (except for exceedingly strong sources where the antenna temperature is much larger than the system temperature).

The integral in equation (A-15) can be easily evaluated if

$$\omega \approx \omega^* \gg 2/T$$

and

$$|\omega - \omega^*| \ll 2/T \qquad .$$

This is almost always true in VLB observations where natural fringe rates are $\sim 10^3$ Hz and residual fringe rates $(\omega - \omega^*)$ are $\sim 10^{-2}$ Hz. The integral in eq. (A-15) can be evaluated if $F(\omega^* t)$ is expressed as a piecewise continuous function of +1's, 0's and -1's, or decomposed into its Fourier components. A Fourier series representation of $F(\omega^* t)$ has a dominant term of $1.18\cos(\omega^* t)$ and low amplitude terms at different frequencies. Only the dominant term substantially contributes to the integral in equation (A-15) which is then easily evaluated:

$$A' = \frac{1.18}{2} AT \qquad \text{(A-16)}$$

The rotated data of the Mk II processor is divided by the "number of bits correlated" which, in this integral case, is the time over which the fringe rotator is non-zero or 3T/4. Thus, the telescope 1 signal, and hence the cross-correlation amplitudes, are reduced by a factor of (1.18/2)/(3/4) = 0.78 from the amplitude expected for an ideal fringe rotator. Thus, the contribution to the cross-correlation amplitude correction factor, b, from the three-level fringe rotator is 1/0.78 = 1.28. If the only significant losses in the cross-correlation procedures are from clipping and fringe rotation,

$$b = 1.28\frac{\pi}{2} = 2.01 \qquad .$$

131

IV. DATA INTERPRETATION

Source structure can be inferred from fringe visibility data in
many ways. The most direct way is to Fourier invert the fringe visi-
bilities as a function of U and V into a spatial map of brightness.
This process requires calibrated fringe phases and a "reasonable"
sampling of the U-V plane. Some spectral-line sources have a spectral
component which can serve as a phase reference. Such a component
must have a residual fringe phase which is only due to an oscillator
instability and a position offset (from the nominal source position
used in processing). In this case, fringe phases of other spectral
components can be calibrated by subtracting the reference component's
phase from each spectral component's phase. Thus, in certain cases,
a direct Fourier inversion of spectral line VLB data is possible,
and for each spectral channel, a map can be constructed whose quality
will be primarily limited by the available coverage of the U-V plane.
The first attempt at this analysis procedure for VLBI data is
presented in Chapter III for the 1612 MHz OH maser emission of
VY CMa.

An alternative approach more generally used to infer source
structure from VLB fringe visibilities is to model either (or both)
the fringe visibility amplitudes and relative fringe phases. The
modeling of fringe visibility amplitudes to obtain source structure
(e.g., source sizes) has been extensively employed in continuum
VLB (e.g., Cohen et al., in press) and will not be discussed here. The
modeling of relative fringe phase data to determine relative position

offsets has been done by Moran et al. (1968), Johnston et al. (1971)
and Knowles et al. (1974). In most of these papers, relative fringe
rates (the time derivative of relative fringe phase) have been
analyzed because of difficulties in resolving the integral turn
ambiguities present in the fringe phase data. In practice, it is
necessary to have almost continuous observations to resolve these
ambiguities. Since relative fringe phase data tends to be at least
an order of magnitude more sensitive than fringe rate data, some
practical considerations for its use will be discussed.

A general procedure for analyzing relative fringe phase data is as
follows. Determine a position offset in right ascension, $\Delta\alpha$, and
declination, $\Delta\delta$, of a spectral component by modeling the relative
fringe rate, $\dot{\Phi}(U,V)$, as

$$\dot{\phi}(U,V) = 2\pi(\dot{U} \cos\delta_s \Delta\alpha + \dot{V}\Delta\delta) \quad . \qquad (A-17)$$

In many cases $\Delta\alpha$ and $\Delta\delta$ determined from fringe rate data may be
sufficiently accurate for experimental purposes. However, if more
precise positions are required, then relative fringe phase data must
be employed. Since the integral turn ambiguities in fringe phase
data are more easily resolved when the phase varies slowly in time,
it may be advisable to remove the effects of the position offset
unambiguously estimated from the fringe rate analysis. Gaps in the
data, for example, caused by going off source to obtain a bandpass
spectrum for auto-correlation, should be as short as possible to insure
proper extrapolation from one phase point to the next to resolve turn
ambiguities.

In general, a time gap should not be so long that the relative phase can change by $\geqslant$ 1/2 turn. Once a continuous connection between adjacent phase points has been made, position offset can be determined by modeling the relative fringe phase (in radians) as

$$\Delta\phi(t) \;=\; 2\pi(U \cos\delta_s \Delta\alpha + V\Delta\delta + c) \qquad\qquad \text{(A-18)}$$

where δ_s is the source declination and c is a constant which is needed to adjust the phases because the first relative phase point is arbitrarily chosen to lie between $-\pi/2$ and $\pi/2$ radians. Thus, a simultaneous least-squares solution for $\Delta\alpha$, $\Delta\delta$, and c is required. If all of the video frequency dependent phase terms (in equation A-4) caused by clock errors have been successfully removed from the data, then c reflects only the <u>initial</u> integral turn ambiguity. With well behaved, low-noise relative phase data continuously covering most of the U, V plane accessible with one baseline, the least-squares solution will usually indicate which integer value c should assume. Using this value for c, the data can be reanalyzed only for $\Delta\alpha$ and $\Delta\delta$ with greater accuracy and lower parameter correlations.

REFERENCES

Christiansen, W.N. and Högbom, J.A. 1969, <u>Radiotelescopes</u> (Cambridge

   University Press, London).

Clark, B.G., Weimer, R., and Weinreb, S. 1972, <u>N.R.A.O. Internal Report</u>,

   No 118.

Cohen, M.H., Moffet, A.T., Romney, J.D., Schilizzi, R.T., Shaffer, D.B.,

   Kellerman, K.I., Purcell G.H., Grove, G., Swenson, G.W. Jr.,

   Yen, J.L. Rinehart, R., Pauling-Toth, I.I.K., Preuss, E.,

   Witzel, A., and Graham, D., <u>Ap. J.</u>, to be published.

Johnston, K.J., Knowles, S.H., Sullivan, W.T. III, Moran, J.M.,

   Burke, B.F., Lo, K.Y., Papa, D.C., Papadopoulos, G.D., Schwartz,

   P.R., Knight, C.A., Shapiro, I.I., and Welch, W.J. 1971, <u>Ap. J.</u>

   <u>(Letters)</u>, <u>166</u>, L21.

Knowles, S.H., Johnston, K.J., Moran, J.M., Burke, B.F., Lo, K.Y., and

   Papadopoulos, G.D. 1974, <u>Astron. Journ.</u>, <u>79</u>, 925.

Moran, J.M. 1975, <u>Methods of Experimental Physics</u>, Vol. II.

Moran, J.M., Burke, B.F., Barrett, A.H., Rogers, A.E.E., Ball, J.A.,

   Carter, J.C., and Cudaback, D.D. 1968, <u>Ap. J. (Letters)</u>, <u>152</u>, L97.

Romney, J. 1975, in preparation.