Astronomical Interferometry at Submillimetre Wavelengths

OLIVER PETER LAY
CLARE COLLEGE
CAMBRIDGE

A dissertation submitted for the degree of
Doctor of Philosophy in the University of Cambridge

September 1994

Astronomical Interferometry at Submillimetre Wavelengths

Oliver Lay

This thesis describes my work on the project to connect the James Clerk Maxwell Telescope (JCMT) and the Caltech Submillimeter Observatory (CSO) to form the world's first astronomical interferometer operating at submillimetre wavelengths. Included are my contributions to the technical design and construction of the instrument and to the subsequent debugging and calibration, and my analysis of the data that was obtained from a number of astronomical objects.

I was responsible for much of the system that brings the signals from the receivers at each antenna to the digital correlator at the JCMT. A key feature is that the signals are carried extensively in single-mode optical fibre. The design of the delay line system is presented; this helps to equalize the travel times from the source to the correlator by inserting different lengths of optical fibre into the two signal paths. Also described is the use of a microcomputer to control a pair of oscillators, where accurate timing is necessary to ensure that a source is tracked correctly as the Earth rotates.

Careful calibration is needed to remove the effects of the instrument from the data. For an interferometer this requires a precise knowledge of many aspects of the system, including the lengths of the baseline components, thermal drift in the fibre and other components, the deformation of the dishes under gravity, the contribution of the Earth's atmosphere and the complex passband. The techniques used to make these measurements on the JCMT–CSO Interferometer are described. There is also an explanation of how software that I have written processes the raw data from the correlator to give spectra and fluxes.

After four rather brief observing sessions, the interferometer is working well and has been used successfully to observe several astronomical sources. Brief descriptions are given of observations and their interpretation for three sources: the recombination-line maser source MWC 349, the black-hole candidate Sagittarius A* at the centre of our galaxy, and the nearby radio-galaxy Centaurus A. This is followed by a more detailed analysis of data for the evolved star VY CMa and observations of discs around young stars.

Spectra of the 321 GHz H_2O line from VY CMa confirm that maser amplification is responsible for the emission. Continuum emission, spatially offset from the masers, was also detected with the interferometer, as was the CO (3–2) line. These data, particularly the offset continuum, are shown to be incompatible with single-star models for VY CMa. Instead, a binary model is proposed. This also helps to account for many previous observational results.

Many protostars are suspected of having accretion discs. The JCMT–CSO Interferometer has resolved the dust continuum emission on scales of < 100 AU for the sources HL Tau and L1551-IRS 5. An analysis of the visibility curves is presented, showing that the emission is extended perpendicular to the outflow direction in each case; radii to half-maximum brightness are determined to be 60 AU and 80 AU respectively. A lower limit to the disc masses of $\sim 0.02\ M_\odot$ is also derived.

Preface

This thesis is the result of my work undertaken at the Mullard Radio Astronomy Observatory in Cambridge between October 1991 and September 1994. This work forms part of a major project involving a number of people, and the observations in Hawaii were made as a team. Nevertheless, I had clearly defined aims and the work described here is my own, unless specifically stated otherwise. It has not, nor has any similar dissertation, been submitted for a degree, diploma or any other qualification at this or any other university. This dissertation does not exceed 60 000 words in length.

I am indebted to many people who have helped to make the last three years such an enjoyable and rewarding period of my life. I must first thank Richard Hills, my supervisor, for his patient attention to my ideas, problems and writing, and for convincing me to become involved in such a stimulating project. He and John Carlstrom have been a constant source of creative ideas and suggestions, and most of what I have learnt is due to them. I enjoyed immensely the periods spent on the interferometer in Hawaii; Mary Fuka, Colin Hall, Chris Mayer, Tom Phillips, Anthony Schinckel, and Taco all worked ridiculous hours, providing expert support and good company.

Thanks must also go to everyone at MRAO for such a relaxed and friendly working environment, in which help is never hard to find. This help often came in the form of Dave Green, always ready to give up his time to point out where I had gone astray on the computers; for this, too, I thank David Odell, Paul Scott and Dave Titterington. Outside Cambridge, I am grateful to Karl Menten at the Harvard–Smithsonian Center for Astrophysics for his hospitality and enthusiasm during my trip to Boston, and to Jeremy Yates who provided many useful comments on Chapter 5.

To my office-mates I owe a great deal: thank you, Jenny, for your cheery plants, healthy working practices, and a desk that made mine look messy; thank you, Kate, for a desk that made mine look tidy, and for keeping me sane; and thank you both, for your friendship and company. Thanks also to Andrew and all my other friends who ensured that there was always life outside work, without which I would have given up long ago. And finally to Clare: for your help, and, most of all, for making me happy.

For my family

Contents

U

Chapter 1

INTRODUCTION

1.1 Background

Radio-astronomy has come a long way since the first detection of cosmic radio emission by Jansky in 1933. The first radio interferometers, where the angular resolution of a single antenna is increased by combining its output with that of others, appeared shortly after the Second World War. The instrument built by Ryle and Vonberg (1946) consisted of two dipole antennas receiving signals at 175 MHz (λ = 1.7 m), spaced by between 10 and 140 wavelengths. Radio technology has evolved rapidly, and observing frequencies have been pushed upwards towards the far-infrared, where radio meets the optical regime. In the last decade, an unexplored region of the ground-based spectrum has been opened as two major telescopes have started observing in the submillimetre wave-band ($\lambda < 1$ mm, $\nu > 300$GHz). These are the James Clerk Maxwell Telescope (JCMT) and the Caltech Submillimeter Observatory (CSO), both located on the summit of Mauna Kea in Hawaii.

The submillimetre sky is dominated by emission from cool objects, with temperatures ranging from ~10 K to ~1000 K. The rotational transitions of simple molecules produce a rich variety of spectral lines, and warm dust is an important source of continuum flux; both can be found in very young stars that are condensing out of their parent molecular clouds, and in old stars that are returning their outer layers to the interstellar medium. Much of the research at submillimetre wavelengths has been concentrated on these areas and there is much that is not understood.

The aim of this project is to connect the JCMT and the CSO to form the first astronomical interferometer operating at submillimetre wavelengths, and to then utilize the unique combination of wavelength and resolution to investigate a range of objects, including young protostars and older, evolved stars.

In this thesis I present my contribution to the project, spanning a broad range of topics. These include the design of hardware incorporating optical fibres and the latest

digital technology, the careful calibration that is vital for successful observation with an interferometer, the software for processing the data, and finally some of the exciting scientific results that have emerged, particularly concerning discs around young stars.

In the rest of this chapter I give a general description of the JCMT–CSO Interferometer, including the design of the system, its expected performance, likely astronomical targets and a brief history of how the project has evolved.

I will start with an introduction to the concepts and jargon of radio interferometry. It is my intention to keep things simple and to restrict the analysis to issues relevant to the JCMT–CSO Interferometer. An exhaustive description of almost all aspects of radio interferometry can be found in Thompson, Moran and Swenson's (1986) book, *'Interferometry and Synthesis in Radio Astronomy'*. Other useful references include *'Tools of Radio Astronomy'* by Rohlfs (1986) and *'Radiotelescopes'* by Christiansen and Högbom (1985). In this chapter, and throughout this thesis, I assume without derivation the basic formulae of radiative transfer—specifically the Planck function for emission from a black-body and the Rayleigh–Jeans limit for low frequencies, the relationships between brightness, luminosity and flux, and the concept of optical depth. Information on these topics can be found in most radio-astronomy texts.

1.2 Basic Interferometry

1.2.1 The response of a simple interferometer

Figure 1.1 depicts a two-element interferometer. The antennas are separated by a baseline of length d and convert incoming electromagnetic radiation to voltages $V_1(t)$ and $V_2(t)$. The output of the correlator, $C_{\rm out}$, is the time-averaged product of the two signals.

Consider a very distant, monochromatic point source located a small angle $\Delta\theta$ away from zenith; the signal, of frequency $\nu_{\rm sky}$, has slightly further to travel to antenna 1 compared to antenna 2 and the voltages can be represented by:

$$V_1(t) = V_{\rm m1}\cos\left\{2\pi\left(\nu_{\rm sky}t - \frac{d\sin\Delta\theta}{\lambda_{\rm sky}}\right)\right\}, \quad (1.1)$$

$$V_2(t) = V_{\rm m2}\cos\left\{2\pi\nu_{\rm sky}t\right\}, \quad (1.2)$$

where $\lambda_{\rm sky}$ is the wavelength of the signals in the atmosphere and $V_{\rm m1}, V_{\rm m2}$ are the maximum amplitudes of the voltage. Using one of the trigonometrical identities, it is easy to show that the product of the two voltages is

$$V_1V_2 = \frac{1}{2}V_{\rm m1}V_{\rm m2}\left[\cos\left\{2\pi\left(2\nu_{\rm sky}t - \frac{d\sin\Delta\theta}{\lambda_{\rm sky}}\right)\right\} + \cos\left\{\frac{2\pi d\sin\Delta\theta}{\lambda_{\rm sky}}\right\}\right]. \quad (1.3)$$

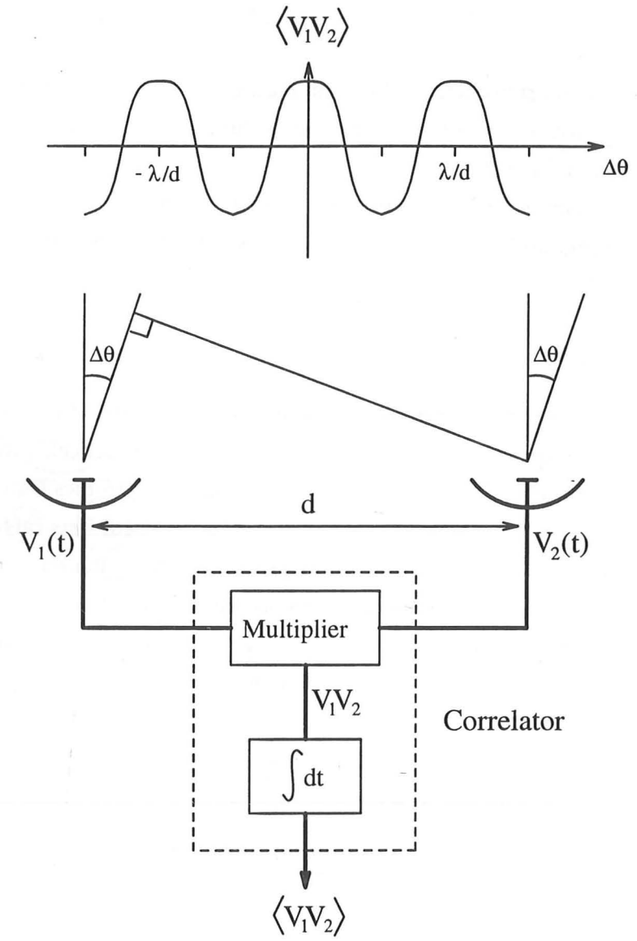


Figure 1.1: A simple interferometer. The response on the sky is shown in the upper half.

After time-averaging, only the second term remains:

$$\langle V_1 V_2 \rangle = \frac{1}{2} V_{m1} V_{m2} \cos\left\{\frac{2\pi d \sin\Delta\theta}{\lambda_{sky}}\right\}. \quad (1.4)$$

For a source close to zenith, $\Delta\theta$ is small and we can approximate with

$$\langle V_1 V_2 \rangle = \frac{1}{2} V_{m1} V_{m2} \cos\left\{\frac{2\pi d \Delta\theta}{\lambda_{sky}}\right\}. \quad (1.5)$$

This response of the interferometer is shown in Figure 1.1. The cosine curve, with a spacing of λ/d between the maxima, is often referred to as the 'fringe'. The fringe-spacing sets the angular resolution of the instrument.

It can be seen from equation 1.5 that the correlator output is determined both by the strength of the signal and the difference in length of the two alternative paths from the astronomical source.

1.2.2 A heterodyne system

In the case just described the signals are correlated at the same frequency as they are received. At high frequencies this becomes increasingly impractical, primarily because the cabling has insufficient bandwidth and the correlator electronics cannot operate fast enough. Instead, the technique of heterodyne detection is used, illustrated in Figure 1.2. The output of a local oscillator (LO), of frequency ν_{lo} at each antenna, is mixed with the incoming signals to produce the sum and difference frequencies. Successful interferometry requires that the two local oscillators must be highly coherent with each other. This is achieved by phase-locking to a common reference signal. In practice, only the much lower, difference frequency is passed; this is termed the intermediate frequency (IF). If the mixer is single sideband then the IF signal can be one of two forms. For upper sideband reception,

$$\cos\left\{2\pi\left(\nu_{sky}t - \frac{x}{\lambda_{sky}}\right)\right\} \longrightarrow \cos\left\{2\pi\left((\nu_{sky} - \nu_{lo})t - \frac{x}{\lambda_{sky}}\right)\right\}, \quad (1.6)$$

while for the lower sideband,

$$\cos\left\{2\pi\left(\nu_{sky}t - \frac{x}{\lambda_{sky}}\right)\right\} \longrightarrow \cos\left\{2\pi\left((\nu_{lo} - \nu_{sky})t - \frac{x}{\lambda_{sky}}\right)\right\}. \quad (1.7)$$

In each case, $\nu_{if} = |\nu_{sky} - \nu_{lo}|$. The two IF signals are then correlated as before to give the same basic cosine response on the sky. However, it is more usual for the mixer to be double sideband; in this case both the upper and lower sideband signals are present simultaneously and we must separate the two contributions.

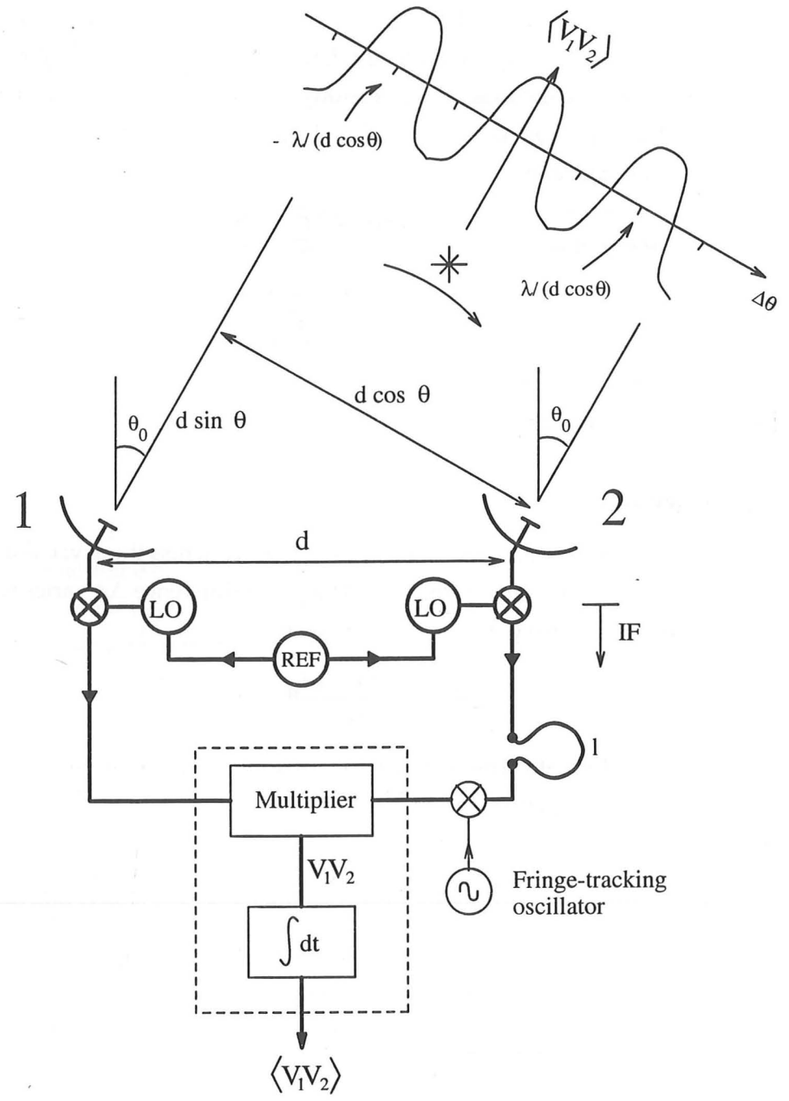


Figure 1.2: A heterodyne interferometer with delay compensation and fringe-tracking oscillator.

1.2.3 Delay lines

We have so far considered the source to be close to the zenith position. In general this is not the case. In Figure 1.2 there is a large, extra path length of $d \sin\theta$ to antenna 1. The time taken for the signals to travel this length is termed the geometrical delay, and we must compensate by using a delay line, of length l, to add extra path to the antenna 2 side. The phase compensation is not exact for a heterodyne system because the signal passes through the delay line at a frequency of $\nu_{\rm if}$. The phase difference between the signals as they enter the correlator is now

$$\Delta\phi = \phi_1 - \phi_2 = \frac{2\pi d \sin(\theta + \Delta\theta)}{\lambda_{\rm sky}} - \frac{2\pi l}{\lambda_{\rm if}} \quad (1.8)$$

$$= 2\pi \left\{ \frac{d \sin\theta}{\lambda_{\rm sky}} - \frac{l}{\lambda_{\rm if}} \right\} + \frac{2\pi \Delta\theta d \cos\theta}{\lambda_{\rm sky}} \quad (1.9)$$

The response is still a cosine curve, but the spacing is now $\lambda_{\rm sky}/(d\cos\theta)$ and in the general case there is no longer a maximum at $\Delta\theta = 0$.

1.2.4 The fringe-rate

As the Earth rotates about its axis, a source appears to move across the sky; θ is a function of t and the geometrical delay slowly changes. The phase difference $\Delta\phi$ varies with time; the rate of change, for a source with $\Delta\theta = 0$, is given by

$$\frac{d\Delta\phi}{dt} = \frac{2\pi d \cos\theta}{\lambda_{\rm sky}} \dot{\theta}, \quad (1.10)$$

and is known as the fringe-rate. The output of the correlator oscillates as the source moves across the response fringe. Equivalently, if the source is moving as depicted in Figure 1.2, then the signal received by antenna 1 will be slightly red-shifted and the signal received by antenna 2 will be blue-shifted. This small difference in frequencies gives rise to the fringe-rate.

When observing a point source, we require the output of the correlator to be constant over time, and an additional oscillator is needed to ensure this. In Figure 1.2, the output phase of the fringe-tracking oscillator is subtracted from the IF phase of antenna 1 using an additional mixer that operates in the upper sideband. If the phase of this oscillator is (cf. equation 1.9)

$$\phi_{\rm osc} = 2\pi \left\{ \frac{d \sin\theta}{\lambda_{\rm sky}} - \frac{l}{\lambda_{\rm if}} \right\}, \quad (1.11)$$

then going into the correlator

$$\Delta\phi = \frac{2\pi \Delta\theta d \cos\theta}{\lambda_{\rm sky}}. \quad (1.12)$$

The response curve, $\cos\Delta\phi$, now has a maximum at $\Delta\theta = 0$. The output of the correlator no longer oscillates and the fringe tracks the source as it moves across the sky. It is also possible to compensate for the fringe-rate by offsetting the LO frequencies.

1.2.5 Phase-switching

A correlator should give zero output when there is no correlation of the input signals. This is rarely the case, since any bias in a correlator will introduce an offset. The technique of phase-switching is used to combat this. At some stage before the correlator, an extra 180° of phase is introduced periodically to one of the IF signals. This is equivalent to reversing the voltage of the signal and flips the sign of the correlator output. If we alternately add and subtract the output of the correlator, in step with the 180° phase-switching, then the resultant output is the same as it was without phase-switching, except that any correlator bias is effectively removed.

1.2.6 Complex visibility

A simple correlator gives zero output for a source at the $\Delta\theta = \lambda/2d$ position; the source is not detected. Such a correlator only measures the quantity $V_{\rm m1}V_{\rm m2}\cos(2\pi d\Delta\theta/\lambda_{\rm sky})$, and we cannot determine both the strength ($V_{\rm m1}V_{\rm m2}$) and position ($\Delta\theta$) of the source. In order to do this a complex correlator is needed. A 'sine' correlation is measured in addition to the normal 'cosine' by introducing an extra phase-shift of $\pi/2$ into one of the IF signals, as shown in Figure 1.3. The output of the 'sine' correlator is given by

$$C_{\sin} = \frac{1}{2} V_{\rm m1}V_{\rm m2} \cos\left\{ \frac{2\pi\Delta\theta d\cos\theta}{\lambda_{\rm sky}} - \frac{\pi}{2} \right\} \quad (1.13)$$

$$= \frac{1}{2} V_{\rm m1}V_{\rm m2} \sin\left\{ \frac{2\pi\Delta\theta d\cos\theta}{\lambda_{\rm sky}} \right\}. \quad (1.14)$$

Both sine and cosine responses are shown in Figure 1.3. These two components of the response to a source form the complex visibility. This has an amplitude proportional to the flux, and a phase that depends on the position of the source. The response of an interferometer with a complex correlator can be pictured as a phase-gradient on the sky, as shown in Figure 1.3 below the sine and cosine responses.

It is instructive to see how the output of the correlator depends on the source size and position and I have tried to illustrate this in Figure 1.4. The full flux is detected for a source whose size is small compared to the fringe-spacing—such a source is said to be unresolved. To calculate the visibility flux of a larger source, we must make the vector addition over all the small, unresolved elements that make up the total brightness distribution. For example, we will detect 64% of the total flux from a source whose brightness is distributed

1.2.3 Delay lines

We have so far considered the source to be close to the zenith position. In general this is not the case. In Figure 1.2 there is a large, extra path length of $d \sin\theta$ to antenna 1. The time taken for the signals to travel this length is termed the geometrical delay, and we must compensate by using a delay line, of length l, to add extra path to the antenna 2 side. The phase compensation is not exact for a heterodyne system because the signal passes through the delay line at a frequency of $\nu_{\rm if}$. The phase difference between the signals as they enter the correlator is now

$$\Delta\phi = \phi_1 - \phi_2 = \frac{2\pi d \sin(\theta + \Delta\theta)}{\lambda_{\rm sky}} - \frac{2\pi l}{\lambda_{\rm if}} \quad (1.8)$$

$$= 2\pi \left\{ \frac{d \sin\theta}{\lambda_{\rm sky}} - \frac{l}{\lambda_{\rm if}} \right\} + \frac{2\pi \Delta\theta d \cos\theta}{\lambda_{\rm sky}} \quad (1.9)$$

The response is still a cosine curve, but the spacing is now $\lambda_{\rm sky}/(d\cos\theta)$ and in the general case there is no longer a maximum at $\Delta\theta = 0$.

1.2.4 The fringe-rate

As the Earth rotates about its axis, a source appears to move across the sky; θ is a function of t and the geometrical delay slowly changes. The phase difference $\Delta\phi$ varies with time; the rate of change, for a source with $\Delta\theta = 0$, is given by

$$\frac{d\Delta\phi}{dt} = \frac{2\pi d \cos\theta}{\lambda_{\rm sky}} \dot{\theta}, \quad (1.10)$$

and is known as the fringe-rate. The output of the correlator oscillates as the source moves across the response fringe. Equivalently, if the source is moving as depicted in Figure 1.2, then the signal received by antenna 1 will be slightly red-shifted and the signal received by antenna 2 will be blue-shifted. This small difference in frequencies gives rise to the fringe-rate.

When observing a point source, we require the output of the correlator to be constant over time, and an additional oscillator is needed to ensure this. In Figure 1.2, the output phase of the fringe-tracking oscillator is subtracted from the IF phase of antenna 1 using an additional mixer that operates in the upper sideband. If the phase of this oscillator is (cf. equation 1.9)

$$\phi_{\rm osc} = 2\pi \left\{ \frac{d \sin\theta}{\lambda_{\rm sky}} - \frac{l}{\lambda_{\rm if}} \right\}, \quad (1.11)$$

then going into the correlator

$$\Delta\phi = \frac{2\pi \Delta\theta d \cos\theta}{\lambda_{\rm sky}}. \quad (1.12)$$

The response curve, $\cos\Delta\phi$, now has a maximum at $\Delta\theta = 0$. The output of the correlator no longer oscillates and the fringe tracks the source as it moves across the sky. It is also possible to compensate for the fringe-rate by offsetting the LO frequencies.

1.2.5 Phase-switching

A correlator should give zero output when there is no correlation of the input signals. This is rarely the case, since any bias in a correlator will introduce an offset. The technique of phase-switching is used to combat this. At some stage before the correlator, an extra 180° of phase is introduced periodically to one of the IF signals. This is equivalent to reversing the voltage of the signal and flips the sign of the correlator output. If we alternately add and subtract the output of the correlator, in step with the 180° phase-switching, then the resultant output is the same as it was without phase-switching, except that any correlator bias is effectively removed.

1.2.6 Complex visibility

A simple correlator gives zero output for a source at the $\Delta\theta = \lambda/2d$ position; the source is not detected. Such a correlator only measures the quantity $V_{\rm m1}V_{\rm m2}\cos(2\pi d\Delta\theta/\lambda_{\rm sky})$, and we cannot determine both the strength ($V_{\rm m1}V_{\rm m2}$) and position ($\Delta\theta$) of the source. In order to do this a complex correlator is needed. A 'sine' correlation is measured in addition to the normal 'cosine' by introducing an extra phase-shift of $\pi/2$ into one of the IF signals, as shown in Figure 1.3. The output of the 'sine' correlator is given by

$$C_{\rm sin} = \frac{1}{2} V_{\rm m1} V_{\rm m2} \cos\left\{ \frac{2\pi \Delta\theta d \cos\theta}{\lambda_{\rm sky}} - \frac{\pi}{2} \right\} \quad (1.13)$$

$$= \frac{1}{2} V_{\rm m1} V_{\rm m2} \sin\left\{ \frac{2\pi \Delta\theta d \cos\theta}{\lambda_{\rm sky}} \right\}. \quad (1.14)$$

Both sine and cosine responses are shown in Figure 1.3. These two components of the response to a source form the complex visibility. This has an amplitude proportional to the flux, and a phase that depends on the position of the source. The response of an interferometer with a complex correlator can be pictured as a phase-gradient on the sky, as shown in Figure 1.3 below the sine and cosine responses.

It is instructive to see how the output of the correlator depends on the source size and position and I have tried to illustrate this in Figure 1.4. The full flux is detected for a source whose size is small compared to the fringe-spacing—such a source is said to be unresolved. To calculate the visibility flux of a larger source, we must make the vector addition over all the small, unresolved elements that make up the total brightness distribution. For example, we will detect 64% of the total flux from a source whose brightness is distributed

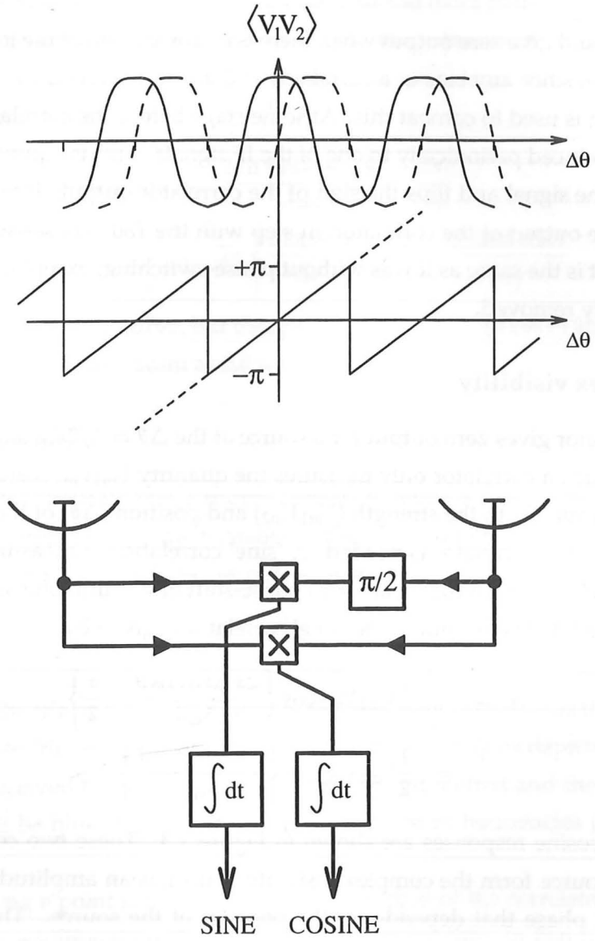


Figure 1.3: The basis of a complex correlator. The sine and cosine responses are shown at the top, with the equivalent phase response below.

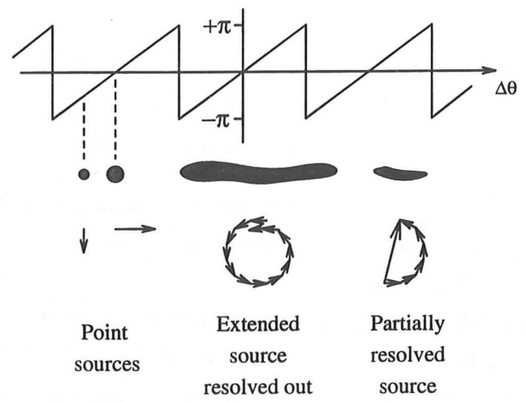


Figure 1.4: The response of a complex correlator to sources of different size (shaded). The arrows represent the vectors that make up the overall response. The length of the vector is proportional to the flux of a small element and the phase depends on the position.

uniformly across half a fringe, and exactly zero if it is evenly distributed across a whole number of fringes. A source is said to be resolved out if it is much larger than the fringe-spacing, since very little of the total flux is detected. An interferometer is therefore not sensitive to extended emission.

1.2.7 Imaging with an interferometer

My treatment of interferometry so far has been set in two-dimensional space for clarity. Extension of the results to three dimensions is straightforward and Figure 1.5 shows the vectors that are needed. $\mathbf{D}_\lambda$ is the baseline vector in units of the wavelength. The phase

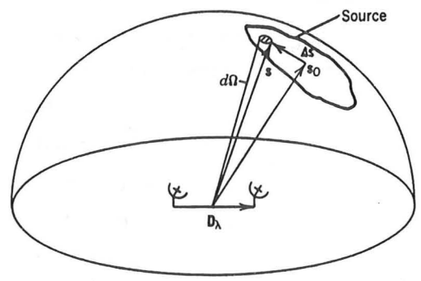


Figure 1.5: Definitions of the vectors $\mathbf{D}_\lambda$, $\mathbf{s}$, $\mathbf{s}_0$ and $\Delta\mathbf{s}$, taken from Thompson, Moran and Swenson (1986).

difference between the signals arriving at the antennas is given by

$$\begin{aligned} \Delta\phi_{\text{ant}} &= 2\pi \mathbf{D}_\lambda \cdot \mathbf{s} && (1.15)\\ &= 2\pi \mathbf{D}_\lambda \cdot (\mathbf{s}_0 + \mathbf{\Delta s}) && (1.16)\\ &= 2\pi \mathbf{D}_\lambda \cdot \mathbf{s}_0 + 2\pi \mathbf{D}_\lambda \cdot \mathbf{\Delta s}. && (1.17) \end{aligned}$$

The first term in equation 1.17 is cancelled by the delay lines and the fringe-tracking oscillator (equation 1.11 must be modified for three dimensions) and the second is equivalent to the term $2\pi\Delta\theta d\cos\theta/\lambda_{\text{sky}}$ of equation 1.9.

An interferometer can usefully be regarded as an instrument that samples particular spatial frequencies of a brightness distribution on the sky. With a complex correlator we are measuring sine and cosine components (or an amplitude and phase) whose period and direction on the sky are set by the orientation of the baseline vector with respect to the source. There is a Fourier-transform relationship between the brightness distribution $B(\mathbf{\Delta s})$, and the complex visibilities measured by the interferometer, $\mathcal{V}(\mathbf{D}_\lambda)$:

$$\mathcal{V}(\mathbf{D}_\lambda) = \int_{4\pi} [G_1(\mathbf{\Delta s})G_2(\mathbf{\Delta s})]^{0.5} B(\mathbf{\Delta s}) e^{-2\pi i \mathbf{D}_\lambda \cdot \mathbf{\Delta s}} d\Omega. \quad (1.18)$$

$G_1(\mathbf{\Delta s})$ is the gain of antenna 1 in direction $(\mathbf{s}_0 + \mathbf{\Delta s})$ relative to the maximum gain in direction $\mathbf{s}_0$. Obviously we can only detect a source with an interferometer if both dishes are pointing at it. Ω represents the solid angle subtended at the Earth by the source.

The fringe-spacing is determined by the apparent length of the baseline as seen by the source, that is $|\mathbf{s}_0 \times \mathbf{D}_\lambda|$. It is conventional to denote the components of $\mathbf{D}_\lambda$ in the plane perpendicular to $\mathbf{s}_0$ by (u, v). I will take u to be positive to the east and v positive to the north; both are expressed as multiples of the observing wavelength. For a source in the sky being observed by a fixed, single-baseline interferometer on the Earth, the baseline vector sweeps out a cone as it rotates about the Earth's axis (Figure 1.6a). The locus of the baseline plotted in the (u, v) plane depends on the declination of the source (Figure 1.6b). The locus reflected through the origin is also marked, since the sense of the baseline vector ($1 \rightarrow 2$ or $2 \rightarrow 1$) is arbitrary. The apparent length of the baseline as seen by the source varies with time; only sources on the celestial equator see the full length as they cross the meridian, and it is here that the interferometer has its minimum fringe-spacing and maximum resolution. Figures 1.6c and 1.6d show the relationship between a point in the (u, v) plane and the corresponding spacing and orientation of the fringes on the sky.

Each point in the (u, v) plane corresponds to a different spatial frequency. In order to reconstruct an image of the brightness distribution, we would like to measure as many of the spatial frequencies as possible. With only a single baseline, however, the (u, v) coverage is poor; for a given source we can only sample a limited range of its spatial components (Figure 1.6c). Although an image can be formed from the Fourier transform

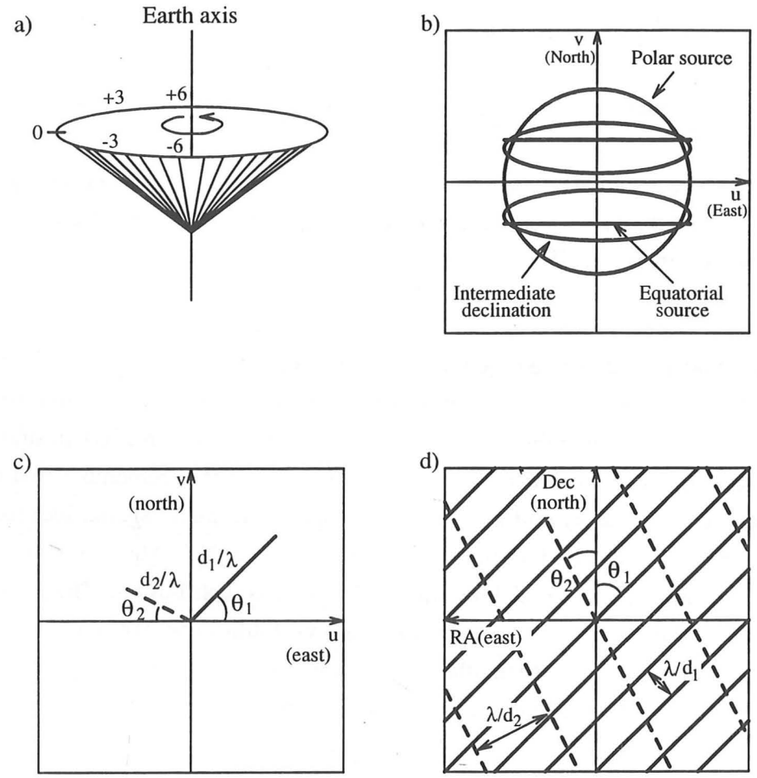


Figure 1.6: Baseline geometry and the (u, v) plane for an interferometer. (a) The cone swept out by the baseline vector as the Earth rotates. (b) The locus of the baseline vector represented in the (u, v) plane for sources at different declinations. (c) Two different projected baselines that give the fringes on the sky depicted in (d). Note that by the convention adopted here, east is to the right in (c) and to the left in (d).

Frequency range /GHz	Wavelength range /mm	Transmission	JCMT designation
200 – 320	1.50 – 0.94	0.96	A
320 – 380	0.94 – 0.79	0.88	B
450 – 530	0.67 – 0.57	0.43	C
620 – 720	0.48 – 0.42	0.38	—
800 – 900	0.38 – 0.33	0.44	—

Table 1.1: Windows of atmospheric transmission. Figures are approximate. Transmission is for an atmosphere with 0.5 mm of precipitable water vapour. See the JCMT user guide (Matthews 1991) for more details.

of the visibility data, in practice there is not much to be gained by doing so. Alternatively, we can say that a fixed, single-baseline interferometer has a poor synthesized beam.

There are two simple methods for obtaining spatial information without forming an image. Firstly, the phase of the complex visibility is a measure of relative positional offset in the direction perpendicular to the fringes. It is important to remember that the phase has an ambiguity of $\pm n\pi$, and one phase corresponds to many spatial locations. I use this method in chapter 5 to determine relative positions for VY CMa. The other technique that I have used involves modelling the sky brightness distribution. The model can be transformed to the (u, v) plane and the expected visibilities compared against the actual data. Some models fit better than others, providing information on the spatial structure of the source. This method is central to the analysis presented in chapter 6.

1.3 The world's first astronomical submillimetre interferometer

1.3.1 Submillimetre technology

The submillimetre region of the electromagnetic spectrum extends from $\lambda = 1$ mm (300 GHz) down to $\lambda \approx 0.1$ mm (3 THz). Observations from ground-based sites are limited to the windows of atmospheric transmission, shown in Table 1.1. Water vapour is the principal cause of absorption in the atmosphere and consequently submillimetre telescopes are sited in high and dry locations. There are currently two large telescopes that were designed to operate primarily at submillimetre wavelengths: the James Clerk Maxwell Telescope (JCMT) and the Caltech Submillimeter Observatory (CSO). Both are sited at an altitude of 4100 m on the summit of Mauna Kea in Hawaii. A third, the Submillimeter Telescope (SMT), will start observing soon on Mount Graham, Arizona. There are several smaller instruments, including the 3-m KOSMA dish at Gornergrat, Switzerland. In addition to these, the 15-m Swedish–ESO Submillimeter Telescope (SEST) in Chile and the 30-m dish of the Institute de Radioastronomie Millimétrique (IRAM) on Pico Veleta in Spain can operate at up to 345 GHz.

Dish surfaces need to be very accurately figured at these frequencies; a surface with an rms error of $\lambda/20$ loses ~40% of the signal. For a dish operating at up to 0.4 mm, for example, a tolerance of 20 μm is desirable.

Submillimetre receivers are of two types. Bolometers directly detect the continuum energy by measuring the temperature of a small piece of crystal with a low heat capacity. The detector must be cooled cryogenically to minimize thermal noise. Until now they have always been single pixel detectors, but bolometer arrays are starting to make an impact on the efficiency of observing. I have already described the principles of heterodyne detection, where a local oscillator signal is used to down-convert the incoming radiation to a more manageable frequency that can be handled by electronics. The development of Superconducting–Insulator–Superconducting (SIS) junction mixers over the last decade has greatly increased the sensitivity. These devices must also be cooled to liquid helium temperatures to make them superconducting. The sources of spectrally-pure local oscillator power have also undergone substantial development; the combination of a Gunn diode and a frequency multiplier has now become established as the standard at the lower frequencies and looks set to push higher as the technology improves. The output of a phase-locked Gunn diode, with a frequency of up to ~150 GHz, can be multiplied up by factors of 2, 3, 4, 5 and 6; together with the wide tuning range that is now possible, this ensures that there is good frequency coverage over most of the band.

1.3.2 The James Clerk Maxwell Telescope

The JCMT is located near the summit of Mauna Kea on Hawaii's Big Island. It is financed and run by the UK, Holland and Canada, and started operation in 1987. The primary mirror has a diameter of 15 m and is housed in a large enclosure with a membrane for protection from the sun and wind. The dish is supported on an alt-az–type mount. A secondary mirror reflects the light back through an aperture at the centre of the primary, after which it hits a flat tertiary mirror that can rotate about the optical axis. This mirror redirects the light to any one of the heterodyne receivers mounted in the cabin on the back of the primary, or to the bolometer at the Nasmyth focus. The principal heterodyne receivers currently in use are denoted A2 (the second receiver built to operate in the 'A' band of Table 1.1), B3 (this has superseded B3i) and C2. Both A2 and B3 down-convert to an IF that spans 1–2 GHz, whereas C2 has an IF in the range 3.5–4.5 GHz.

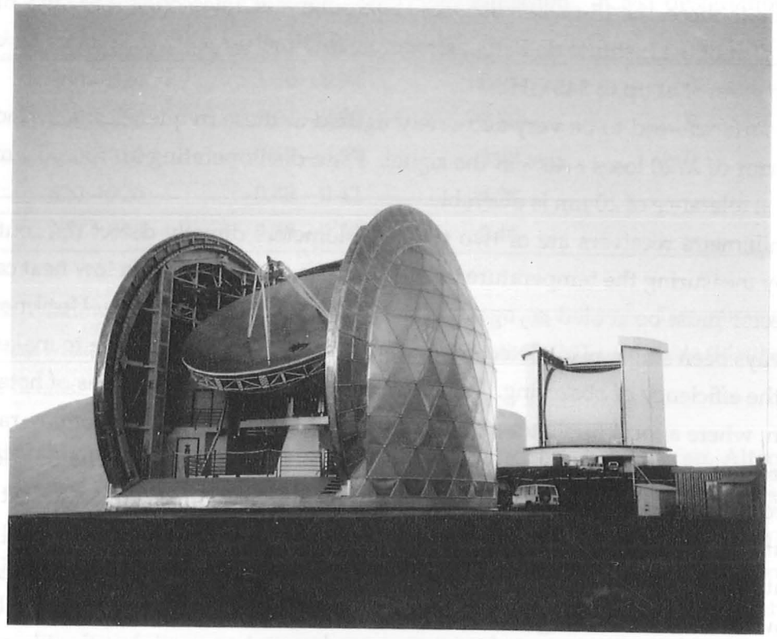


Figure 1.7: The JCMT–CSO Interferometer (the CSO is in the foregound).

There are two spectrometers at the JCMT; the Acousto-Optical Spectrometer (AOS) has now been largely superseded by the Dutch Autocorrelation Spectrometer (DAS). We use the latter for interferometry, and part of chapter 2 explains how it works.

1.3.3 The Caltech Submillimetre Observatory

The CSO became operational at about the same time as the JCMT and is funded both privately and by the National Science Foundation. It has a 10-m primary mirror on an alt-az mount which, although enclosed, does not have a protective membrane. Instruments can be located either at the Cassegrain focus or in the side-cabin, on a plate that rotates in elevation with the dish. There are heterodyne receivers for operating in the 'A', 'B', 'C' and higher frequency bands (although they are not referred to by this JCMT terminology), all of which generate IF signals at 1–2 GHz. There is currently no digital correlator at the CSO.

The CSO is located approximately 164 m to the east-southeast of the JCMT (Figure 1.7).

This relative location of the two sites was chosen with the prospect of interferometry in mind.

1.4 The design of the JCMT–CSO Interferometer

I now describe the design of the interferometer in more detail. The aim was to make as much use of existing hardware as possible—this includes the heterodyne receivers at each telescope and the digital correlator at the JCMT. The main additions are a new phase-lock system for the two local oscillators and a new IF system to deliver the signals from the two receivers to the correlator at the JCMT. I have been responsible for implementing most of the IF system, including the design of the delay lines and the interfacing to a microcomputer. Although I did not make any significant contribution to the hardware of the phase-lock system, it is relevant to much of the fringe-tracking software that I wrote.

1.4.1 The phase-lock system

The purpose of the phase-lock system is to ensure that the local oscillators at the two telescopes, separated by over 150 m, are highly coherent with one another at frequencies of up to 500 GHz and beyond. Most of this work was done by John Carlstrom at Caltech, so I will not describe this system in great detail.

Figure 1.8 shows the relevant components of the system in schematic form. The primary output of the synthesizer, with $\nu_{\rm synth} \sim 1$ GHz, is initially multiplied by six, and basically determines the final output frequencies of the local oscillators. We used a Hewlett-Packard 8662 to guarantee the requisite high stability. The CSO control computer instructs the synthesizer to generate the appropriate frequency, taking into account the radial velocity of the source with respect to the telescopes. The HP8662 is also used to produce outputs at 10 MHz and 640 MHz. Both the 6 GHz and the 640 MHz are transmitted to the JCMT via a fibre-optic link. The use of fibres is a major feature of the project and is discussed at greater length in chapter 2. The 640 MHz is divided by a factor of 64 to give 10 MHz at the JCMT. Ideally, the 10-MHz signal should be transmitted directly, but the link has a filter that rejects frequencies below ~100 MHz. The frequency-division at the JCMT has proved to be problematic (see chapter 3).

The chains of phase-locked loops (PLLs) at the CSO and JCMT are almost identical. The control box of a PLL has three inputs, and one output that controls the frequency of the oscillator in the loop. It tries to maintain the phase of the oscillator so that $\phi_{\rm osc} = \phi_{\rm in} \pm \phi_{\rm ref}$. The sign depends on whether the loop is operating in an upper or lower sideband mode. The PLL is used to clean up and amplify the input signal and behaves like a fly-wheel, smoothing out phase-jitter and irregularities in the input. Each telescope has two loops:

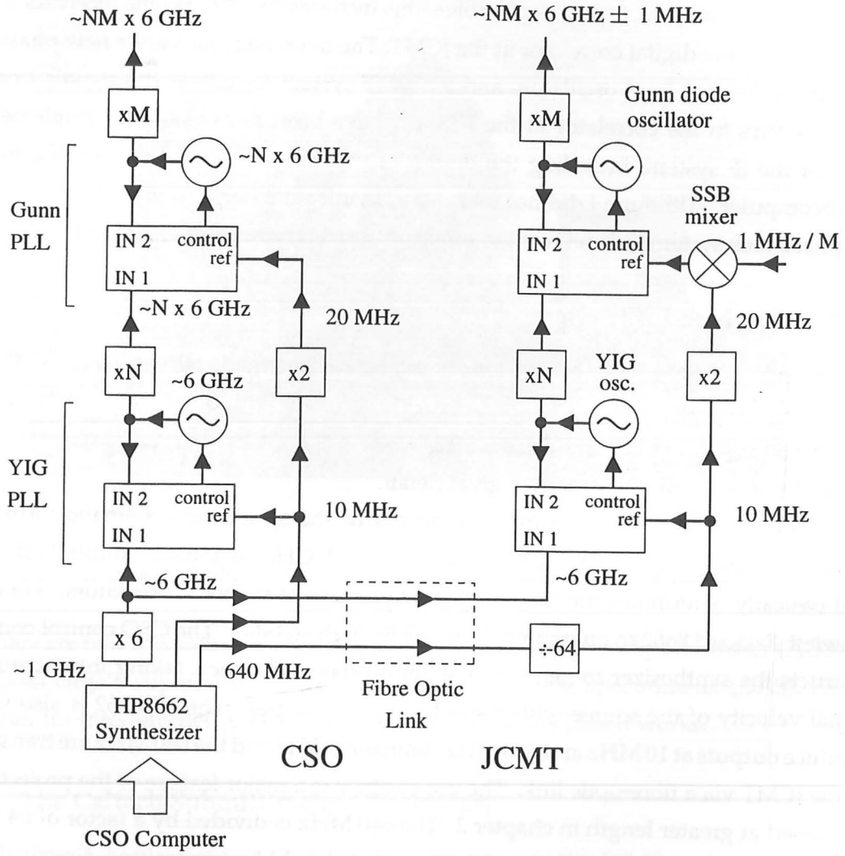


Figure 1.8: The phase-lock system that maintains phase-coherence between the local oscillators at the JCMT and the CSO.

the YIG loop with a frequency of ~6 GHz and the Gunn loop working at ~90 GHz. More detailed information on PLLs can be found in Gardner (1979).

A harmonic mixer in each arm of the system is used to increase the frequency by a factor N. A further increase, by a factor M, is achieved with a frequency multiplier, to give the final LO frequency. In a typical case for operation at 'B' band, $N = 14$ and $M = 4$. With these values, a starting frequency of 6 GHz is multiplied up to 344 GHz. An additional 1 MHz is added or subtracted from the final output on the JCMT side using an extra single sideband mixer at the Gunn PLL; this is used to reject one of the two sidebands generated by down-conversion to the IF (see §1.5.1).

The phases of the two outputs of the local oscillator system can be represented by

$$\phi_{\rm clo} = M\{N(\phi_{\rm synth} + \phi_{10}) + 2\phi_{10}\} + \phi_{\rm const1}, \quad (1.19)$$

$$\phi_{\rm jlo} = M\left\{N(\phi_{\rm synth} + \phi_{10}) + 2\phi_{10} + (sb)\frac{\phi_1}{M}\right\} + \phi_{\rm const2}, \quad (1.20)$$

where ϕ_{10} and ϕ_1 are the phases corresponding to frequencies of 10 MHz and 1 MHz. $\phi_{\rm const1}$ and $\phi_{\rm const2}$ are phase offsets that are functions of the local oscillator frequency $\nu_{\rm lo}$. For observing in the upper sideband, $sb = +1$; for the lower sideband $sb = -1$. So

$$\phi_{\rm jlo} = \phi_{\rm clo} + (sb)\phi_1 + \Delta\phi(\nu_{\rm lo}). \quad (1.21)$$

1.4.2 The IF system

Figure 1.9 shows the path that signals take from the sky to the correlator. Existing heterodyne SIS junction receivers at the two telescopes down-convert the frequency to an IF in the range 1–2 GHz using the coherent local oscillators described in the last section. The IF signals are then delivered to the control room at the JCMT using optical fibres. The fibre from CSO to JCMT was pulled through a conduit between the two sites and has a total length of ~300 m. There is a loop of fibre on the JCMT side to compensate for this. Coarse delay compensation is achieved by switching different lengths of fibre into the signal path using optical fibre switches—this is the first use of such a system on an astronomical interferometer. The signals are then converted back from modulations of light intensity to electrical waveforms and pass into the correlator. On the way, the frequency of the JCMT IF is shifted upwards by an amount (1 MHz $\pm\nu_{\rm fringe}$) using a Direct Digital Synthesizer (DDS) and a single sideband mixer. The $\pm\nu_{\rm fringe}$ component takes care of the fringe-tracking (see §1.2.4), with the sign determined by the observing sideband. The 1 MHz, in conjunction with the 1-MHz offset of the JCMT LO frequency, is used for sideband rejection (§1.5.1). A microcomputer controls both the switching of the delay lines and the output of the DDS.

Chapter 2 is devoted to a more complete description of the correlator, the delay lines, the use of optical fibre and the control of the DDSs.

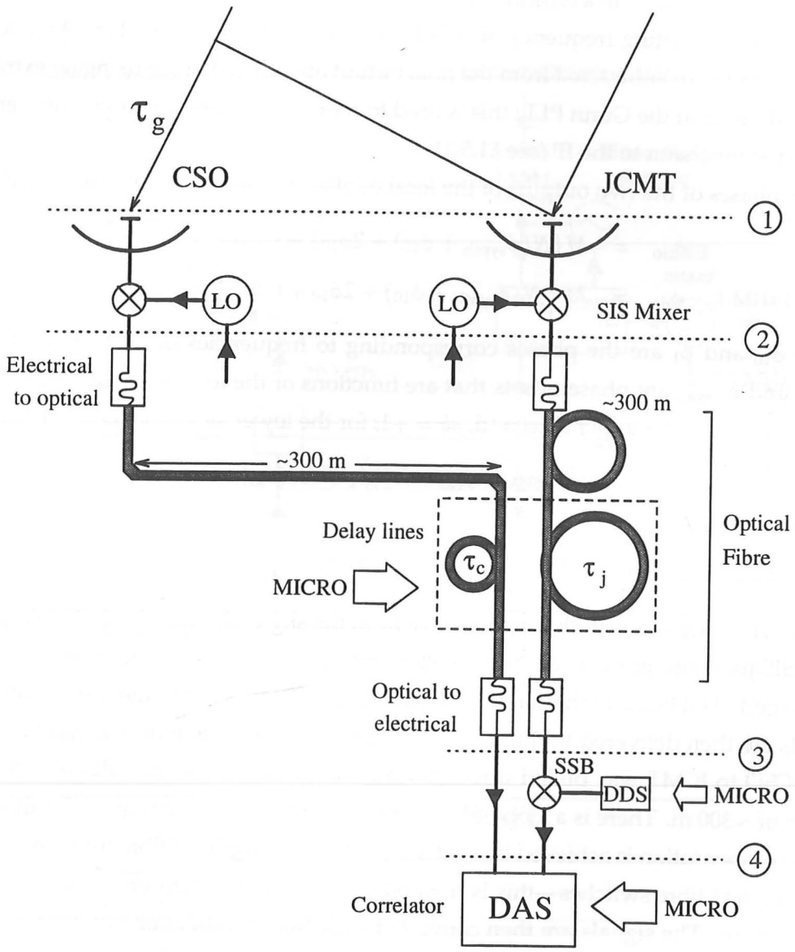


Figure 1.9: The two paths followed by signals from sky to correlator. The numbers are referred to in the main text.

1.5 Getting the phase right

1.5.1 The correlator output

This section is intended to illustrate how the whole system works, by following the phases of the signals as they travel from the source to the correlator. I use ϕ_{j1} to refer to the phase of the JCMT signal at position (1) on Figure 1.9, ϕ_{c2} for the CSO phase at position (2), etc. For an incoming signal of frequency ν_{sky}:

$$\phi_{j1} = 2\pi\nu_{sky}t, \tag{1.22}$$

$$\phi_{c1} = 2\pi\nu_{sky}(t-\tau_g), \tag{1.23}$$

where τ_g is the geometrical delay. At the point just after the mixer, there are two possibilities for producing a particular intermediate frequency; the relevant phases are

$$\phi_{j2} = \pm\{2\pi\nu_{sky}(t-\Delta t_{j12})-\phi_{jlo}\}, \tag{1.24}$$

$$\phi_{c2} = \pm\{2\pi\nu_{sky}(t-\tau_g-\Delta t_{c12})-\phi_{clo}\}. \tag{1.25}$$

The plus sign applies to signals with $(\nu_{lo}+1\,\text{GHz}) \le \nu_{sky} \le (\nu_{lo}+2\,\text{GHz})$; these are in the upper sideband. The minus sign is for signals in the lower sideband, with $(\nu_{lo}-2\,\text{GHz}) \le \nu_{sky} \le (\nu_{lo}-1\,\text{GHz})$. I have used Δt_{j12} to represent the time taken for the signals to propagate between points (1) and (2) on the JCMT side and similarly for Δt_{c12} at the CSO. The local oscillator phases ϕ_{jlo} and ϕ_{clo} are as defined in equation 1.21. Now that they have been down-converted, the signals have frequencies of ν_{jif} and ν_{cif} for the JCMT and CSO respectively. There are two possible sky frequencies for a given intermediate frequency, one from each sideband. After propagation through the delay lines,

$$\phi_{j3} = \pm\{2\pi\nu_{sky}(t-\Delta t_{j12})-\phi_{jlo}\}-2\pi\nu_{jif}(\tau_j+\Delta t_{j23}), \tag{1.26}$$

$$\phi_{c3} = \pm\{2\pi\nu_{sky}(t-\tau_g-\Delta t_{c12})-\phi_{clo}\}-2\pi\nu_{cif}(\tau_c+\Delta t_{c23}). \tag{1.27}$$

In this case Δt_{j23} and Δt_{c23} represent the propagation times, excluding the delay lines which introduce additional delays τ_j and τ_c. As the two signals enter the correlator they have

$$\phi_{j4} = \pm\{2\pi\nu_{sky}(t-\Delta t_{j12})-\phi_{jlo}\}-2\pi\nu_{jif}(\tau_j+\Delta t_{j23}+\Delta t_{34})+\phi_{dds}, \tag{1.28}$$

$$\phi_{c4} = \pm\{2\pi\nu_{sky}(t-\tau_g-\Delta t_{c12})-\phi_{clo}\}-2\pi\nu_{cif}(\tau_c+\Delta t_{c23}+\Delta t_{34}). \tag{1.29}$$

The output of the correlator depends on the difference between the phases and should not be dependent on time over short periods.

$$\begin{aligned}\phi_{j4}-\phi_{c4} &= \pm 2\pi\nu_{sky}(\tau_g+\Delta t_{c12}-\Delta t_{j12})\pm(\phi_{clo}-\phi_{jlo})+2\pi(\nu_{cif}(\tau_c+\Delta t_{c24})\\ &\quad -\nu_{jif}(\tau_j+\Delta t_{j24}))+\phi_{dds} \qquad (1.30)\\ &= \pm 2\pi\nu_{sky}\tau_g \mp (sb)\phi_1+2\pi(\nu_{cif}\tau_c-\nu_{jif}\tau_j)+\phi_{dds}+\phi_{const}. \qquad (1.31)\end{aligned}$$

Terms involving a Δt have been combined into the term $\phi_{\rm const}$. Now we can express frequencies and delays as offsets from their central values:

$$\nu_{\rm sky} = \nu_{\rm sky0} + \Delta\nu, \quad (1.32)$$

$$\tau_{\rm g} = \tau_{\rm g0} + \Delta\tau_{\rm g}, \quad (1.33)$$

$$\nu_{\rm jif} = \nu_{\rm jif0} \pm \Delta\nu, \quad (1.34)$$

$$\nu_{\rm cif} = \nu_{\rm cif0} \pm \Delta\nu. \quad (1.35)$$

The plus and minus again refer to signals in the upper and lower sidebands respectively. $\nu_{\rm jif0}$ and $\nu_{\rm cif0}$ are the fixed centres of the IF bands for JCMT and CSO. $\nu_{\rm sky0}$ depends on the sideband under consideration. Substituting these values into equation 1.31 gives

$$\begin{aligned}\phi_{\rm j4} - \phi_{\rm c4} &= \pm 2\pi(\nu_{\rm sky0} + \Delta\nu)(\tau_{\rm g0} + \Delta\tau_{\rm g}) \mp (sb)\phi_1 \\ &\quad + 2\pi\{(\nu_{\rm cif0} \pm \Delta\nu)\tau_{\rm c} - \nu_{\rm jif0} \pm \Delta\nu)\tau_{\rm j}\} + \phi_{\rm dds} + \phi_{\rm const} \quad (1.36)\\ &= \pm 2\pi\nu_{\rm sky0}\Delta\tau_{\rm g} \pm 2\pi\nu_{\rm sky0}\tau_{\rm g0} \pm 2\pi\Delta\nu(\tau_{\rm g0} + \tau_{\rm c} - \tau_{\rm j}) \pm 2\pi\Delta\nu\Delta\tau_{\rm g} \\ &\quad \mp (sb)\phi_1 + 2\pi(\nu_{\rm cif0}\tau_{\rm c} - \nu_{\rm jif0}\tau_{\rm j}) + \phi_{\rm dds} + \phi_{\rm const}. \quad (1.37)\end{aligned}$$

We now choose

$$\phi_{\rm dds} = \phi_1 - (sb)2\pi\nu_{\rm sky0}\tau_{\rm g0}, \quad (1.38)$$

with the $\nu_{\rm sky0}$ appropriate to the sideband that we wish to observe. If $sb = +1$, then for signals down-converted to the IF in the upper sideband (the plus signs)

$$\begin{aligned}\phi_{\rm j4} - \phi_{\rm c4} &= +2\pi\nu_{\rm sky0}\Delta\tau_{\rm g} + 2\pi\Delta\nu\tau_{\rm err} + 2\pi\Delta\nu\Delta\tau_{\rm g} \\ &\quad + 2\pi(\nu_{\rm cif0}\tau_{\rm c} - \nu_{\rm jif0}\tau_{\rm j}) + \phi_{\rm const}, \quad (1.39)\end{aligned}$$

where the delay error $\tau_{\rm err} = \tau_{\rm g0} + \tau_{\rm c} - \tau_{\rm j}$. The first term in equation 1.39 gives rise to the normal fringe response (cf. equation 1.12). The second term is zero if there is exact delay compensation; if not, there is a phase component proportional to the frequency offset from the centre of the band, an aspect explored further in chapter 2. The third term sets a limit to the field of view of the interferometer (in practice the primary beams of the antennas are more restrictive). It arises because we can only stop the fringe exactly for one particular sky frequency, in this case for $\nu_{\rm sky0}$. For our system, $\Delta\nu \leq 0.5$ GHz and $\Delta\tau_{\rm g} \simeq 27$ ps for a position on the sky that is 10 arcseconds away from the centre of the field; for a continuum source in such a position there is a phase variation of about 10° from one side of the 1-GHz band to the other. The fourth term changes as the delay lines are switched. The lengths of the delay lines should be known accurately to compensate for this. The final term, although denoted as a constant, changes with the observing frequency, including the updates to account for the changing radial velocity of the source with respect to the antennas.

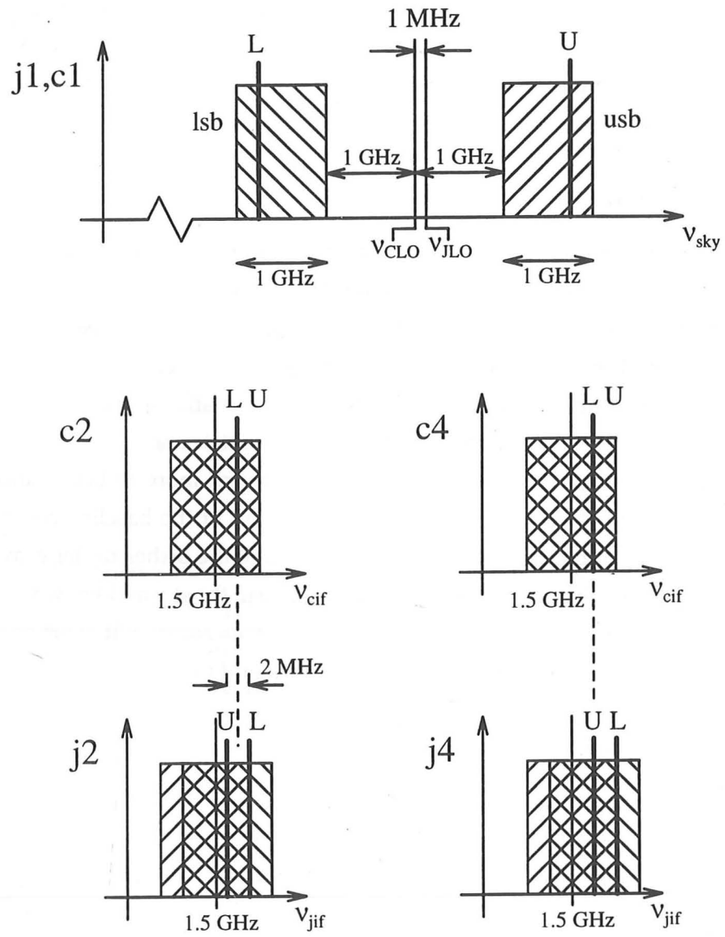


Figure 1.10: The relationship between the upper and lower sidebands, illustrating how one of them—in this case the upper sideband—is rejected by shifting the JCMT local oscillator by 1 MHz compared to the CSO, and then later subtracting it from the IF signal. The shaded regions show the passband, both on the sky (top) and in the IF (lower four diagrams). The two sky frequencies that give rise to a particular IF on the CSO side are also included. Note that the 1-MHz shift is not drawn to scale. The labels J1, C1, etc. are explained in the main text.

The phase difference for signals in the lower sideband when $sb = -1$ has a form identical to equation 1.39, except that the first three terms have minus signs. For signals in the upper sideband when $sb = -1$ or in the lower sideband when $sb = +1$, there is an extra phase term of $2\phi_1$, corresponding to a frequency of 2 MHz. Time-averaging in the correlator ensures that the signal in these two cases does not make a contribution. By choosing $sb = +1$ we reject the lower sideband; $sb = -1$ rejects the upper sideband (see Figure 1.10).

1.5.2 Calibration

There are two major aspects to the calibration of an interferometer: understanding the phase response as a function of time, and dividing out the effects of the complex spectral passband of the system. The latter is relatively straightforward; some parts of the passband experience more attenuation than others, requiring an appropriate frequency-dependent scaling of the amplitudes for correction. Any phase variations across the passband that are due to the instrument can be removed by simple subtraction.

Understanding the phase vs. time response is much more difficult, and requires a detailed knowledge of, for example, the precise length of the baseline components, the thermal behaviour of the cabling and the distortions of the dishes under gravity. Ideally, we should be able to observe a point source anywhere in the sky and know what the phase of the correlator output will be. The current reality falls somewhat short of this goal; in chapter 3 I describe how I have approached this task and the progress made.

1.5.3 Data reduction

The length of an individual integration with the JCMT–CSO Interferometer is kept to the relatively short value of 10 seconds to avoid decorrelation effects in the correlator (see chapter 2). The data-reduction software has several stages, the most important being:

1. Correct each 10 s integration for the second, fourth and fifth terms in equation 1.39.
2. Correct each integration for the known phase contributions of the instrument, for example the baseline errors and dish deformation.
3. Correct each integration for the complex passband and convert to units of effective antenna temperature.
4. Add a number of 10 s integrations together.
5. Average over specified frequency intervals if required.
6. Display the resulting complex spectrum.

Chapter 4 describes the reduction process and includes examples of the different sources that we observed.

1.6 Expected performance of the interferometer

In this section I first calculate the sensitivity and spatial resolution expected from the JCMT–CSO Interferometer, before comparing it to existing single-dish telescopes and other interferometers. Subsequently, I discuss some of the astronomy that should be possible with our unique combination of wavelength and resolution.

1.6.1 Sensitivity

The sensitivity, σ, is defined as the flux of a point source needed to produce a signal-to-noise ratio of unity. To save space and algebra, I will simply quote the result for the sensitivity of an interferometer that is derived in chapter 6 of Thomson, Moran and Swenson (1986). For a coherent integration of length $t_{\rm int}$ over a bandwidth $\Delta\nu$,

$$\sigma = \frac{\sqrt{2}kT_{\rm sys}}{\eta A_{\rm e}(t_{\rm int}\Delta\nu)^{0.5}}, \tag{1.40}$$

where k is the Boltzmann constant, $T_{\rm sys}$ is the system temperature (see chapter 3) and $A_{\rm e}$ is the geometrical mean of the surface areas of the two dishes. For the 15-m diameter JCMT and the 10-m CSO, $A_{\rm e} = 120\ {\rm m}^2$. The overall efficiency of the system is given by η, which should be ~50%. Equation 1.40 assumes that we know the phase of the source. This will be true for a perfectly calibrated instrument, or if we have other information that can be used as a phase reference, such as a bright spectral feature. If this is not the case, then the ability to detect signal in the noise is reduced and we must include another factor of $\sqrt{2}$ in the numerator.

Inserting appropriate values into equation 1.40 we obtain

$$\sigma_{\rm cont} = 26\ {\rm mJy}\left(\frac{\eta}{0.5}\right)^{-1}\left(\frac{t_{\rm int}}{1000\ {\rm s}}\right)^{-0.5}\left(\frac{\Delta\nu}{1\ {\rm GHz}}\right)^{-0.5}\left(\frac{T_{\rm sys}}{800\ {\rm K}}\right) \tag{1.41}$$

for a continuum source, and

$$\sigma_{\rm line} = 770\ {\rm mJy}\left(\frac{\eta}{0.5}\right)^{-1}\left(\frac{t_{\rm int}}{1000\ {\rm s}}\right)^{-0.5}\left(\frac{\Delta v}{1\ {\rm kms}^{-1}}\right)^{-0.5}\left(\frac{\nu_{\rm sky}}{345\ {\rm GHz}}\right)^{-0.5}\left(\frac{T_{\rm sys}}{800\ {\rm K}}\right) \tag{1.42}$$

for a narrow line feature.

1.6.2 Spatial resolution

The baseline of the JCMT–CSO can be described by three components. The component along the Earth's axis has length 43 m; the east-west separation is 158 m and the other

orthogonal component, pointing away from the Earth and towards the celestial equator, is almost zero. The total length is about 164 m, giving a minimum fringe-spacing of

$$\Delta\theta_{\rm min} = 1.1'' \left(\frac{\nu_{\rm sky}}{345\,{\rm GHz}} \right)^{-1}. \tag{1.43}$$

With only one fixed baseline, there is no meaningful imaging ability.

We can also estimate the temperature of the weakest thermal continuum source that can be detected in a given time at the maximum resolution. The limiting flux for a 1000 s integration is given by equation 1.41 as 26 mJy. An object much larger than $\Delta\theta_{\rm min}/2$ will be substantially resolved, so the following calculation is based on a source of size $0.5'' \times 0.5''$. In the Rayleigh–Jeans limit, the flux S_ν of a black-body source at frequency ν, subtending solid angle $\Delta\Omega$, is given by

$$S_\nu = \frac{2kT_{\rm b}\nu^2\Delta\Omega}{c^2}. \tag{1.44}$$

The brightness temperature, $T_{\rm b}$, must be at least 1.2 K to give a flux exceeding 26 mJy at 345 GHz from an object of the specified size. To detect a spectral line at a resolution of 1 km s^{-1} requires a line temperature of at least 35 K.

1.6.3 Comparison to other instruments

In Table 1.2, I compare the main instruments operating at millimetre and submillimetre wavelengths. Although the JCMT–CSO Interferometer has only one fixed baseline, it is the only interferometer operating at submillimetre wavelengths at the present time. The Harvard–Smithsonian Center for Astrophysics is designing a 6-element submillimetre array (to be known as the SMA) that will be sited on the summit of Mauna Kea, close to the JCMT. The array, comprising six dishes of diameter 6 m, will only have the same total collecting area as the single dish of the JCMT, so it is envisaged that the JCMT and/or the CSO will be connected to the array when high sensitivity is needed.

1.7 Astronomy with a single-baseline interferometer

How can a single, fixed-baseline interferometer be used to make a useful contribution to astronomy? From the calculations in section 1.6.2, it is apparent that sources for the JCMT–CSO Interferometer should be compact and fairly bright.

1.7.1 Protostars

One of the prime motivations for building the interferometer was to study protostars—young stars still in the process of forming from their parent molecular clouds. Most of the emission from millimetre to infrared wavelengths comes from warm dust, heated initially

Instrument	# dishes	dish diameter	wavelength range	Max. E-W baseline	Min. E-W fringe-spacing
JCMT–CSO	2	15 m/10 m	0.8 to 1.3 mm	158 m	1.1″ @ 0.87 mm
OVRO	5 (+1)	10 m	1.3 to 9 mm	200 m	1.3″ @ 1.3 mm
BIMA	6 (+3)	6 m	2.7 to 3.4 mm	300 m	1.8″ @ 2.7 mm
NMA	6	10 m	2.0 to 3.4 mm	560 m	0.7″ @ 2.0 mm
IRAM	4	15 m	2.7 to 3.4 mm	288 m	1.9″ @ 2.7 mm
JCMT	1	15 m	0.3 to 1.3 mm	—	(14″ @ 0.87 mm)
SEST	1	15 m	0.8 to 3.4 mm	—	(14″ @ 0.87 mm)
IRAM 30 m	1	30 m	0.8 to 7 mm	—	(7″ @ 0.87 mm)

Table 1.2: A comparison of the JCMT–CSO Interferometer with other instruments. (OVRO: Owens Valley Radio Observatory Millimeter Array, California; BIMA: Berkeley–Illinois–Maryland–Association Array, California; NMA: Nobeyama Millimeter Array, Japan; IRAM: Millimeter interferometer, Plateau de Bure, France; SEST: Swedish–ESO Submillimeter Telescope, Chile; IRAM 30 m: Pico Veleta, Spain.)

by the collapse of material onto a dense core, and then subsequently by the onset of nuclear burning. In the Rayleigh–Jeans regime, the flux increases rapidly with frequency, owing to both the quadratic rise in the Planck function and the increase in the dust emissivity. Much of the molecular cloud in which a source is embedded is optically thick and impenetrable to wavelengths shorter than ~0.1 mm. The submillimetre wave-band is therefore ideal for observing the events taking place in the heart of newly-forming stars, and interferometry delivers the resolution needed to resolve the important structures.

Chapter 6 describes the observations that we made of accretion discs in two well-known protostars. The planets of our solar system condensed from such a disc around the early sun, and for the first time we have been able to resolve these structures easily and determine their sizes and orientations.

The interferometer is also well-suited to searching for the signature of infalling material around young stars. A convincing observation has proved elusive; much of the evidence currently available is based on the interpretation of the detailed shapes of molecular lines. With the interferometer we can look for red-shifted absorption lines against the warm dust continuum emission from the core. Figure 1.11 illustrates this; a red-shifted absorption line can only result from infalling material along the line of sight to the core. A single-dish telescope can be used in principle, but in practice the absorption lines are 'filled in' by emission from more extended regions of gas at the same velocity. It is the interferometer's

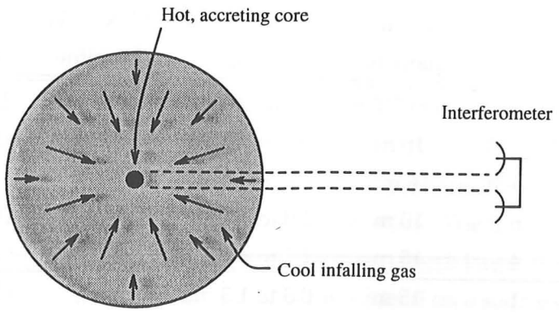


Figure 1.11: The geometry giving rise to red-shifted absorption lines from protostars. The interferometer resolves out the extended emission.

insensitivity to this extended emission that makes it ideal for this type of observation.

We observed one protostellar source, IRAS 05338, to search for red-shifted absorption at the HCN (4–3) transition. There was no obvious evidence for an absorption line in this source, although this may be due in part to the fact that the local oscillators were not in a 'doppler tracking' mode to account for the changing radial velocity of the source with respect to the Earth-bound telescopes. Further observations will be made.

1.7.2 Masers and evolved stars

Masers (Microwave Amplification by the Stimulated Emission of Radiation) are obvious candidates for observation, since they are very bright and very compact. They are commonly associated with old, evolved stars whose outer envelopes are being blown off in a wind. There are several strong maser transitions in the submillimetre spectrum, all only recently discovered. Chapter 5 is a detailed study of VY Canis Majoris, one of the strangest and most well-studied evolved stars. The observations made with the interferometer revealed some unexpected features. I show how these can be accounted for by a simple model that also helps to explain much of the anomalous behaviour associated with the source.

1.7.3 Galactic Nuclei

Galactic Nuclei also fulfil our requirement for bright and compact sources. The nearest of these is about 8 kpc away, at the centre of our own galaxy. The central source, Sagittarius A*, is suspected to be a black hole with a mass of about 10^6 $M_\odot$. There has been much debate as to whether the strong continuum emission from this source at millimetre and

submillimetre wavelengths is due to warm dust, perhaps in an accretion disc, or to a non-thermal synchrotron source. In chapter 4 I present our observations of Sgr A* and show how they have helped to resolve this issue. In the same chapter I also show spectra of the nearest known radio-galaxy, Centaurus A. We detected narrow, red-shifted absorption lines against the bright, nuclear continuum source, in a manner analogous to that of Figure 1.11. The absorption must arise from molecular clouds that are moving towards the nucleus, where there is thought to be a large black hole.

1.8 A brief history of the project

We faced a multitude of challenging problems in the course of getting the JCMT–CSO Interferometer to produce useful results. Most interferometers are designed and built to be dedicated full-time to interferometry; debugging, calibration and observing can be carried out systematically over long periods of time. Both the JCMT and the CSO are used almost exclusively for single-dish observing, and the time available to us has so far been limited to just 24 nights, spread over four observing runs. That we have already been able to produce good scientific results in such a short period is a tribute to the dedicated and expert support that we have received.

1.8.1 Design and general construction

The potential for interferometry with the JCMT and CSO telescopes was appreciated some time before either of them was built. A detailed discussion of the options that were available is presented in the report compiled by a panel for the JCMT Board (1990). Most of the basic design, including the use of optical fibre for the signal transmission and delay lines and DDSs for fringe-tracking and sideband rejection, was formulated by Richard Hills and John Carlstrom. Some preparatory work was completed in early 1991, but it was not until October 1991 that final approval was obtained and the detailed design and construction could begin in earnest. I was initially given the job of designing and building the delay line system using optical fibre, which I completed at MRAO with help from the Optoelectronics group of the Cambridge University Engineering Department.

Interferometry was allocated 3 nights of telescope time at the start of June 1992. Much of the hardware came together in a two-week spell of frantic activity at Caltech in the last half of May. The phase-lock system for the local oscillators was built and tested on the bench and appeared to perform well. I learnt to program in C and spent most of my time interfacing the microcomputer to the DDS oscillators. The next chapter describes a large part of my work up until this point.

1.8.2 Run 1: June 1992

In the few days preceding our allotted time we installed the new instrumentation on the telescopes and carried out some basic tests. There were problems with the micro's real-time clock and both the digital input and output cards used for interfacing. The baseline was surveyed by a group from the University of Hawaii and a submillimetre test transmitter was built and located at the UKIRT site, some 800 m away on a ridge overlooking the JCMT and the CSO.

Not surprisingly, our first three nights of telescope time were plagued with problems. It was a major challenge to get the receivers, the IF system, the correlator and at least three computers working together in harmony. We observed several bright quasars and the water maser U Her in the 'B' band, but failed to detect any of them consistently. There was, however, good evidence in the data that we were detecting U Her, and we suspected a bug in the software responsible for tracking the fringe. This was eventually found and corrected. The errors in the baseline were also substantial at this stage, and contributed to the almost complete decorrelation that we were experiencing.

The data-reduction software for this first run was written by Chris Mayer. Although initially this was perfectly adequate, our demands soon began to outstrip its capabilities. I spent a large fraction of the time before the next run developing a more flexible, command-driven program, and made extensive improvements to the microcomputer program.

1.8.3 Run 2: August 1992

We received another four nights of time not long after the first run. The weather was not ideal and we were restricted to operation in the low frequency 'A' band. This time we were able to detect most bright sources, although the output of the correlator drifted rapidly with time because of the poor calibration. We observed quasars and the unique hydrogen recombination maser source MWC 349.

These observations provided the data required for a good calibration of the system, particularly the length of the baseline. I wrote software to fit a baseline error to the variation of phase and succeeded in pinning it down to the nearest centimetre or so. This work is presented as part of chapter 3.

The clock on the microcomputer was still causing trouble. The installation of a communications link between the JCMT VAX and the micro solved this problem and simplified the system in general. This link was ready for the next outing to Hawaii.

1.8.4 Run 3: March 1993

With seven nights on the telescopes we were able to make more progress and obtain useful astronomical results. We spent the first four nights working in the 'A' band, one of which was lost when an Ortel laser diode (electrical-to-optical transducer) went 'soft'. A replacement was shipped out from Owens Valley and arrived the next day. The weather improved for the remaining three nights and we observed in the higher frequency 'B' band. The calibration was still far from ideal and we were only able to integrate coherently on bright objects.

The principal sources were Sagittarius A*, Centaurus A, and VY CMa, all of which proved very interesting. I describe the first two in chapter 4 as examples of the data-reduction process. The whole of chapter 5 is devoted to VY CMa. We had been taking measurements with hot loads for the first time, allowing the correlated output to be scaled to effective antenna temperature. In the subsequent data reduction I discovered that the sensitivity of the interferometer was substantially worse than that predicted by equation 1.41—we seemed to be losing about 50% of the signal—and it was not until near the end of the next run that the reason became clear (see next section). There also appeared to be jumps in the phase as a source was tracked across the sky. The transmission of the 640-MHz signal from the CSO to the JCMT was suspected to be responsible for these jumps, and they hampered my attempts to make further refinements on the baseline solution. I also made substantial changes to the reduction software.

1.8.5 Run 4: November 1993

This has been our longest period of observing to date, with ten nights on the two telescopes. Most of the system worked well and we were fortunate to have good weather. We were frustrated by receiver problems over the first two nights, but once over this we spent the majority of the available time observing protostars, in an attempt to resolve their suspected accretion discs. The results from this work are presented in chapter 6. We observed VY CMa once more, as well as the proto-planetary nebula CRL 618.

The poor sensitivity was traced to an error with the 180° phase-switching in the correlator (see chapter 2). This was corrected, but the resulting sensitivity was still not up to the expected level. The 10-MHz signal was confirmed as the cause of the phase jumps and a temporary arrangement was set up to circumvent the problem. We also gained a better understanding of the phase vs. time calibration by including terms to account for the non-intersection of the elevation and azimuth axes at each telescope, and a term to account for the refractive index of the atmosphere (chapter 3).

1.8.6 Future observations

The interferometer has now been made available to the general astronomical community. The instrument is not intended for common-user operation, so all observations will be carried out as a collaboration between the 'consortium' of people who built it and the observers who have made successful proposals. The first such period of observing is scheduled for October 1994.

During this run we intend to make the first observations in the 'C' band, where the higher thermal fluxes and resolution should prove especially useful for the protostellar work. Our improved sensitivity will make observations much more efficient: doubling the sensitivity reduces the observing time required by a factor of four. The better phase calibration will lead to longer periods of coherent integration, and we will no longer be restricted to relatively bright sources. In the longer term, we look forward to collaborating with the Harvard–Smithsonian Submillimeter Array, due to become operational in about four years' time.

There are many valuable observations to be made with the interferometer. A survey of protostellar systems could be very important in establishing the stages in the evolution of young stars, including the relative incidence and formation times of accretion discs and multiple systems. We may also extend the observations to include medium-mass protostars. Both evolved stars and extra-galactic sources have proved to be fruitful areas of study and this work also looks set to continue.

1.9 Summary

This introduction has covered the basic theory of interferometry to a level needed to understand the design of the JCMT–CSO submillimetre interferometer. I have calculated the expected performance of our instrument, compared it against others and outlined the astronomy that is possible. Finally I described the stages that the project has been through and the future prospects.

Chapters 2, 3 and 4 elaborate on some of the major technical aspects of the project. This is followed by an analysis of the data for the evolved star VY CMa in chapter 5, and for protostellar discs in chapter 6. I briefly summarize and conclude in chapter 7.

Chapter 2

TRACKING THE FRINGE

2.1 Introduction

The travel time for signals from a source in space to the correlator should be independent of the antenna at which they are collected. This introduces the need for compensating delay, as described in the previous chapter. Ideally this should be continuously variable, so that the geometrical delay can be equalized for any position on the sky.

Practical delay systems are of three basic types:

- A continuously variable, 'trombone'-style system; this is used on a few optical interferometers—COAST[1], SUSI[2] and the future Keck I–Keck II, for example—and is implemented with mirrors mounted on movable trolleys.
- Random Access Memory (RAM); RAM-based systems involve digitizing the signals for each antenna and reading them into memory. Each is then retrieved at a different time, to compensate for the geometrical delay, and correlated. The Ryle Telescope uses this technique and the RAM is an integral part of the correlator.
- A discrete set of delay lines; described in more detail below.

The last two systems tend to be used more at longer wavelengths. When correlation is required over large bandwidths the electronics must be very fast, and the use of switchable delay lines becomes more attractive than RAM. Different lengths of cable are switched into the signal paths to provide delay compensation. The smallest step size needed is dictated by the bandwidth of the correlation $\Delta\nu$ and the signal loss that can be tolerated (see §2.3). For example, to correlate signals from a single sideband over a bandwidth of 1 GHz with less than 1% requires delay lines with a resolution of ~100 ps, or ~30 mm of electrical

[1]Cambridge Optical Aperture Synthesis Telescope

[2]Sydney University Stellar Interferometer

length. A large number of steps is needed to span the range of possible geometrical delays (± baseline length) with this resolution. A discrete delay line system has recently been installed on the Owens Valley Millimeter Array (Soares and Padin, 1994). The delays are arranged in a binary sequence, with the longest steps in optical fibre and the shortest in rapidly switchable strip-line circuits.

Spectral line correlators determine the cross-correlation function of two signals, which involves correlating them over a range of relative delays. We use the Dutch Auto-correlation Spectrometer (DAS) to correlate the signals from the interferometer. This is normally used for single-dish spectral line work on the JCMT, but was designed with the ability to perform cross-correlation. This is described in §2.2. Such a correlator greatly reduces the resolution of the delay lines that is needed, for reasons set out in §2.3.

A major feature of the project was to use optical fibres to deliver the IF signals from the two receivers, via the delay lines, to the correlator, and to carry the phase-lock reference frequencies from the CSO to the JCMT. The benefits and drawbacks of choosing optical fibre over the more traditional coaxial cable are discussed in §2.4, followed in §2.5 by my design for implementing the delay lines in optical fibre.

The doppler shift between the signals arriving at the two antennas (§1.2.4) must also be corrected. In §2.6 I describe how we use a pair of Direct Digital Synthesizers (DDSs) to achieve this, and how in addition they are are used for phase-switching and sideband rejection.

The delay lines and the DDSs are under the control of a microcomputer, for which I did all the interfacing and wrote the software. The timing is critical; §2.7 and §2.8 relate how I approached the problem and describe the sequence of events during an integration.

2.2 The correlator

The DAS was designed and built by a Dutch team led by Albert Bos. It is central to the whole operation of the interferometer and deserves some description.

2.2.1 Theory

An auto-correlation spectrometer measures the frequency spectrum $F(\nu)$ of a time-varying signal $V_1(t)$ by computing the auto-correlation function $G_{11}(\tau)$:

$$G_{11}(\tau) = \langle V_1(t) V_1(\tau - t) \rangle, \tag{2.1}$$

where τ is a time delay. From the Wiener–Khintchine theorem, the Fourier transform of $G_{11}(\tau)$ gives the power spectrum $F(\nu)$,

$$G_{11}(\tau) \overset{F-T}{\Longleftrightarrow} F(\nu). \tag{2.2}$$

$G_{11}(\tau)$ is real and symmetric, so $F(\nu)$ must itself be real. Since $G_{11}(\tau)$ is symmetric, it is only necessary to measure it over positive τ. If we have N measurements of G_{11}, equally spaced in τ, then a Fast Fourier Transform algorithm will give the power at N evenly spaced frequencies.

For interferometry, we require the cross-correlation function of two signals $V_1(t)$ and $V_2(t)$, given by

$$G_{12}(\tau) = \langle V_1(t) V_2(\tau - t) \rangle. \tag{2.3}$$

$G_{12}(\tau)$ is no longer symmetric and the Fourier transform gives a complex power spectrum,

$$G_{12}(\tau) \overset{F-T}{\Longleftrightarrow} \tilde{F}(\nu). \tag{2.4}$$

At each frequency we now have an amplitude and a phase. The interpretation of a complex power was introduced in §1.2.6, where it was shown that the phase is a measure of position on the sky. Since it lacks symmetry, G_{12} must be measured over both positive and negative values of τ; N measurements of G_{12} will transform to give $N/2$ amplitudes and phases in the frequency domain (note that the total number of independent values is conserved by the Fast Fourier Transform).

2.2.2 The DAS in practice

A detailed description of the DAS can be be found in a series of technical reports (Bos, 1988; Bos, 1986a; Bos, 1986b; Bos, 1986c). Figure 2.1 illustrates in a simplified and schematic form how the DAS operates. We follow the path of a signal from the time that it enters the correlator:

1) Filtering into subbands: The correlator can cope with input bandwidths of up to 1 GHz, centred on 1.5 GHz. Electronic technology cannot currently operate fast enough to process the whole 1 GHz as a unit, so the signal is broken up into up to eight chunks, or subbands, each with a bandwidth of 160 MHz. Adjacent subbands overlap by 35 MHz. Local oscillators LO1, LO2, etc. are used to down-convert the full input signal and the low-pass filters pass bandwidths of 160 MHz. The LOs must be coherent between the two IF arms, and in reality the down-mixing is achieved in two stages. The choice of ν_{LO1}, ν_{LO2}, etc. defines the correlator configuration and is discussed in §2.2.3. For example, if $\nu_{\mathrm{LO1}} = 1.500$ GHz then IF signals with $1.500\,\mathrm{GHz} \leq \nu_{\mathrm{IF}} \leq 1.660\,\mathrm{GHz}$ are selected.

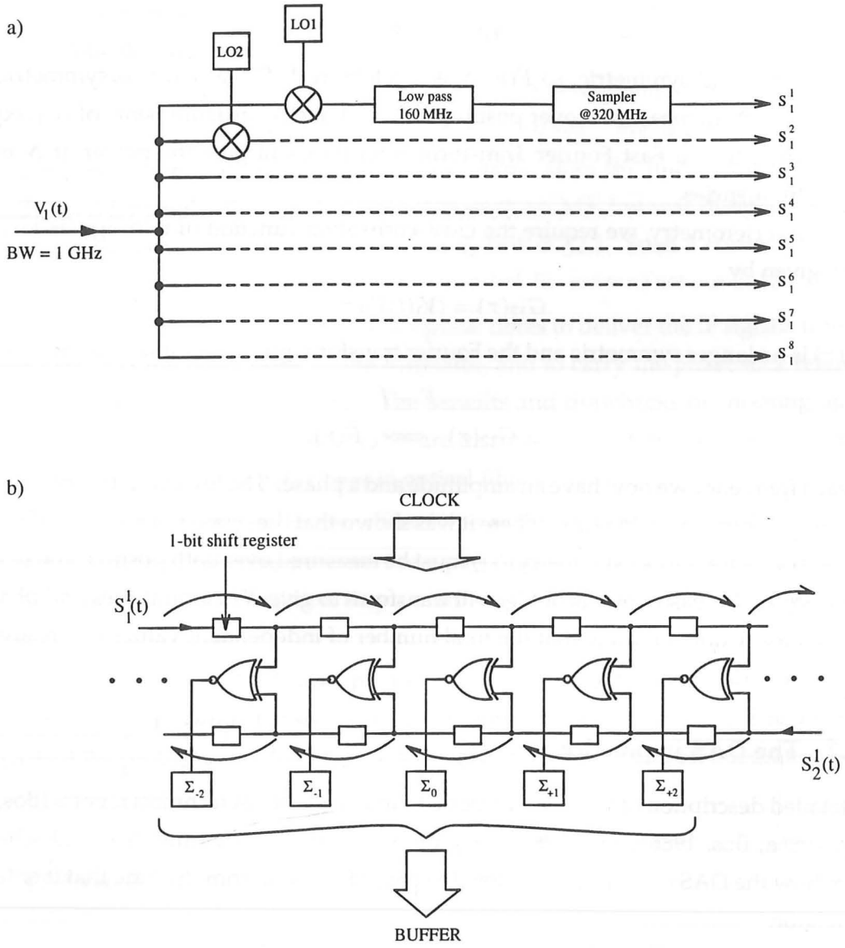


Figure 2.1: Basic principles involved in the operation of the DAS. (a) Each of the two IF input signals is split into subbands that are digitized to give the outputs S_1^1, etc. Only one IF is shown. (b) A highly schematic illustration of correlation; five lags are shown and each signal only has two levels, 1 and 0. At each clock cycle the signals advance to the next node in the directions indicated. See main text for the truth table of the logic gates.

2) Sampling: An analogue-to-digital converter samples each 160-MHz subband at the Nyquist rate[3] of 320 MHz. The DAS actually samples with 2-bit resolution (4 levels) but for clarity I describe the correlation process as though the sampling was 1-bit (=1 for positive signal; =0 for negative signal). When performing a cross-correlation of the two analogue input signals, V_1 and V_2, there will be up to 16 digital signals. I have labelled them S_1^1 for the first subband of V_1, S_2^1 for the first subband of V_2, etc. S_1^1 will now be correlated with S_2^1, S_1^2 with S_2^2, and so on.

3) On-chip correlation: Figure 2.1b shows schematically how two 1-bit signals are correlated. S_1^1 and S_2^1 are a series of 1s and 0s, with one digit at each node. The digits facing one another at each pair of nodes are correlated every clock cycle by performing a simple logical operation. The truth table is:

	0	1
0	1	0
1	0	1

The result is added to the value stored in the appropriate accumulator (Σ_1, Σ_2, etc.). At the next clock cycle the digits are all moved along by one node; the digits belonging to S_1^1 are shifted to the right and S_2^1 to the left. The above process is then repeated and the values stored in the accumulators increase. If the input signals are uncorrelated, then after N clock cycles the accumulators will on average each hold the value $N/2$. If the signals are correlated the number is greater than this; if they are anti-correlated the number will be less. If the clock period is T, then the correlator measures the cross-correlation function at intervals of $\Delta\tau = 2T$.

The DAS is substantially more complex than this simple picture. Firstly, it employs a parallel architecture so that it can process the inputs (which were sampled at 320 MHz) with a basic clock rate of 40 MHz. In addition, the 2-levels of amplitude resolution, which increase the correlator efficiency from 64% to 87%, require a larger truth table. The weightings used in the DAS are:

	0	1	2	3
0	6	4	2	0
1	4	3	3	2
2	2	3	3	4
3	0	2	4	6

[3] in the DAS there are two 160 MHz samplers taking interleaved samples for each subband, so that the effective sampling rate is 320 MHz

All of this occurs on one chip, each of which can perform 16 correlations. With 128 chips, the DAS can generate up to 2048 correlation values at different values of τ. The on-chip accumulators that hold the values become full after 13.1 ms of integration.

4) Transfer to buffers: To prevent overflow, the numbers contained in the accumulators are off-loaded into much larger memory buffers at regular intervals. This interval, known as a correlation cycle, can be set by the user but must be shorter than 13.1 ms. The correlator issues a 'correlator ready' (CRDY) signal at the end of each correlation cycle. This 6.4-μs long pulse is used for timing by the interferometry microcomputer.

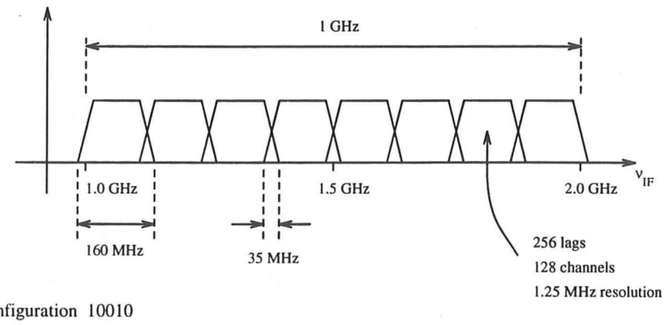
There are four memory buffers, one pair for each of the two phase-switchings (or on/off-source for single-dish work). One buffer of each pair accumulates the current data in a correlation cycle, while the other, containing the previous cycle's data, is read out to one of the JCMT VAX computers. The rôles are then swapped, so that integration is not interrupted for data read-out. The S/R input line to the DAS controls whether data are accumulated into one of the in-phase buffers (Signal) or one of the anti-phase buffers (Reference), each of which can be used for up to 53.6 s. The S/R line is toggled regularly for phase-switching.

5) Fourier transformation: The VAX that controls the DAS performs the Fast Fourier Transform on the lag data to give the power spectrum. There is the option of incorporating a tapering function in the lag domain (this is effectively the same as applying Hanning smoothing in the frequency domain), but the resolution is degraded slightly and we do not use it.

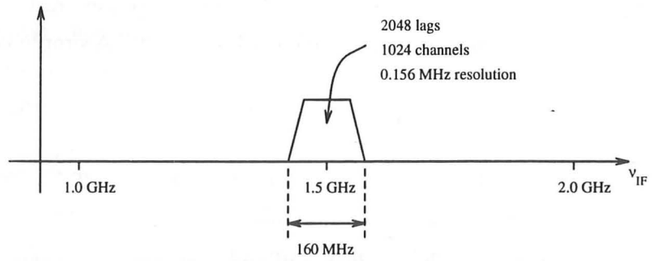
2.2.3 Correlator configurations

The correlator is very flexible, and different configurations can be realised by choosing the frequencies $\nu_{LO1}, \nu_{LO2}, \ldots \nu_{LO8}$ and the connections between the correlator chips. Figure 2.2 shows examples of the configurations that we have used. The total number of lags available is fixed at 2048, but they can be used in various combinations so that the interferometer always has 1024 frequency channels in total, ignoring overlaps. Configuration 10030 (the names are rather arbitrary) uses all eight subbands, each of which has 256 lags. The resulting spectrum covers the full 1 GHz of bandwidth with a resolution of 1.25 MHz. Configuration 10010 devotes all of the available lags to just one subband, generating a high-resolution spectrum spanning 160 MHz at a resolution of 0.156 MHz. Configuration 10050 is a hybrid; 256 lags are used for each of four subbands spanning a total of 512 MHz at a resolution of 1.25 MHz. The other 1024 lags produce a higher resolution spectrum of the central 160 MHz. We used configuration 10050 quite extensively.

a) Configuration 10030



b) Configuration 10010



c) Configuration 10050

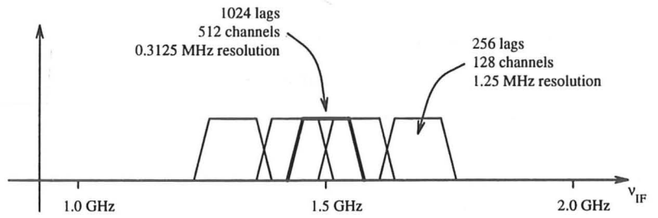


Figure 2.2: Three examples of correlator configurations in the DAS, showing the arrangement of the different subbands.

2.3 How many delay lines?

A correlator like the DAS, with its range of built-in delays, greatly reduces the resolution required for the external delay compensation system. It measures the cross-correlation function $G_{12}(\tau)$ over a finite range of τ, say $-\tau_{\text{max}}$ to $+\tau_{\text{max}}$ (the lag window). If the external delay compensation is exact, then G_{12} will be centred within the lag window. However, if we are using discrete delay lines, then in general there will be some delay error, τ_{err}. G_{12} will no longer be centred and there will be a phase-gradient across the frequency spectrum. The question is, how much delay error can we tolerate before the losses become significant?

2.3.1 A simple, approximate approach

The phase-gradient itself can be corrected by software, but some signal will be irretrievably lost. To estimate this loss, consider a channel in the frequency spectrum of width $\Delta\nu$ (much narrower than any spectral line in the signal) with a phase-gradient across it. If the signal strength in this channel is A when there is no delay error or phase-gradient, then when the delay error is introduced, the amplitude will be reduced to A''. A simple vector integration over $\Delta\nu$ gives

$$A'' = A\,\text{sinc}(\pi\Delta\nu\tau_{\text{err}}). \tag{2.5}$$

If we decide that we can tolerate a loss of amplitude of up to 1% in the frequency domain, then we can easily determine the maximum allowable τ_{err} for a given $\Delta\nu$, and therefore the resolution needed for the delay lines. For example, for $\Delta\nu = 1.25$ MHz we require $\tau_{\text{err}} < 62$ ns.

This analysis is valid for the old style of correlators that used large filter banks to divide the signals into small frequency bins, but I realised that a slightly different approach was needed to deal with a digital cross-correlator like the DAS, where the signals are correlated in the time domain before being transformed to frequency.

2.3.2 A more rigorous approach to delay error

To reiterate, the cross-correlation function of two signals is given by

$$G_{12}(\tau) = \langle V_1(t)V_2(\tau - t)\rangle, \tag{2.6}$$

where τ can range from $-\infty$ to $+\infty$. The correlator can only measure G_{12} within its available lag window $L(\tau)$. So

$$G'_{12}(\tau) = G_{12}(\tau)L(\tau). \tag{2.7}$$

Fourier transforming,

$$\tilde{F}'(\nu) = \tilde{F}(\nu) \star H(\nu), \tag{2.8}$$

where $\star$ denotes convolution. If no tapering is applied, then $L(\tau)$ is a top-hat function with unit amplitude in the range $-\tau_{\text{max}} \leq \tau \leq +\tau_{\text{max}}$. The Fourier transform of this function, $H(\nu)$, is a sinc function that sets the resolution of the spectrum $\tilde{F}'(\nu)$ at $(2\tau_{\text{max}})^{-1}$ (the separation of the peak and first zero of the sinc function).

When there is a delay error τ_{err}, a different cross-correlation function is measured;

$$G''_{12}(\tau) = G_{12}(\tau - \tau_{\text{err}})L(\tau). \tag{2.9}$$

This introduces a phase-gradient in the frequency domain;

$$\tilde{F}''(\nu, \tau_{\text{err}}) = \tilde{F}(\nu)e^{i2\pi\nu\tau_{\text{err}}} \star H(\nu). \tag{2.10}$$

Figure 2.3 illustrates the relationship between $\tilde{F}$, $\tilde{F}'$ and $\tilde{F}''$. To keep it simple, only the upper envelopes of the cross-correlation functions have been included, within which the function oscillates between positive and negative values.

Software is used to correct for the phase-gradient in $\tilde{F}''$, equivalent to shifting the correlation function back to the centre of the lag window, but as τ_{err} becomes larger, more signal is lost and the amplitude of $\tilde{F}''$ becomes significantly lower than $\tilde{F}'$. We can calculate the magnitude of this loss. For a Gaussian spectral line with centre frequency ν_0 and half-width σ, originating from a point source and with no noise,

$$\tilde{F}(\nu) = Ae^{-\frac{(\nu-\nu_0)^2}{2\sigma^2}}. \tag{2.11}$$

If there is a delay error of τ_{err} and the lag window is not tapered, then after convolution the new maximum amplitude of the line has

$$A''(\sigma, \tau_{\text{err}}) \propto \int_{-\infty}^{+\infty} Ae^{-\frac{(\nu-\nu_0)^2}{2\sigma^2}} e^{i2\pi\nu\tau_{\text{err}}} \text{sinc}\{2\pi\tau_{\text{max}}(\nu - \nu_0)\}d\nu. \tag{2.12}$$

When $\tau_{\text{err}} = 0$, the correlation function is centred and A'' is a maximum. The fractional loss when $\tau_{\text{err}} \neq 0$ can be defined by

$$\gamma(\sigma, \tau_{\text{err}}) = 1 - \frac{A''(\sigma, \tau_{\text{err}})}{A''(\sigma, 0)}. \tag{2.13}$$

Parseval's theorem relates the 'power' in one transform domain to the other. In our case,

$$\int_{-\infty}^{+\infty} |G_{12}(\tau)|^2 d\tau \propto \int_{-\infty}^{+\infty} |\tilde{F}(\nu)|^2 d\nu, \tag{2.14}$$

and similarly for G'_{12} and G''_{12}. This gives a useful insight into the loss of signal. If, in Figure 2.3, area is lost under the correlation function in the lag domain (which implies that we also lose area once the function is squared), then area is also lost under the spectral line. In other words, the amplitude of the line is reduced. It is easy to predict how the loss will vary with τ_{err} for the two extremes of σ:

2.3 How many delay lines?

A correlator like the DAS, with its range of built-in delays, greatly reduces the resolution required for the external delay compensation system. It measures the cross-correlation function $G_{12}(\tau)$ over a finite range of τ, say $-\tau_{\text{max}}$ to $+\tau_{\text{max}}$ (the lag window). If the external delay compensation is exact, then G_{12} will be centred within the lag window. However, if we are using discrete delay lines, then in general there will be some delay error, τ_{err}. G_{12} will no longer be centred and there will be a phase-gradient across the frequency spectrum. The question is, how much delay error can we tolerate before the losses become significant?

2.3.1 A simple, approximate approach

The phase-gradient itself can be corrected by software, but some signal will be irretrievably lost. To estimate this loss, consider a channel in the frequency spectrum of width $\Delta\nu$ (much narrower than any spectral line in the signal) with a phase-gradient across it. If the signal strength in this channel is A when there is no delay error or phase-gradient, then when the delay error is introduced, the amplitude will be reduced to A''. A simple vector integration over $\Delta\nu$ gives

$$A'' = A\,\text{sinc}(\pi\Delta\nu\tau_{\text{err}}). \tag{2.5}$$

If we decide that we can tolerate a loss of amplitude of up to 1% in the frequency domain, then we can easily determine the maximum allowable τ_{err} for a given $\Delta\nu$, and therefore the resolution needed for the delay lines. For example, for $\Delta\nu = 1.25$ MHz we require $\tau_{\text{err}} < 62$ ns.

This analysis is valid for the old style of correlators that used large filter banks to divide the signals into small frequency bins, but I realised that a slightly different approach was needed to deal with a digital cross-correlator like the DAS, where the signals are correlated in the time domain before being transformed to frequency.

2.3.2 A more rigorous approach to delay error

To reiterate, the cross-correlation function of two signals is given by

$$G_{12}(\tau) = \langle V_1(t)V_2(\tau - t)\rangle, \tag{2.6}$$

where τ can range from $-\infty$ to $+\infty$. The correlator can only measure G_{12} within its available lag window $L(\tau)$. So

$$G'_{12}(\tau) = G_{12}(\tau)L(\tau). \tag{2.7}$$

Fourier transforming,

$$\tilde{F}'(\nu) = \tilde{F}(\nu) \star H(\nu), \tag{2.8}$$

where $\star$ denotes convolution. If no tapering is applied, then $L(\tau)$ is a top-hat function with unit amplitude in the range $-\tau_{\text{max}} \leq \tau \leq +\tau_{\text{max}}$. The Fourier transform of this function, $H(\nu)$, is a sinc function that sets the resolution of the spectrum $\tilde{F}'(\nu)$ at $(2\tau_{\text{max}})^{-1}$ (the separation of the peak and first zero of the sinc function).

When there is a delay error τ_{err}, a different cross-correlation function is measured;

$$G''_{12}(\tau) = G_{12}(\tau - \tau_{\text{err}})L(\tau). \tag{2.9}$$

This introduces a phase-gradient in the frequency domain;

$$\tilde{F}''(\nu, \tau_{\text{err}}) = \tilde{F}(\nu)e^{i2\pi\nu\tau_{\text{err}}} \star H(\nu). \tag{2.10}$$

Figure 2.3 illustrates the relationship between $\tilde{F}$, $\tilde{F}'$ and $\tilde{F}''$. To keep it simple, only the upper envelopes of the cross-correlation functions have been included, within which the function oscillates between positive and negative values.

Software is used to correct for the phase-gradient in $\tilde{F}''$, equivalent to shifting the correlation function back to the centre of the lag window, but as τ_{err} becomes larger, more signal is lost and the amplitude of $\tilde{F}''$ becomes significantly lower than $\tilde{F}'$. We can calculate the magnitude of this loss. For a Gaussian spectral line with centre frequency ν_0 and half-width σ, originating from a point source and with no noise,

$$\tilde{F}(\nu) = Ae^{-\frac{(\nu-\nu_0)^2}{2\sigma^2}}. \tag{2.11}$$

If there is a delay error of τ_{err} and the lag window is not tapered, then after convolution the new maximum amplitude of the line has

$$A''(\sigma, \tau_{\text{err}}) \propto \int_{-\infty}^{+\infty} Ae^{-\frac{(\nu-\nu_0)^2}{2\sigma^2}} e^{i2\pi\nu\tau_{\text{err}}} \text{sinc}\{2\pi\tau_{\text{max}}(\nu - \nu_0)\}d\nu. \tag{2.12}$$

When $\tau_{\text{err}} = 0$, the correlation function is centred and A'' is a maximum. The fractional loss when $\tau_{\text{err}} \neq 0$ can be defined by

$$\gamma(\sigma, \tau_{\text{err}}) = 1 - \frac{A''(\sigma, \tau_{\text{err}})}{A''(\sigma, 0)}. \tag{2.13}$$

Parseval's theorem relates the 'power' in one transform domain to the other. In our case,

$$\int_{-\infty}^{+\infty} |G_{12}(\tau)|^2 d\tau \propto \int_{-\infty}^{+\infty} |\tilde{F}(\nu)|^2 d\nu, \tag{2.14}$$

and similarly for G'_{12} and G''_{12}. This gives a useful insight into the loss of signal. If, in Figure 2.3, area is lost under the correlation function in the lag domain (which implies that we also lose area once the function is squared), then area is also lost under the spectral line. In other words, the amplitude of the line is reduced. It is easy to predict how the loss will vary with τ_{err} for the two extremes of σ:

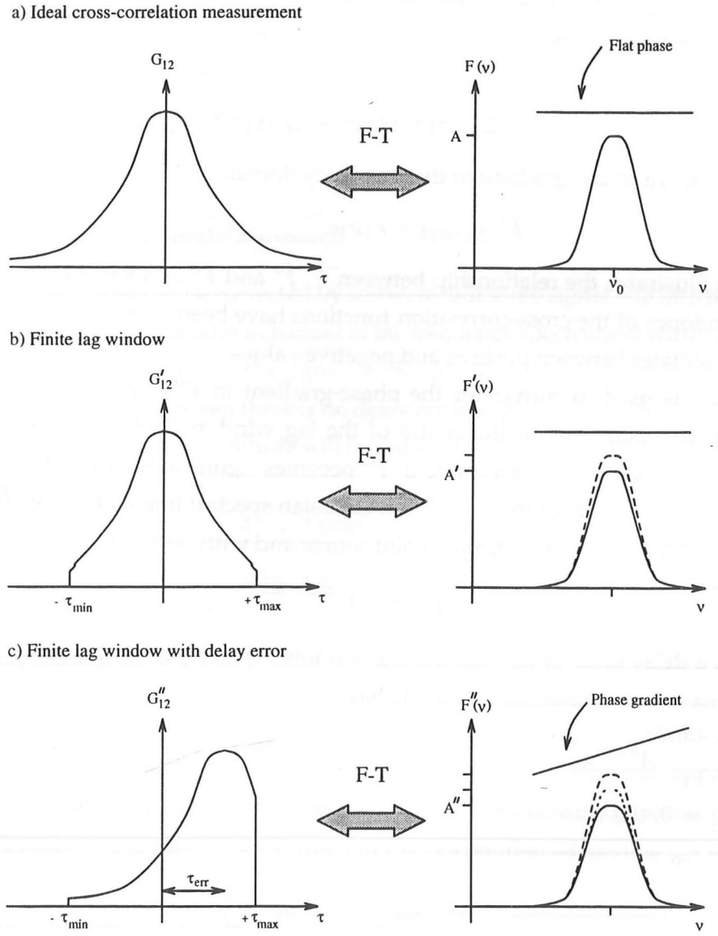


Figure 2.3: The loss of amplitude of a Gaussian spectral line when measured by a correlating spectrometer.

- **A very narrow spectral line,** σ very small. The envelope of the correlation function is very broad and practically independent of τ. Shifting it within the lag window has little effect on the area under the envelope within the window. From Parseval's theorem there will therefore be very little loss in amplitude of the narrow line and $\gamma \simeq 0$.

- **A very broad spectral line,** σ very large. The correlation function will be very narrow and none of it is lost until $\tau_{\rm err} > \tau_{\rm max}$, at which point it is shifted entirely out of the lag window. For $\tau_{\rm err} < \tau_{\rm max}$, $\gamma \simeq 0$; for $\tau_{\rm err} > \tau_{\rm max}$, $\gamma \simeq 1$.

With very little loss at the two extremes of line width, there must be a maximum loss for some intermediate value of σ. I evaluated γ numerically over a range of σ and $\tau_{\rm err}$. Figure 2.4 plots log γ as a function of the delay error in units of $(\tau_{\rm err}/\tau_{\rm max})$ for different line widths. Figure 2.5 is an alternative presentation of the data; log γ is plotted against $(2\sigma\tau_{\rm max})$, which is the ratio of the line width to the effective spectral resolution of the correlator. In both cases, I have illustrated on the plot that, irrespective of the line width, the loss is always less than 1%, if $\tau_{\rm err} < 0.15\tau_{\rm max}$. So long as $\tau_{\rm err} < \tau_{\rm max}$, the maximum loss occurs when the line width is approximately equal to the resolution.

The DAS can be configured in many different ways (§2.2.3). The smallest value possible for $\tau_{\rm max}$ occurs when there are 256 lags assigned to 160 MHz. In that case, $\tau_{\rm max} = 0.4\,\mu$s, and we must ensure that $\tau_{\rm err} < 60$ ns $(= 0.15 \times 0.4\,\mu$s). External delay compensation in steps of 120 ns satisfies this requirement. The delays must span the range $\pm$(baseline length/c) to cover all possible source locations. For the JCMT–CSO Interferometer this corresponds to $\pm$540 ns. Nine delays are needed; 0, $\pm$120 ns, $\pm$240 ns, $\pm$360 ns and $\pm$480 ns will guarantee that the correlation function is always centred in the lag window to within $\pm$60 ns, keeping losses in the resulting spectrum below 1%. Note that we have arrived at a result very similar to that of the simple, naive approach of §2.3.1.

The above treatment is based on continuous functions, whereas the DAS measures the cross-correlation function at discrete values of τ. It is straightforward to show that this analysis is still correct in the discrete case, provided the analogue signals have been sampled adequately to avoid aliasing. I have ignored any small effects due to the discrete nature of the sampling levels, however. The important result is that only nine different delays are required.

2.3.3 The need for short integrations

The maximum length of an individual integration is limited by 'smearing' of the cross-correlation function. Since the geometrical delay continually changes, the centre of the correlation function shifts during an integration. This effect is equivalent to convolution with

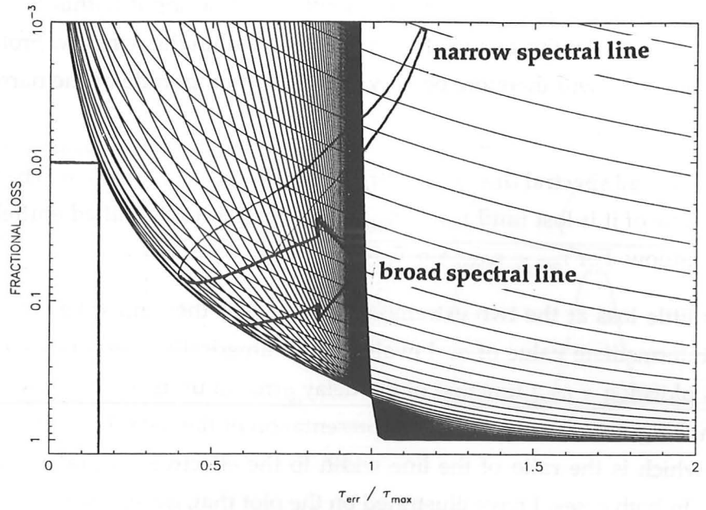


Figure 2.4: The fractional loss in amplitude of a spectral line as a function of the line width and the ratio of delay error to the half-width of the lag window.

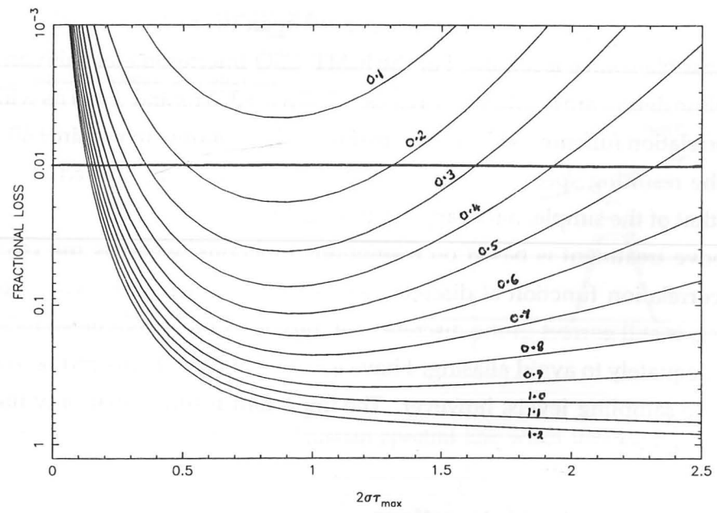


Figure 2.5: The fractional loss of amplitude of a spectral line. The numbers on the different curves give τ_{err}/τ_{max}.

a top-hat function of width $\frac{d\tau_g}{dt}t_{int}$ in the lag domain, or multiplication by sinc $\left(\pi\frac{d\tau_g}{dt}t_{int}\nu_{sub}\right)$ in the frequency domain. There is a loss of sensitivity towards the high (intermediate) frequency edge of each subband; with a small-angle approximation for the sinc function, the loss can be expressed as

$$1-\eta_{smear} \approx \frac{1}{6}\left(\pi\frac{d\tau_g}{dt}t_{int}\nu_{sub}\right)^2. \qquad (2.15)$$

The rate of change of the geometrical delay, $\frac{d\tau_g}{dt}t_{int}$, varies up to $\sim 5\times10^{-11}$ s s^{-1} for the JCMT–CSO baseline, and ν_{sub} has a maximum value in each subband of 160 MHz. For $t_{int} = 10$ s the maximum loss is 1%; for $t_{int} = 20$ s this rises to 4% and then 9% for $t_{int} = 30$ s. It is for this reason that we restrict the length of an individual integration to 10 s.

2.4 Using optical fibres

It was decided at an early stage in the project to use optical fibre instead of co-axial cable for the transmission of the IF signals and the 6-GHz reference tone for the local oscillators. Optical fibre is extruded from glass; the core region has a higher refractive index than the surrounding cladding and this ensures that light propagating along the core undergoes total internal reflection at the boundary. For single-mode fibre, the core region is so small that only one mode of propagation is possible. This minimizes the dispersion and gives the highest possible bandwidth. Single-mode fibre for operating at a wavelength of 1.31 μm typically has a core diameter of 9 μm and a potential bandwidth of hundreds of gigahertz. A fibre-based system needs a modulator to convert electrical signals into variations of either the light intensity or the phase, and a detector to make the conversion back again.

There are a number of advantages of optical fibre over co-ax, and several disadvantages:

√ Fibre is much more compact than co-axial cable. Bare fibre has a diameter of about 250 μm. A protective sheath of kevlar increases this to about 3 mm. Good quality, low-loss co-ax has a diameter of about 10 mm and is much less flexible than the equivalent fibre. This was the major deciding factor, since it is almost impossible to install thick co-ax cables in the existing conduit between the JCMT and the CSO.

√ Fibre-based systems have much larger bandwidths than those using co-ax, since the signals are modulated on a very high carrier frequency. The speed of the modulator is the main limitation, but systems with bandwidths up to 20 GHz or so are now available commercially. Co-ax often requires circuitry to re-balance the frequency response.

√ Signals propagating in fibre are not prone to electrical interference.

√ Standard single-mode fibre, costing about £0.60 per metre, is much cheaper than co-ax.

√ Special fibre can be used that has a temperature coefficient of expansion as low as 0.04 ppm/K. This is important for minimizing thermal drifts, and is particularly relevant to the transmission of the LO reference. Specially designed phase-stabilized co-ax has a temperature coefficient of ~1 ppm/K.

√ Lengths of fibre can be connected together with very little loss. In the fusion-splicing technique, two ends of fibre are very carefully aligned and then fused together with an electrical discharge arc. Insertion losses of less than 0.1 dB are obtained consistently.

× The modulators and detectors are very expensive. A laser diode / photodetector pair with a bandwidth of 10 GHz costs ~£10,000.

× The connectors available are less robust than their co-axial equivalents and care is required when making or breaking a connection.

× The total system has a relatively high insertion loss. The typical electrical loss is about 38 dB, so amplification is required.

We purchased the laser diode / photodetector pairs from the Ortel Corporation. The light intensity is varied by direct modulation of the laser current. The bandwidth is 10 GHz, but we use only a small part of this. It is important that the laser diodes are maintained at the correct operating temperature and each unit has a built-in Peltier heater/cooler. Status LEDs indicate any problems. One of the output pins on the photodiode units carries a voltage that is proportional to the light level being received, to which I connected LED bar displays. These proved useful for showing that the light paths were not attenuated by bad connections or breaks.

We used special temperature-compensated fibre from Sumitomo for the long stretches running between the two telescopes. One cable, containing four of these fibres, was pulled through the underground conduit that connects the two sites. These fibres have an outer cladding with a negative temperature coefficient of expansion, giving the claimed overall thermal expansion coefficient of 0.04 ppm/K (the figure *guaranteed* by Sumitomo is more like 0.5 ppm/K). This fibre was also used for the compensating length at the JCMT.

2.5 Design and construction of the delay lines

2.5.1 Choosing the basic configuration

The delay lines were implemented in optical fibre, to take advantage of the many benefits stated above, especially the small size and high stability. The conclusion of §2.3 was that at least nine different relative delays are needed between the two IF signal paths. The delay step of 120 ns corresponds to about 25 m of fibre (glass has a refractive index of ~1.5). There is a variety of commercially available switches for use with single-mode optical fibre, including 2-way, 3-way and cross-over types. An electrical relay is used to insert a small glass prism in the light path to redirect it. This demands a high level of precision, which is reflected in the price. Figure 2.6 shows three different possible configurations using the different types of switches. The important requirements, on which the different designs can be judged, are listed here:

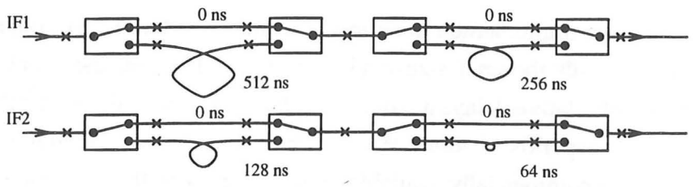
- It is critical that there is good isolation between the two IF signals. Figure 2.7 shows two ways of introducing spurious correlation between the signals. The dominant source of IF power is the noise, but this is not correlated between the two IFs (unlike the signal). If, however, there is leakage from one IF to the other, then there can be substantial spurious correlation.
- The number of splices, connectors and switches in the signal path should be kept to a minimum to reduce loss and back-reflection.
- The cost should be minimized where possible.

Table 2.1 compares the three different designs presented in Figure 2.6. It is clear that design C is most suited to our needs. The final design is shown in Figure 2.8 and has an extra modification; one of the 3-way switches is replaced by a passive, 3-way coupler in each IF arm. These couplers were made by Canstar and consist of three fibres fused together at a common point. They are designed so that a signal at any one of the three inputs is split evenly between the exit fibres. Two of the exits are terminated to give very little back-reflection, and the other provides the single output. They are much cheaper than the switches and have no moving parts that might fail. The penalty is an insertion loss of 1.5 dB, compared to 0.5 dB for the switches. These values refer to the optical power—the equivalent loss to the RF power is twice this value.

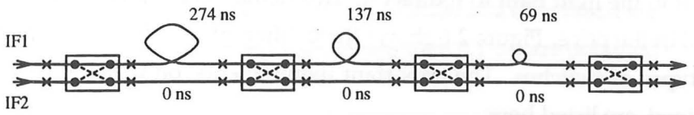
The two 3-way switches were purchased from BT&D, a subsidiary of British Telecom. The switch position is determined by two TTL levels; I connected LEDs to indicate which delay lines are in use. Switching time is less than 10 ms.

The fibres were spliced together using the fusion splicer at the Optoelectronics group of Cambridge University Engineering Department. Tubular metal sleeves, secured at each

a)



b)



c)

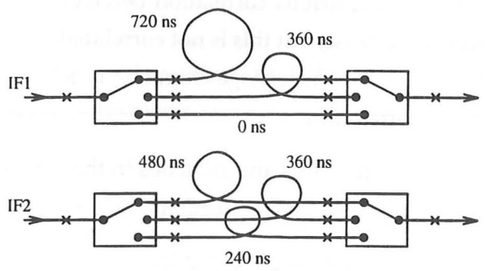
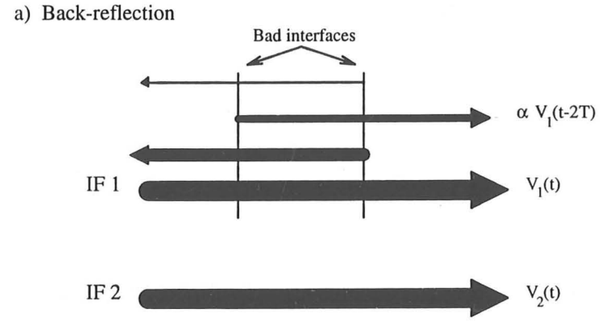


Figure 2.6: Three possibilities for the delay line system: (a) uses simple 2-way switches, (b) uses cross-over switches and (c) uses 3-way switches. Numbers indicate the delay provided by each length of fibre. Crosses show where splices are needed.

	# relative delays	# switches	cost per switch	# splices in path	# switches in path	isolation
A	16	8 × 2-way	~£700	7	4	√
B	8	4 × cross-over	~£1000	8	4	×
C	9	4 × 3-way	~£1100	4	2	√

Table 2.1: Comparison of different delay line designs

a) Back-reflection



b) Leakage between the two IF paths

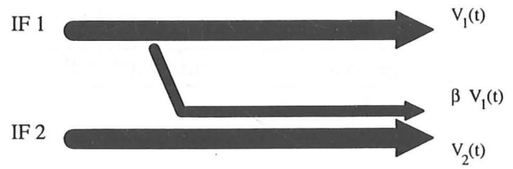


Figure 2.7: Spurious correlations introduced by (a) back-reflections from two bad interfaces and (b) leakage from one IF path to the other. The latter is the main cause for concern.

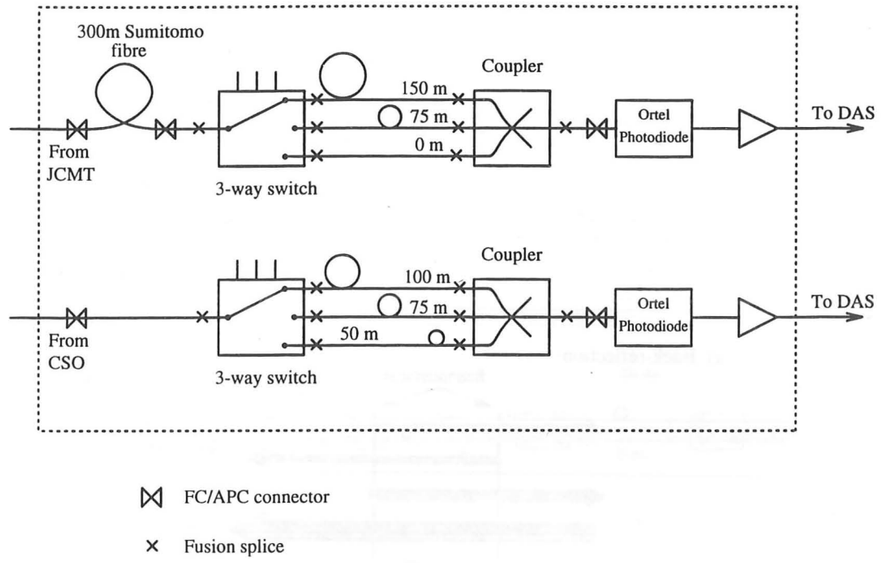


Figure 2.8: The schematic design adopted for the delay lines using two 3-way switches and two couplers. The numbers indicate lengths of fibre.

end with epoxy, protect the joins. Connectors are fitted at both inputs and outputs so that the delay lines can be removed from the IF path. I chose to use high-performance FC/APC connectors (Angled-Physical Contact single-mode Fibre Connector, conceived by Seikoh). These have a typical insertion loss of 0.2 dB and a back-reflection rating of –68 dB (again, these figures are for the optical power).

2.5.2 Packaging the delay lines

The incorporation of this design into a physical package was an interesting problem. My main concern was that the fibres should be well-protected; any damage to the fibres would be very hard to repair once the system was installed in Hawaii. I settled on the solution depicted in Figure 2.9, which shows one of the two IF arms.

The appropriate lengths of bare fibre were measured out and laid in the different slots cut into 1-inch packing foam. The foam was glued onto a sheet of hardboard and the coupler mounted at the centre. The pieces of foam that were removed to form the slots were then replaced to protect the fibres and coupler underneath. The switch is mounted externally to allow ease of access for the electrical TTL connections to the microcomputer. The fibres entering and leaving the foam unit are sheathed in kevlar for protection. Two

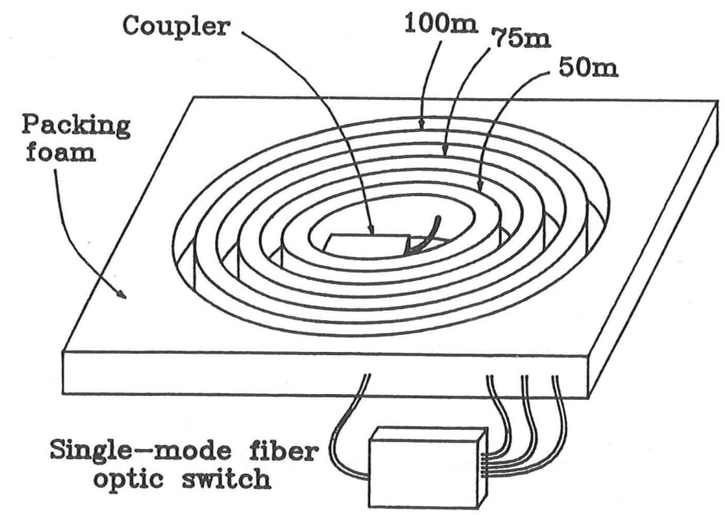


Figure 2.9: The packaging of the components for one arm of the IF delay system.

of these units are required—they sit one on top of the other in the bottom of a 19-inch rack box. The total size, excluding the switches, is just $50 \times 35 \times 5$ cm.

2.5.3 Operation in practice

There have been no problems so far with the delay lines. The (total RF power) insertion loss for the optical system is about 45 dB. There has been no indication of spurious correlation, although no formal tests to determine this have been carried out. Because the ends of the fibres were trimmed for fusion splicing, the exact values of the relative delays were not known and had to be measured using the procedure described in §3.1.2.

2.6 Tracking the fringe

The rôle of the two Direct Digital Synthesizers (DDSs) was explained in the last chapter. They are required to produce the following phases as a function of time:

$$\begin{aligned} \phi_{\text{IFDDS}} &= 2\pi\nu_1 t + A\sin\left(\frac{2\pi t}{T_\star}\right) + B\cos\left(\frac{2\pi t}{T_\star}\right) + C + \pi P(t), \\ \phi_{\text{LODDS}} &= 2\pi\frac{\nu_1}{M}t, \end{aligned}$$

where $\nu_1 = 1$ MHz, t is the time since midnight, Universal Time (UT), $T_\star$ is the period of a sidereal day and M is the frequency multiplication factor in the Gunn phase-locked loop. The coefficients A, B and C are constants for a particular source on a particular day, and

are defined in §3.1.1; they account for the fringe-rate, which varies up to about 15 Hz when observing in the 'B' band. The phase-switching function $P(t)$ alternates regularly between +1 and 0.

The next sections describe how a DDS works and the way in which it is made to deliver the required phases.

2.6.1 How a DDS works

The DDSs that we use are a manufactured by Stanford Telecom. The outputs are sine and cosine waveforms whose frequencies are determined by digital inputs. Output frequencies of up to 25 MHz are possible when the maximum clock speed of 60 MHz is used. The frequency required is specified by a 32-bit binary number, giving a resolution of 14 mHz with the 60-MHz clock. The resolution is improved at lower clock frequencies; we operate with a 10-MHz clock to attain a frequency resolution of 2 mHz. Both phase and frequency modulation are controlled digitally.

Each DDS comes on a circuit-board of size 6.5 × 3.5 inches. Most of the work is done by the STEL-1177 Numerically Controlled Oscillator chip. Central to its operation is a 32-bit accumulator that holds the current value of the phase in the form of a binary number, such that $2^{32} \equiv 360°$. On every clock cycle, the value in the phase accumulator is incremented by another 32-bit number that is proportional to the output frequency. The 13 most significant bits of the 32-bit phase are used to address a look-up table in memory that contains values of the sine and cosine functions. These digital amplitudes are then converted into analogue form and smoothed by low-pass filters to remove aliased components. It is possible to modulate the phase of the sine and cosine channels independently, whereas frequency modulation is common to both.

One of the two 25-pin 'D' type connectors on each DDS is entirely devoted to frequency modulation, for which we have no use. Digital input information regarding the frequency required and the phase offset is delivered, 8 bits at a time, using 8 data lines. Three address lines indicate whether the 8 bits of data refer to a new phase offset or to a new frequency that is to be set. The data are read into an input buffer; the output of the DDS is not affected by the new values until the 'load frequency' or 'load phase' commands are given using separate lines. At this point the new phase offset and/or frequency is transferred from the input buffer to the Arithmetic Logic Unit and the output waveforms change accordingly. I wired most of the inputs on the two DDSs in parallel to the interface cards in microcomputer. The exception in each case is the 'chip select' line that makes the circuit-board respond to inputs. Appropriate levels applied to these lines allow the DDSs to be controlled either individually (e.g. for giving them their different frequencies) or simultaneously (e.g. to start them off in phase with one another).

2.6.2 Following the phase function

The biggest challenge was to get the IF DDS to generate the correct phase required for fringe-tracking. The fringe phase, corresponding to the fringe frequency, is given by

$$\phi_{\text{fringe}}(t) = A \sin\left(\frac{2\pi t}{T_\star}\right) + B \cos\left(\frac{2\pi t}{T_\star}\right) + C. \tag{2.16}$$

This can be approximated by a Taylor expansion about time t_0;

$$\phi_{\text{fringe}}(t - t_0) \simeq D + E(t - t_0) + F(t - t_0)^2, \tag{2.17}$$

where

$$D = \phi_{\text{fringe}}\Big|_{t_0}, \qquad E = \frac{d}{dt}\phi_{\text{fringe}}\Big|_{t_0}, \qquad F = \frac{1}{2}\frac{d^2}{dt^2}\phi_{\text{fringe}}\Big|_{t_0}. \tag{2.18}$$

A DDS being clocked at 10 MHz has a resolution of 2 mHz, so that the difference between the frequency specified by the linear term E and the frequency that the DDS actually delivers is at worst 1 mHz. Assuming for the moment that the quadratic term is negligible, then the largest phase error that can accrue over a 10 s integration is 0.01 turns, or 4° (Figure 2.10a). This is not too serious. The value of F, however, can vary up to $\sim 3 \times 10^{-3}$ rad s^{-2}, and after 10 s the discrepancy between the phase that we need and the phase that is delivered by the DDS can be up to 20° (Figure 2.10b). For a 20 s integration this increases to 80°.

To prevent this phase error I decided to employ the scheme shown in Figure 2.10c, where regular phase offsets are applied to keep the output of the IF DDS close to the required curve. Notice that there is also an initial offset. The micro knows the curve that is required and keeps track of the phase at the output of the DDS, so that it is able to demand the appropriate phase corrections.

It should now be apparent that very accurate timing is needed in the control of the IF DDS. For example, a 1 ms discrepancy between the actual time and the time according to the computer would result in a phase error of up to about 5°. This timing requirement is fundamental to any interferometer that has a maximum fringe-rate of 15 Hz. The method by which it is attained is addressed in the next two sections, firstly in conjunction with a description of the micro and then as part of the overall sequence of events that takes place in an integration.

2.7 The microcomputer

The interferometry micro has three basic tasks:

- to coordinate the fringe-tracking and phase-switching activities of the two DDSs

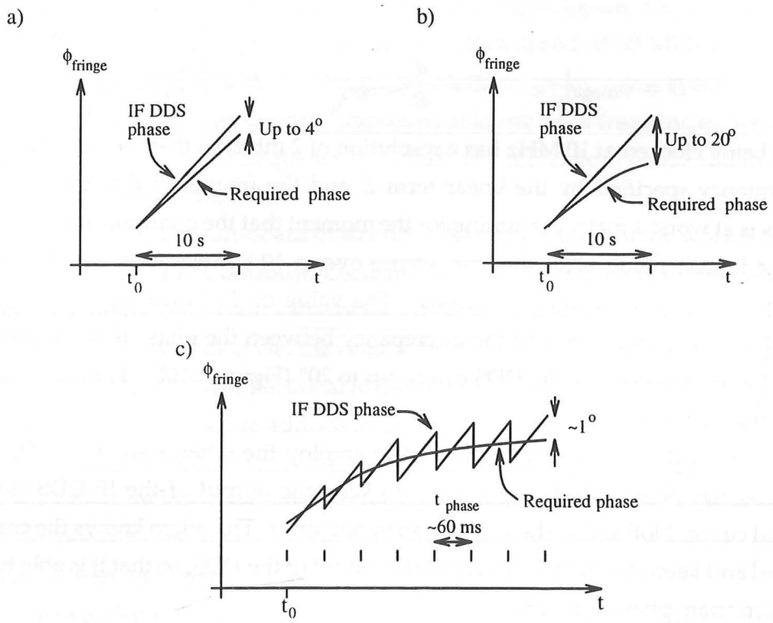


Figure 2.10: Tracking the fringe phase with the IF DDS. (a) Shows the discrepancy that can arise from the linear component when using a constant DDS frequency and phase offset. (b) As (a), except for the quadratic component. (c) Minimizing the error using regular updates of the phase.

- to instruct the DAS which buffer should be used to accumulate the current integration (this must be synchronous with the phase-switching)
- to switch the delay lines when required

It is located in the control room at JCMT and has the following components: a 7U high, 19-inch wide crate with fan tray, a 350 W switched-mode power supply, a VME bus with 12 slots to take 6U cards, a 2 Mb 3.5" floppy drive, an 83 Mb hard drive, a Mizar 7130 CPU card with real-time clock, two 16-channel digital output cards from PEP, one 16-channel digital input card from PEP, a Mizar GPIB communication card, the OS 9 operating system and a C compiler. The following paragraphs describe important aspects of the system and highlight some of the many problems that I had to overcome.

The digital output cards: Outputs are connected to the DDSs, the delay line switches, a 'D' type latch (see below) and the S/R line of the correlator that selects the buffer. The response speed of the output boards was limited by the performance of the opto-isolation circuitry. Errors resulted when the software tried to switch the outputs too quickly (on consecutive lines of code, for example) and the switching software had to be slowed down by introducing pauses in the code.

The digital input cards: Only two inputs were needed; they were used to monitor the 1-Hz JCMT clock and the CRDY pulses from the DAS (via the 'D' type latch). The board only responded to signals exceeding 13 V. I bypassed the Zener diodes on the inputs and added extra resistors to make the board compatible with TTL levels. De-bounce circuitry for rejecting pulse lengths shorter than 5 ms proved very restrictive, so I replaced the capacitors on the de-bounce chips with smaller ones to reduce the rejected pulse length threshold to 1 ms.

The real-time clock: The micro must know the current time to the nearest second, so that it can correctly assign times to the ticks of the 1-Hz telescope clock. Originally, I opted to use the real-time clock on the CPU card; I soon discovered, however, that it ran fast by about 1 part in 10^4, but only when the power supply to the micro was turned on. At 14 000 ft the CPU card runs much hotter than normal; the clock is operating outside its temperature rating and runs fast. This problem was avoided once communication with the VAX was established, as the VAX simply tells the micro which second of the day it will be at the next tick of the 1-Hz clock.

The 'D' type latch: The CRDY pulses from the DAS (§2.2.2) that signify the end of each correlation cycle are only 6.4 μs wide—too short to be picked up by the input board. I used

the rising edge of each pulse to change the output of a 'D' type latch from high to low. This falling edge triggers the DDSs and, because the output of the latch remains in the low state after the end of the CRDY pulse, there is enough time for the digital input boards and the micro to register that such a pulse has occurred. After this, the micro resets the latch so that it is ready for the next pulse.

The software: I wrote a program in C to run on the micro. It coordinates the various events during an integration by reading the input cards and writing to the output cards. The basic steps and calculations should be clear from the sequence of events described in §2.8.

Communication with the VAX: There was no connection between the JCMT VAX and the micro for the first two runs, and a number of parameters had to be entered by hand. The coefficients A, B and C, the observing frequency and sideband, the phase-switching frequency and the integration length all had to be typed in when a new source was observed. Before the March 1993 run, we fitted a GPIB communications card that allowed both computers to talk to each other using Steve Brooke's COMHNDL software. This in itself was the cause of many problems which have since been ironed out. The VAX now passes the parameters to the micro. In addition, it sends the delay line settings needed and the time of the next 1-Hz tick in seconds after UT midnight.

2.8 The sequence of events for an integration

Events during an integration are synchronized using two timing cues. The integration start time, t_0, always coincides with a tick (the low to high transition) of the 1-Hz JCMT telescope clock. The VAX that controls the telescope knows the time of each tick, and before the start of each integration it sends the value of t_0 to the interferometry micro, via the GPIB link. It then instructs the DAS to start on that tick. Within each integration, the CRDY pulses from the DAS are used as timing marks for the fringe-tracking and phase-switching operations. CRDY pulses denote the end of each correlation cycle, whose period, $T_{\rm cor}$, is set by the user in units of 6.4 μs. We use $T_{\rm cor}$ = 12.0 ms (= 1875 × 6.4 μs). The period between phase-switches is $N_{\rm phase}T_{\rm cor}$, where $N_{\rm phase}$ is also set by the user. The length of an integration is $N_{\rm int}N_{\rm phase}T_{\rm cor}$ (where $N_{\rm int}$ is the number of phases per integration) and should be an even number so that there are equal periods of integration for the 0° and 180° phase-switchings. Usually we have $N_{\rm phase}$ = 5, and $N_{\rm int}$ = 162, giving a phase-switching frequency of 8.3 Hz and an integration time of 9.72 s. We normally carry out integrations in groups of 10, with the next integration starting on the 1-Hz tick that follows the end of

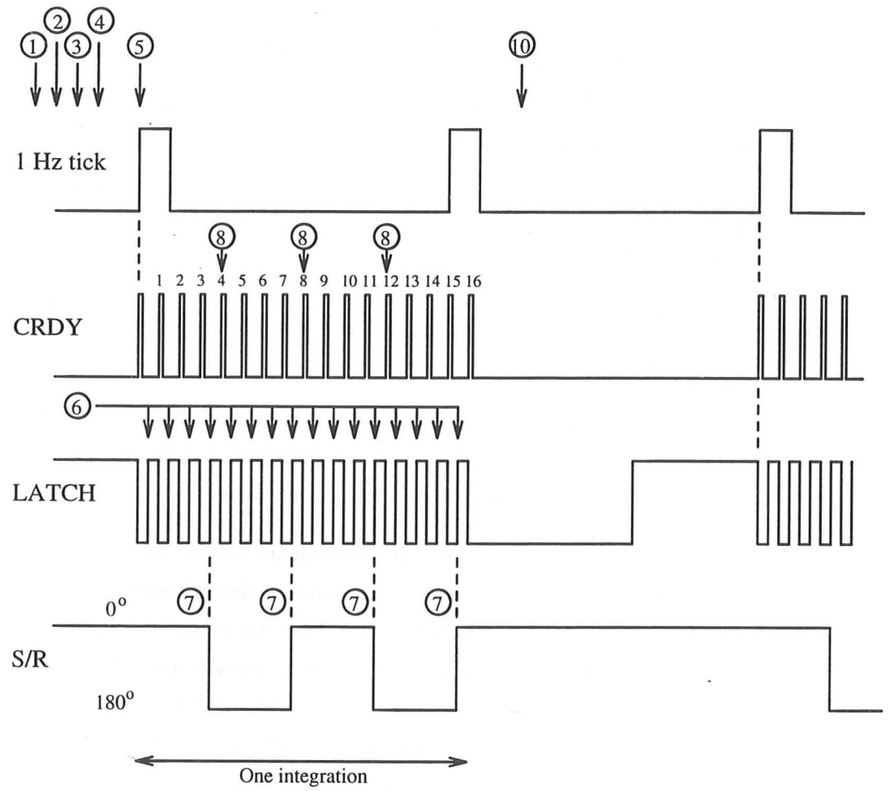


Figure 2.11: The sequence of events during an integration with $N_{\rm phase}$ = 4 and $N_{\rm int}$ = 4. The circled numbers correspond to the steps described in the main text.

the previous one. This gives an observing efficiency of 97%. The results for each set of 10 integrations are held in the same data file and share common header information (see chapter 4).

The basic steps during an integration are listed below with numbers corresponding to those in Figure 2.11, which illustrates the specific case of $N_{\rm phase}$ = 4, $N_{\rm int}$ = 4.

1. The VAX calculates A, B, C, and the delay line setting needed for the source currently being observed. These are sent to the micro, along with the time of the next tick in seconds after UT midnight, the sky frequency and the sideband.

2. The micro sets the delay lines.

3. The VAX commands the DAS to start an integration on the next tick, and the micro

calculates the phase offset and fringe frequency required for an integration starting at this time. The sine and cosine outputs of the LO DDS are effectively switched if the observations are upper sideband, so that the output frequency is added to, rather than subtracted from, the 20 MHz (Figure 1.9). The DDSs are both reset and the new values for phase and frequency are loaded into the input buffers.

4. The micro looks for the latch to go low, indicating that the DAS has started. If the latch stays high for over 1 s then there is no integration, in which case the number of previous successful consecutive integrations is reported back to the VAX and the micro waits for stage 1.

5. When the latch goes low, the new DDS frequencies and phases come into effect.

6. The micro resets the latch and waits for the next CRDY pulse. This is repeated until there have been $(N_{\text{phase}} - 1)$ CRDY pulses in the current phase setting (excluding the pulse that marks the very start of the integration).

7. The micro toggles the S/R line. There is a pipeline delay in the DAS that postpones the effect of this change by one correlation cycle, so that the values accumulated for the current cycle go to the buffer denoted by the state of the S/R line on the *previous* CRDY pulse. This subtle behaviour of the DAS was only recognized in the November 1993 run. Prior to this, we did not account for the delay, and the switching of the buffers was a whole correlation cycle out of phase with the switching of the phase by the IF DDS. This resulted in the loss of over 40% of the signal and was the cause of the low sensitivity that we had been experiencing (§1.8). The micro knows that the time of the next CRDY pulse will be $t_0 + (n+1)N_{\text{phase}}t_{\text{cor}}$ where n is the number of phase-switches that there have been so far in the integration. It calculates the new phase offset required for fringe-tracking and adds an additional 180° for the phase-switch. The new values are loaded into the input buffers of the IF DDS, but are not used yet.

8. At the next CRDY pulse the latch goes low and the new phase offset comes into effect.

9. Stages 6, 7 and 8 are repeated until N_{int} phases have been completed.

10. The micro calculates the time of the next 1-Hz tick and goes to step 3.

2.9 Summary

This chapter has concentrated on many of the technical aspects of the interferometer. I have discussed how the correlator works and shown how it reduces the number of discrete

delay lines that are needed. My calculations showed that 9 relative delays are needed for the JCMT–CSO Interferometer to ensure that the loss is less than 1%. These delays were implemented using 3-way switches to select lengths of optical fibre, and packaged in a robust and compact unit. A pair of Direct Digital Synthesizers are used to track the fringe with a precision of about ±1°, to provide 180° phase-switching and to reject one of the signal sidebands. Both the DDSs and the delay lines are controlled by a microcomputer that I interfaced and programmed. All of the hardware that I have described is now working as intended.

The only addition planned is a down-converter to shift the IF output of receiver C2 at the JCMT from a centre frequency of 4 GHz to 1.5 GHz. A corresponding shift of the JCMT LO frequency is also needed. A unit to do both these tasks has been built by John Carlstrom and should be used for the first time in October 1994.

Chapter 3

CALIBRATION

The correct calibration of an interferometer is critical to its performance. There are several aspects that need attention. The phase of the correlator output drifts over time; this must be corrected in order to make long, coherent integrations on faint sources. We must also determine the system temperature to convert the correlation levels into effective antenna temperatures. These are scaled in turn to units of flux, by a factor that depends on the collecting area and the efficiency of the system. Finally, we must measure the complex passband (the frequency-dependent gain and phase of the instrument) so that it can be removed from the astronomical spectra. I address each of these issues in turn.

3.1 Phase calibration: the challenges

The aim is to predict the correlator output when observing a point source in any position and at any time. This is equivalent to knowing the exact difference in path length between the two alternative routes from the source to the correlator. Achieving this goal requires a detailed knowledge of the system, including

- the exact length of the baseline,
- the lengths of the delay lines and other cabling,
- the effects of thermal drift on the phase of the local oscillators,
- the amount by which each dish deforms under gravity,
- the offset between the azimuth and elevation axes of each antenna,
- several atmospheric effects, such as the refractive index,
- any other phase variations that are introduced as the dishes move, for example bumps in the azimuth tracks.

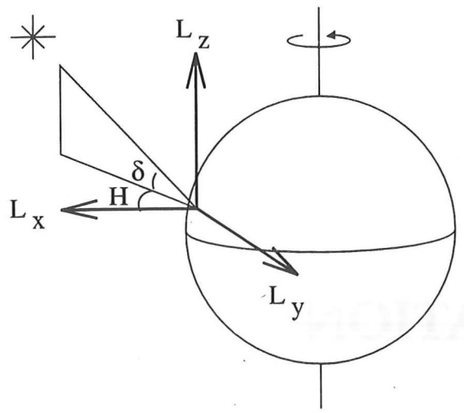


Figure 3.1: The convention used for the three components of the baseline. The vectors are with respect to the local longitude of the interferometer, with L_z in the polar direction, L_y pointing to the east and L_x outwards to the celestial equator.

The most accurate measurements of the baseline components and the offsets between the azimuth and elevation axes are made by modelling the way in which the phase changes with time for a number of astronomical sources. In this section I analyse each of the contributions in more detail, but I defer the results of the model-fitting for the baseline and the axes to §3.2.

3.1.1 Baseline length

The baseline vector **D**, measured between the points of intersection of the azimuth and elevation axes at each antenna, can be described by three orthogonal components, L_x, L_y and L_z. We have defined them as shown in Figure 3.1. As a first estimate, the baseline was surveyed by a group from the University of Hawaii, who measured the distances to corner-cube reflectors mounted on the tertiary mirrors of the JCMT and CSO using lasers. The baseline components were calculated by triangulation with a claimed uncertainty of about ±1 cm.

It is straightforward to calculate the effect of an error in the baseline on the output of the correlator. The phase difference between the signals as they reach the antennas is given by

$$\Delta\phi_{\rm ant} = 2\pi \frac{\mathbf{D}}{\lambda_{\rm sky}} \cdot \mathbf{s}, \tag{3.1}$$

where **s** is the unit vector in the direction of the source. For a source at local hour angle H

and declination δ (Figure 3.1), we can expand the vectors to give

$$\Delta\phi_{\rm ant} = \frac{2\pi}{\lambda_{\rm sky}} \begin{pmatrix} L_x \\ L_y \\ L_z \end{pmatrix} \cdot \begin{pmatrix} \cos\delta\cos H \\ -\cos\delta\sin H \\ \sin\delta \end{pmatrix} \tag{3.2}$$

$$= \frac{2\pi}{\lambda_{\rm sky}} (L_x \cos\delta\cos H - L_y \cos\delta\sin H + L_z \sin\delta). \tag{3.3}$$

The local hour angle H can be expressed in terms of the hour angle of the source at UT midnight, $H_{\rm ut0}$, and the number of seconds that have elapsed since then, t:

$$H = H_{\rm ut0} + 2\pi \frac{t}{T_*}, \tag{3.4}$$

where T_* is the period of a sidereal day. Expanding the sine and cosine, we obtain

$$\cos H = \cos H_{\rm ut0} \cos\left(2\pi\frac{t}{T_*}\right) - \sin H_{\rm ut0} \sin\left(2\pi\frac{t}{T_*}\right), \tag{3.5}$$

$$\sin H = \sin H_{\rm ut0} \cos\left(2\pi\frac{t}{T_*}\right) + \cos H_{\rm ut0} \sin\left(2\pi\frac{t}{T_*}\right), \tag{3.6}$$

so that equation 3.3 can be written in the form

$$\Delta\phi_{\rm ant} = \frac{2\pi}{\lambda_{\rm sky}} \left\{ A \sin\left(2\pi\frac{t}{T_*}\right) + B\cos\left(2\pi\frac{t}{T_*}\right) + C \right\}, \tag{3.7}$$

where

$$A = -\cos\delta(L_x \sin H_{\rm ut0} + L_y \cos H_{\rm ut0}), \tag{3.8}$$

$$B = \cos\delta(L_x \cos H_{\rm ut0} + L_y \sin H_{\rm ut0}), \tag{3.9}$$

$$C = L_z \sin\delta. \tag{3.10}$$

These are the three parameters given to the microcomputer (§2.6.2). If there are baseline errors ΔL_x, ΔL_y and ΔL_z, then from equation 3.3, the corresponding phase error is given by

$$\phi_{\rm base} = \frac{2\pi}{\lambda_{\rm sky}} (\Delta L_x \cos\delta\cos H - \Delta L_y \cos\delta\sin H + \Delta L_z \sin\delta). \tag{3.11}$$

For a given source, the phase varies sinusoidally with period T_*, with an amplitude that depends on ΔL_x and ΔL_y. There is also an offset proportional to ΔL_z. A baseline error of $\lambda_{\rm sky}$ (~1 mm) gives an amplitude of 2π for a source at zero declination. The results of modelling the baseline errors are presented in §3.2.

3.1.2 Lengths of the delay lines and cabling

We can directly measure the lengths of the delay lines and other fixed propagation lengths in the system. Accurate lengths are needed for the software correction of the delay error

(§4.1.4). In equation 1.30, the delay lines introduce delays τ_j and τ_c; other cable lengths are accounted for by the Δt terms, which are then combined in the phase term ϕ_{const}. Equation 1.30 shows that we only need to know the relative lengths of the delay lines; if the delays on the JCMT side are actually 1.7 ns, 253.2 ns and 498.4 ns, they can be represented by values of –251.5 ns, 0 ns and +245.2 ns, with the extra 253.2 ns absorbed into the term Δt_{j24}. We can now deal with ϕ_{const}:

$$\phi_{const} = \pm 2\pi\nu_{sky}(\Delta t_{c12} - \Delta t_{j12}) + 2\pi(\nu_{cif}\Delta t_{c24} - \nu_{jif}\Delta t_{j24}). \quad (3.12)$$

Writing $\nu_{sky} = \nu_{sky0} + \Delta\nu$, $\nu_{jif} = \nu_{jif0} \pm \Delta\nu$ and $\nu_{cif} = \nu_{cif0} \pm \Delta\nu$ (as in §1.5.1) and collecting terms together gives:

$$\phi_{const} = \pm 2\pi\nu_{sky0}(\Delta t_{c12} - \Delta t_{j12}) \pm 2\pi\Delta\nu(\Delta t_{c14} - \Delta t_{j14}) \pm 2\pi(\nu_{cif0}\Delta t_{j24} - \nu_{jif0}\Delta t_{j24}), \quad (3.13)$$

where $\Delta t_{c14} = \Delta t_{c12} + \Delta t_{c24}$ and similarly for Δt_{j14}. The first term is quite large but is also very nearly constant; it varies by a small amount to compensate for the changing velocity of the telescopes with respect to the source. The magnitude of this variation is of order 0.5° over a night's observing and is not worth correcting. The second term arises from the path difference within the instrument, excluding the delay lines, and must be corrected, whereas the final term is a constant and is ignored.

We must measure the relative lengths of the delay lines and any additional path offsets that make up the Δt_{j14} and Δt_{c14} terms, including cabling and the primary-to-secondary mirror distances. Errors will result in a phase-gradient across the passband; to keep the phase response flat to within ±5° across 1 GHz, we must know the path difference to within 8 mm. The delay line lengths, however, need to be more accurately determined than this, to ensure that the phase of a source does not jump when the delay lines are switched. The lengths should all be known to the nearest 3 mm to prevent jumps of more than 5°. Although I measured the lengths of fibre carefully when I was building the delay system, small amounts were cut off in the splicing process.

Figure 3.2 shows the set-up that I devised to determine the delay line lengths after installation. A pair of delay lines is selected and used as the arbitrary zero of length. In addition to a variation in amplitude, the complex spectrum of the correlated noise obtained with these delay lines in position will have a phase-gradient, arising from the difference in length of the two paths from the CSO receiver to the correlator. This spectrum is saved and used as the passband calibration (§3.4) for the spectra obtained on the other delay line settings. These passband-corrected spectra should be flat in amplitude, and have a phase-gradient that gives the path difference between the delay lines being measured and the nominal zero. Once the rough lengths of the delays are accounted for and the gradients have mostly been removed, the phase offset at the centre IF frequency becomes

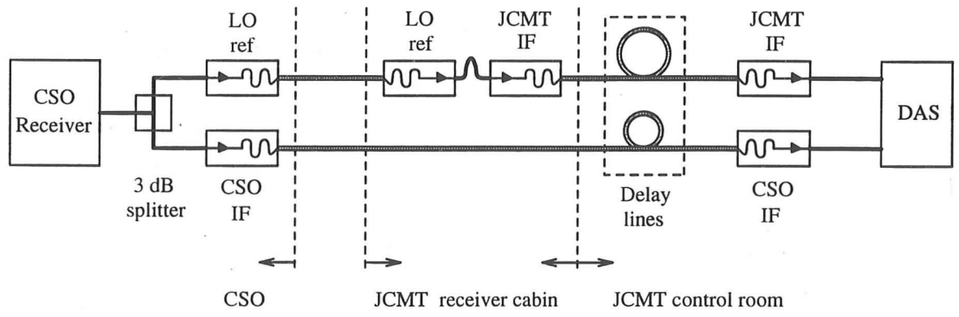


Figure 3.2: Measuring the lengths of the delay lines. One of the receivers at the CSO is used as a source of noise which is then split and transmitted down both of the fibre links between the telescopes. In the JCMT receiver cabin, the output of the link that is normally used to transmit the local oscillator reference tone is connected to the input of the JCMT IF link. Both the CSO and the JCMT fibres now carry the same signal. This passes through the delay lines and into the correlator.

a more accurate indicator of the remaining error. The values obtained using the set-up in Figure 3.2 are listed in Table 3.1.

The residual path offset (the second term of equation 3.13) cannot be determined using this method. Once the relative delay line lengths have been established, we must observe a distant, bright, continuum point source. A quasar is ideal. Any phase-gradient present is now due to the residual path offset only, assuming that the geometrical delay is accurately known. It is not really necessary to measure this path offset explicitly, since it is automatically corrected for by the passband calibration process.

3.1.3 Thermal drifts

Temperature changes cause small variations in the phase of the correlator output, partly due to the thermal expansion and contraction of the fibre carrying the local oscillator reference. The frequency of this signal is multiplied up from ~6 GHz to near the observing frequency, and any phase variations in the fibre are increased by the same factor. If the fibre is stretched so that its electrical length increases by an amount λ_{sky}, then the JCMT local oscillator phase will change by 360°, as will the correlator output.

To minimize the impact of thermal changes we use special temperature-compensated optical fibre, made by Sumitomo with a thermal coefficient of expansion of 0.04 ppm/K. The fibre running in a covered conduit between the two antennas has an electrical length of ~500 m (physical length ~300 m), so a temperature change of 10 K should give a

	Nominal glass lengths (Fig. 2.8)	Relative delay	Relative vacuum path length
JCMT	0 m	–362.873 ns	–108.786 m
	75 m	0 ns	0 m
	150 m	+369.032 ns	+110.633 m
CSO	50 m	–131.402 ns	–39.393 m
	75 m	0 ns	0 m
	100 m	+121.713 ns	+36.489 m

Table 3.1: Relative propagation times through the delay lines. The uncertainty is ±5 ps.

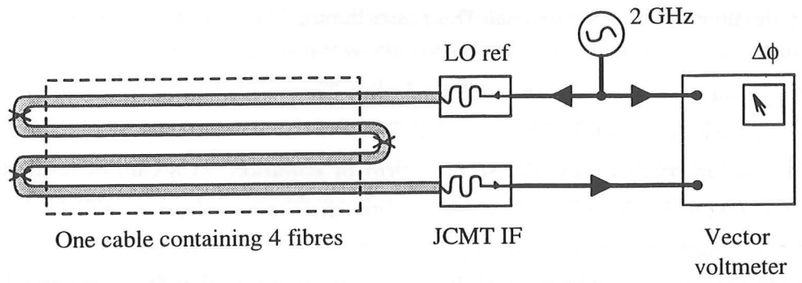


Figure 3.3: Round-trip experiment used to monitor thermal changes in the LO reference fibre and the effect of slews in azimuth and elevation. The measured changes in phase need to be scaled up to the observing frequency and then divided by four to account for the double return journey.

phase change of ∼70° when $\lambda_{sky} = 1$ mm. The Ortel transmitters and receivers, the coaxial cabling, and components making up the phase-lock system are also sensitive to temperature.

We performed the round-trip experiment shown in Figure 3.3 to establish the approximate magnitude of the thermal drift. There is a typical peak-to-peak phase variation of ∼200° over a night (calculated for observations at 345 GHz), suggesting that the fibre itself is not the dominant source of drift. This slow variation in phase can be offset by regularly observing a phase calibrator source, as described in §3.2.2.

3.1.4 Dish deformation

As each dish moves in elevation, its structure is distorted by gravity, introducing small but significant changes to the path from source to correlator. The antennas of most interferometers are built to the same design and the deformation is the same for each dish, so that the extra paths cancel. Since our two antennas are different, we must account for their distortions.

This task is not as hard as it first appears. Both dishes are designed to deform homologously under gravity (i.e. they always maintain a paraboloid form), and it is only the focal length of the primary that changes. This can be compensated by moving the secondary mirror to re-focus. Figure 3.4 illustrates this process, showing, for example, how a dish sags at the edges when pointed at zenith. This increases the focal length, f_p, of the primary, and the secondary must be moved out by an amount Δf_z to maintain the focus at the receiver. Focusing is controlled by the telescope computers, and the value of Δf_z for each antenna is recorded in the GSD data files (§4.1.1). When pointing at the horizon, the dish has the same focal length as it would in zero gravity, but the optical axis is offset. This is corrected by moving the secondary by an amount Δf_x, perpendicular to the axis. The change in path introduced by movement of the secondary mirror is given by $2\Delta f_z$, to first order. The JCMT focus moves out by ≈1.5 mm from horizon to zenith; for the CSO this figure is ≈1.6 mm.

Figure 3.4d illustrates how the change in shape of the primary can introduce extra path, as the distance between the apex of the primary and the receiver is increased by up to Δy_{max}. The magnitude of Δy_{max} can be calculated in terms of the primary mirror dimensions and Δf_z, to give an upper limit to this contribution to the path. This is shown below.

The combination of a primary and secondary mirror, with focal lengths f_p and f_s, respectively, gives a resulting focal length f_{ps} such that

$$\frac{1}{f_{ps}} = \frac{1}{f_p} + \frac{1}{f_s}. \tag{3.14}$$

Changes in f_p are therefore related to changes in f_{ps} by

$$\frac{\Delta f_{ps}}{f_{ps}^2} = \frac{\Delta f_p}{f_p^2} \tag{3.15}$$

But $\Delta f_{ps} = 2\Delta f_z$ and $f_{ps} \approx 2f_p$, so that $\Delta f_p \approx \Delta f_z/2$. Now the equation of a parabola can be written in terms of its focal length as

$$y = \frac{1}{4f_p}x^2. \tag{3.16}$$

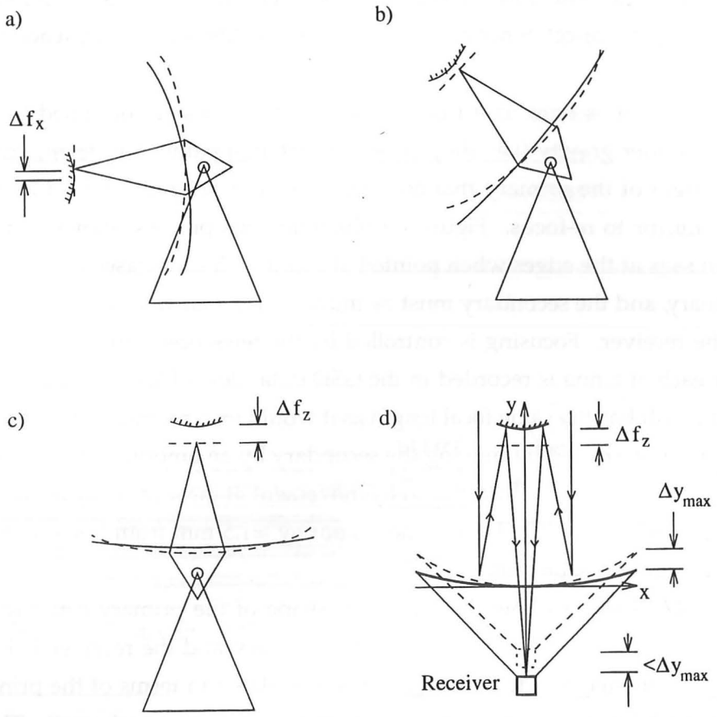


Figure 3.4: The effect of gravity on a parabolic antenna. Both the JCMT and the CSO are designed to deform homologously, retaining their parabolic shape and requiring only a change in focus. The dashed lines denote the hypothetical positions under zero-gravity. (a) When pointing at the horizon, the secondary mirror must be shifted sideways by an amount Δf_x. (b) At intermediate elevations the secondary is shifted both sideways and along the optical axis. (c) The focal length of the primary is a maximum when pointing at zenith and the secondary moves out by Δf_z to compensate. (d) The flexing of the primary can also change the path length. This illustrates an extreme case, where the distance from the apex of the primary to the receiver (or tertiary mirror) is increased by the same amount as the maximum deflection of the primary, Δy_{max}. The total extra path is $\sim 2\Delta f_z + \Delta y_{max}$. We can estimate Δy_{max} (see main text).

Changing the focal length by Δf_p needs a change of shape given by

$$\Delta y = -\frac{x^2 \Delta f_p}{4 f_p^2} \tag{3.17}$$

$$\approx -\frac{x^2 \Delta f_z}{8 f_p^2}. \tag{3.18}$$

The JCMT has $f_p = 5.4$ m and $x_{max} = 7.5$ m; the CSO has $f_p = 4.2$ m and $x_{max} = 5.2$ m. Therefore, for both dishes,

$$\Delta y_{max} \sim 0.2 \Delta f_z. \tag{3.19}$$

The extra path length is $\sim 2\Delta f_z + \Delta y_{max}$, so even in this extreme case, the flexing of the primary only contributes 10% of the total. In reality the value is substantially less. The original numerical simulations of the JCMT dish predict that the path length from the primary apex to the tertiary mirror is essentially fixed. Much more significant is that when moving from horizon to zenith the whole antenna—the primary, secondary and the receiver cabin—shifts downwards by ≈0.4 mm with respect to the fixed elevation axis. This gives at worst a 180° change of phase over a period of 6 hours (for observations at a wavelength of 0.8 mm). There is likely to be a similar shift at the CSO, which will tend to cancel that at the JCMT, but I am not currently aware of its magnitude.

During the fourth run, a simple correction of $2\Delta f_z$ for each antenna was applied. We made one attempt to check this, by observing the bright maser source MWC 349 (§4.2). The phase of the strong maser line was recorded for various focus settings on each telescope and the results were in good agreement with a correction of $2\Delta f_z$.

3.1.5 Non-intersection of axes

The baseline is measured between the points of intersection of the azimuth and elevation axes at each antenna. If they do not actually intersect, then an additional error is introduced, as shown in Figure 3.5. The variable amount of extra path for one antenna is given by

$$\Delta d_{axes} = \Delta a \cos \epsilon, \tag{3.20}$$

where ϵ is the elevation. For two dishes, the resulting effect on the phase is

$$\Delta \phi_{axes} = 2\pi \frac{(\Delta a_1 - \Delta a_2)}{\lambda_{sky}} \cos \epsilon. \tag{3.21}$$

Physical measurement of Δa for the JCMT is difficult, but not impossible, and the value is thought to be less than 3 mm. It is not possible, however, to make a direct measurement at the CSO. We must therefore include these as parameters in the baseline model, and determine the value of $\Delta a_1 - \Delta a_2$ from fitting to the phase variations of astronomical sources. This is described in §3.2.

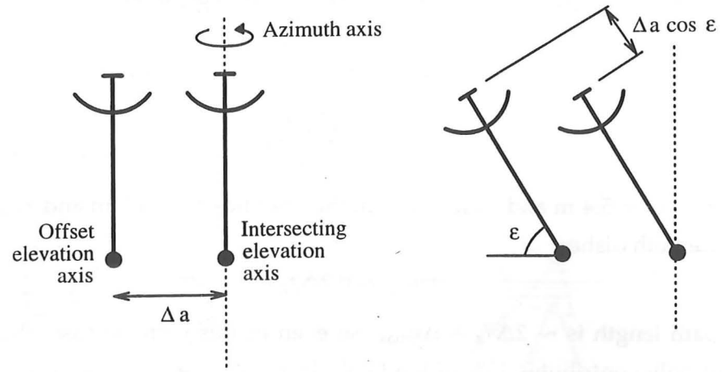


Figure 3.5: An extra path length is introduced if the elevation and azimuth axes do not intersect. The diagram shows the path offset when an antenna is at elevation ϵ instead of zenith, relative to the ideal, intersecting case.

3.1.6 Atmospheric effects

The Earth's atmosphere lies between an astronomical source and the antennas. It affects the interferometer phase in three ways:

Random phase fluctuations

These are primarily caused by the different amounts of water vapour present above the two dishes. If all of the water vapour in the atmosphere were condensed into liquid form, then for every 1 mm of liquid, the corresponding water vapour introduces an extra path of 6.3 mm (p.411 of Thompson, Moran and Swenson 1986). The amount of water vapour present above either of the two dishes varies over relatively short timescales (Masson, 1994), introducing fluctuations in the path difference and the output phase of the correlator. The weather is often the limiting factor in making observations; when the phase fluctuations are large, we are limited to working at low frequencies. In good observing conditions the phase fluctuation for a sequence of 10 s integrations is ~30° at 345 GHz. Several groups are working on systems to measure the amount of water vapour above each antenna, allowing dynamic correction of the atmospheric phase.

Refractive index

The difference in altitude of the two dishes can also cause errors in the path lengths. There would be no problem if the atmosphere had a refractive index, $n_{\rm atm} = 1$, since the

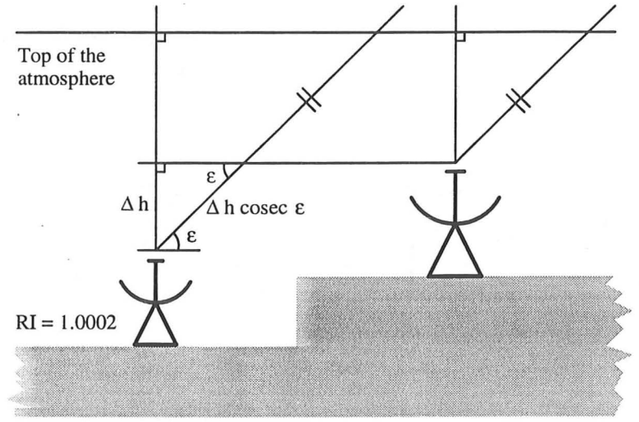


Figure 3.6: The origin of extra path from the difference in altitude of two antennas. Signals received by the lower antenna need to travel an extra distance Δh cosec ϵ through the atmosphere which has a refractive index slightly higher than unity.

difference in altitude is implicit in the three baseline components. At an altitude of 4100 m, the atmosphere has an average pressure of 59 kPa and a refractive index of $n_{\rm atm} = 1.0002$ [calculated from figures given in Kaye and Laby (1986)]. The lower telescope looks through more of the atmosphere than the higher one, by an amount $\Delta h / \sin\epsilon$ (see Figure 3.6), changing the phase between the signals by

$$\Delta\phi_{\rm ri} = 2\pi \frac{\Delta h (n_{\rm atm} - 1)}{\lambda_{\rm sky} \sin\epsilon}. \tag{3.22}$$

The JCMT is 13 m higher than the CSO, so that the path difference at zenith is about 2.6 mm; for $\epsilon = 20°$, this increases to 7 mm. This is therefore an important effect. I have not taken into account the curvature of the Earth, which will render equation 3.22 invalid for small values of ϵ.

Differential refraction

This results from the curvature of the Earth: two telescopes located at equal altitudes but different longitudes observe a given source at slightly different elevations. They therefore look through different amounts of atmosphere. The effect is much smaller than that for the difference in altitude. See p.414 of Thompson, Moran and Swenson (1986) for the derivation of the result

$$\Delta\phi_{\rm dr} = 2\pi \frac{c\tau_{\rm g}\mathcal{L}_0}{r_0\lambda_{\rm sky}} {\rm cosec}^2\epsilon, \tag{3.23}$$

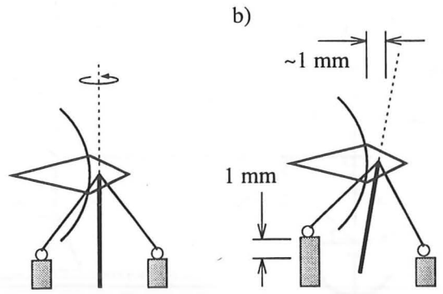


Figure 3.7: (a) The pointing and position of the JCMT dish is influenced by the track over which it rolls in azimuth. (b) Bumps in the track cause the dish to tip, changing the path length to the source, even though the pointing is corrected.

where c is the speed of light in a vacuum, τ_g is the geometrical delay, $\mathcal{L}_0$ is the excess path in the zenith direction (for one antenna, over and above the vacuum path) and r_0 is the radius of the Earth. Typically, $\mathcal{L}_0 \sim 1$ m on Mauna Kea, so for a source at an elevation of 20°, aligned with the length of the baseline, there is an extra effective path length of ≈0.2 mm.

3.1.7 Azimuth bumps and cabling

The JCMT rotates in the azimuth direction using four wheels resting on a circular track, as depicted in Figure 3.7a. Any small bumps in this track tend to tip the azimuth axis away from vertical (Figure 3.7b). Although the pointing is corrected, there is a change in the path length to the source; to first order, a 1-mm bump in the track can cause up to a 1-mm change in the path. The worst bumps are currently believed to be ~0.5 mm. The pointing accuracy of the JCMT has been carefully measured using inclinometers, and it should be fairly straightforward to translate the set of elevation pointing offsets into path offsets for the interferometer at different azimuth values. I have not yet attempted to do this. The CSO uses an enclosed bearing, which is less prone to these problems, but a similar technique should be possible for correction, if this is required.

In addition to the thermal effects that I have already discussed, the local oscillator phase-reference fibre is subjected to stresses as the antennas move. The fibre is routed from the conduit in the ground to the receiver cabins on the telescopes. The JCMT cabin is mounted on the rear of the dish and moves in both elevation and azimuth; at the CSO, the signal originates in the sidecabin, which only moves in azimuth. Cable-benders are used to make the transition from the fixed conduit to the rotating cabins. The CSO employs a

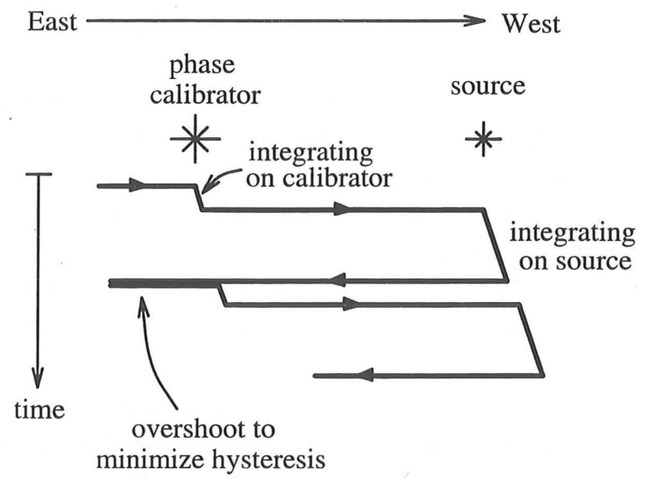


Figure 3.8: Scheme for moving the dishes between a faint source and a phase calibrator, used to minimize hysteresis in the JCMT azimuth cable turner.

simple helical bender, whereas the JCMT uses a more complicated 'hinged arm' to take up the slack. For both antennas, rotation in azimuth has a much larger impact on the phase than tipping in elevation.

I measured the variation in phase as a function of azimuth for each antenna, using the same round-trip configuration with which we examined the thermal drift (Figure 3.3). The helical bender at the CSO gives rise to a peak-to-peak variation of about 100° for operation at 345 GHz. Although large, the variation is smooth and predictable, as the cables gradually coil up into a slightly tighter configuration. In contrast, the JCMT's arm gives a peak-to-peak range of only 45°, but the variation with azimuth is quite irregular, reflecting at any given time the configuration of the many hinges. More importantly, there is hysteresis in the system; the phase at a particular azimuth position of the JCMT depends on the direction in which it is approached. The discrepancy can be up to about 10°.

I also simulated the backwards-and-forwards motion of each dish between a source and calibrator (§3.2.2), separated by 50° in azimuth, and found that the phase at each azimuth position was reproducible to within 5°. The hysteresis at the JCMT can be eliminated if we ensure that each source is always approached from the same direction; since the sources are always being tracked from east to west, we should adopt the strategy depicted in Figure 3.8.

3.1.8 Phase jumps

Having reduced the data from the second and third runs, I noticed that there appeared to be discrete jumps in the phase of the system, typically by about 90° every 10 minutes or so. They did not seem to coincide with any particular event—switching delay lines, updating the local oscillator frequency or changing the focus, for example—and I suspected problems with the transmission of the 10-MHz reference signal from the CSO to the JCMT, used to clock the DDSs and to offset the YIG and Gunn phase-locked loops. The Ortel optical link is unable to carry a frequency as low as 10 MHz. Instead, we transmit at 640 MHz and use a prescaler at the JCMT to divide by 64 to produce 10 MHz. Although we always had doubts over this aspect of the design, it performed without problems on the bench. The weakness is that if the prescalar misses a count, or counts extra because of a noise spike, then the phase of the resultant 10 MHz will either gain or slip by 1/64 of a cycle (∼6°). After all the multiplication factors in the local oscillator chains are included, these small phase slips can easily account for the much larger jumps observed.

On the fourth run, we monitored the 10 MHz signal at the JCMT and soon found that slips were actually occurring at about the same rate as the jumps had been seen. It was harder to pin down the cause, but we eventually noticed that turning the lights on near the CSO receiver cabin often coincided with a jump. It is not clear how the electrical noise was entering the system, but the power supplies to amplifiers or the Ortel transmitters is a possible route.

The temporary solution for the rest of the run was to eliminate the prescalar by imposing a 10-MHz modulation on the 640 MHz, from which a coherent 10-MHz signal could be extracted at JCMT by detecting and filtering. This is not very satisfactory in the long term and we are considering other options to avoid the problem. One idea (due to Richard Hills) is to pass a 10-MHz square wave through a high-pass filter at the CSO and transmit the resulting waveform over the link. The 10 MHz can then be recovered using a Schmitt trigger at the JCMT.

3.2 Phase calibration: correction

The contributions to the phase vs. time response that are described above can be divided into five categories. (1) The effects of the delay lines and cable lengths, dish deformation, atmospheric refractive index and differential refraction can be accounted for using relatively simple calculations and measurements, as I have already shown. (2) The exact baseline length and the offsets between the azimuth and elevation axes cannot be measured directly. They must be determined by fitting a model to the phase as a function of time for different astronomical sources. (3) The effects of temperature and mechanical stress

on the cabling cannot be determined analytically, but can be offset by observing a phase calibrator source. (4) Bumps in the azimuth tracks must be corrected using a look-up table of values for each azimuth. (5) Phase fluctuations from water vapour in the atmosphere cannot be corrected by simple means, and are currently one of the main limiting factors to the performance of interferometers at millimetre and submillimetre wavelengths.

In practice, at any given time there are contributions from all of these effects. The challenge is to disentangle them, so that each can be corrected individually by the required amount. In the first instance, the baseline is likely to be much the largest source of error for which we can get a rough estimate by ignoring thermal drifts and the non-intersection of axes.

3.2.1 Calibrating the baseline

From §3.1.1 the extra phase generated by errors in the three baseline components L_x, L_y and L_z is given by

$$\phi_{\text{base}} = \frac{2\pi}{\lambda}(\Delta L_x \cos\delta \cos H - \Delta L_y \cos\delta \sin H + \Delta L_z \sin\delta), \tag{3.24}$$

where H is the hour angle of the source and δ the declination. By observing how the phase of a bright source varies with time, we can fit for the values of ΔL_x and ΔL_y. Finding ΔL_z is harder, since it gives a constant phase offset for a given source.

I wrote a least-squares fitting program to determine ΔL_x and ΔL_y. There are several problems with this. The first is that the output of the correlator is in the range −180° to +180°, with jumps of 360° at the wrap-around. These must be removed so that the phase follows a smooth curve, over many turns if necessary. The effect of the phase jumps described in §3.1.8 is much more serious and disruptive. Clearly we cannot fit a curve of the form of equation 3.24 across such jumps. Instead, I split the data into relatively short chunks of perhaps 10 minutes' duration, within which there were no jumps, and then ignored the relative phase between each chunk. This also reduces the impact of any thermal gradient, but greatly limits the power of the fitting process. Each chunk is used to constrain the possible values of ΔL_x and ΔL_y; the fitting program finds the best-fit value for each and also calculates the appropriate error ellipse. Figure 3.9 shows an example of several such ellipses for different sources, illustrating how they can pinpoint a solution. The ellipses nominally represent the 90% regions of confidence for the fit to each period of data; that they do not all intersect shows that there are non-Gaussian sources of error. Horizontal ellipses give good constraints for L_y and come from data taken near zenith where $H \sim 0$. In my initial attempts, it proved very difficult to find a unique solution at a level of precision better than 1 cm or so.

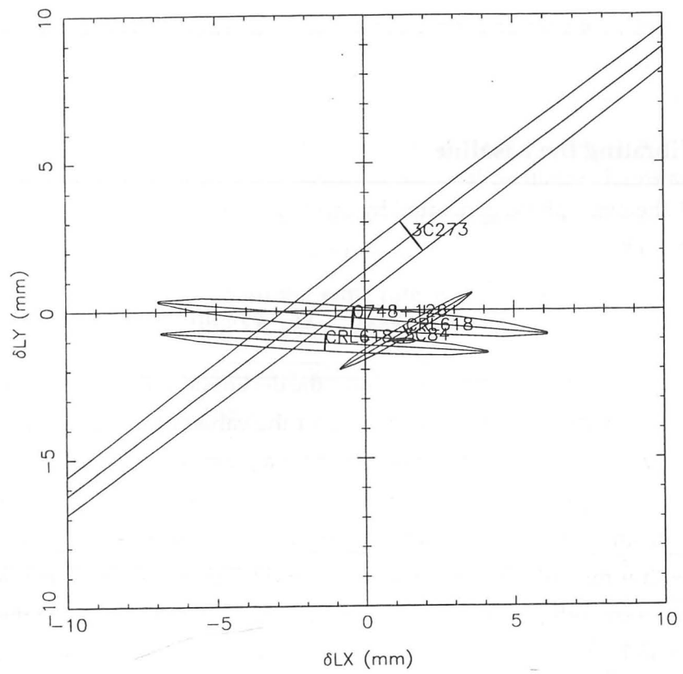


Figure 3.9: Error ellipses from five chunks of data on Night 7 of the fourth run in November 1993. Each ellipse shows the constraints imposed on the errors in the lengths of the L_x and L_y components of the baseline. The agreement here is good to within about 2 mm.

The reason for this was that the non-intersection of axes was much larger than expected. Modifying the fitting procedure to include equation 3.21, I found that the combined distance between the axes ($\Delta a_1 - \Delta a_2$) was about 6 mm. It is unclear what fraction of this is due to each antenna.

With this term included, the overall consistency of the baseline solution is much improved. Our current best values for the baseline components (the vector is from the CSO to the JCMT) are:

$$\begin{aligned} L_x &= (-1.933 \pm 0.002)\ \text{m}, \\ L_y &= (-157.914 \pm 0.002)\ \text{m}, \\ L_z &= (+43.378 \pm ?)\ \text{m}. \end{aligned}$$

I have not yet attempted an accurate determination of L_z, which requires observations of two sources with similar declinations. The values of L_x, L_y and L_z are given to the VAX at the JCMT which computes the A, B and C for each source, using equations 3.8 to 3.10. Any subsequent refinements can be accounted for with suitable phase corrections in the reduction software. Figure 3.10 shows an example of the phase variation over time for the bright source CRL 618, with the current best values for the baseline and combined offset of axes. There still appear to be problems with sources at low elevations for reasons that are not yet clear.

3.2.2 Using a phase calibrator

Regular observations of a bright phase calibrator source can be used to eliminate the effect of a slow thermal gradient, and to offset the azimuthal changes to the cabling. If the calibrator is observed every 20 minutes, then a curve can be fitted through the phase values and extrapolated to indicate the phase of the system during observations of the fainter source in the intervening periods.

We have not yet used a phase calibrator for observing, since we have only recently gained a good understanding of the baseline, offset of the axes and the phase jumps. Now that these are well-characterized, the use of a phase calibrator should become routine and will be vital for long, coherent integrations on faint sources.

3.3 Flux calibration

There are two stages to this process. The correlation levels are first converted to effective antenna temperature by multiplying by the system temperature (§4.1.6). This automatically accounts for any absorption by the Earth's atmosphere. The effective antenna temperature

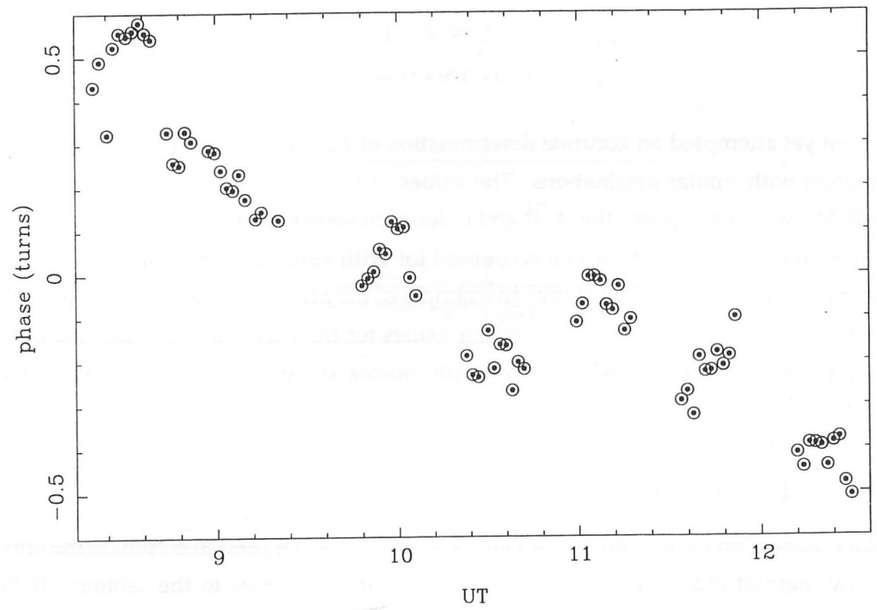


Figure 3.10: Phase vs. time (in hours) for the source CRL 618 from the fourth observing run in November 1993. There is a slow change of about 400° over 4.5 hours, and a clear jump by about 90° at UT = 10.8.

can then be converted to units of flux, using a scaling factor dependent on the collecting area and efficiency of the interferometer.

3.3.1 Measuring the system temperature

The IF signal is auto-correlated for a single-dish measurement. In this case the IF noise power is proportional to the square of its voltage:

$$P_{\rm noise,IF} \propto V_{\rm IF}^2 \tag{3.25}$$

$$\propto T^*_{\rm sys}. \tag{3.26}$$

The system temperature $T^*_{\rm sys}$ is said to be referenced to a position above the atmosphere, such that a black-body source at this temperature, located above the atmosphere and filling the beam of the antenna, generates the same power in the IF as the noise does. The system temperature can be determined by measuring the IF power with and without a hot load (a piece of absorber at the ambient temperature) placed in the beam. This is known as the two-load calibration method, and is described by Ulich and Haas (1976). The result is

$$T^*_{\rm sys} = \frac{P_{\rm sky}}{P_{\rm load} - P_{\rm sky}} T_{\rm amb}. \tag{3.27}$$

$P_{\rm sky}$ is the IF power when the antenna is pointing at a blank region of sky, $P_{\rm load}$ is the power obtained with the hot load in the beam and $T_{\rm amb}$ is the ambient temperature, usually assumed to be 280 K. The units of the IF power measurement are not important.

For the interferometer we cross-correlate two IF signals, so the effective IF noise power is given by

$$P_{\rm noise,IF} \propto V_{\rm IF1} V_{\rm IF2} \tag{3.28}$$

$$\propto \sqrt{T^*_{\rm sys1} T^*_{\rm sys2}}. \tag{3.29}$$

The effective system temperature for the interferometer is the geometric mean of the two single-dish values, which can be measured independently using the two-load method.

Some refinements are needed, however. Firstly, we are sensitive to noise entering in both the upper and lower sidebands, but only to signal in one or the other. This effectively doubles the system temperature that we must use to convert correlation into antenna temperature. In addition, if one sideband of the receiver is more sensitive than the other, then a further correction is needed. Both of these effects can be accounted for by expressions that include the relative signal gains of the two sidebands, denoted by $g_{\rm u}$ and $g_{\rm l}$. For signals in the upper sideband

$$T^*_{\rm sys,usb} = \frac{g_{\rm l} + g_{\rm u}}{g_{\rm u}} \sqrt{T^*_{\rm sys1} T^*_{\rm sys2}}, \tag{3.30}$$

and for signals in the lower sideband

$$T^*_{\rm sys,lsb} = \frac{g_{\rm l} + g_{\rm u}}{g_{\rm l}} \sqrt{T^*_{\rm sys1} T^*_{\rm sys2}}. \tag{3.31}$$

There is another potential complication. If observations are made at a frequency close to one of the absorption lines of the Earth's atmosphere, then one sideband is likely to suffer more absorption than the other and have a higher system temperature. To correct for this effect we need a model for the profile of the absorption line. The 321-GHz VY CMa H_2O maser observations that I discuss in chapter 5 occur close to the corresponding H_2O line in the atmosphere; according to Karl Menten, a 10% correction factor is necessary [see also Menten and Melnick (1990)].

The system temperatures for each antenna and in each subband are measured before every 100 s of integration. Hot loads are inserted into the two beams and the IF powers in each subband are recorded by total power detectors in the correlator. The geometric mean for each subband is found and used to convert the correlation levels to antenna temperatures. Typical system temperatures at 345 GHz during the fourth run were 1000 K for the JCMT and 500 K for the CSO, giving a mean of ~700 K.

3.3.2 Antenna temperature to Jansky

For a single-dish telescope of effective receiving area A, observing in one polarization with efficiency[1] η, the flux S, and antenna temperature $T_{\rm A}$, are related by

$$\frac{1}{2}\eta S A = k T_{\rm A}, \tag{3.32}$$

where k is the Boltzmann constant. For an interferometer, A must be replaced with the geometric mean of the areas for the two dishes, $\sqrt{A_1 A_2}$. The mean collecting area for the 15-m JCMT and the 10-m CSO is 118 m^2. If the efficiency is 50% then we predict that $S/T_{\rm A} \approx 50$ Jy K^{-1}.

In reality, many things can increase this value (for example, short-term phase fluctuations of the atmosphere or the local oscillators), so it must be determined by observation. A flux calibrator, usually a quasar, is observed with both the interferometer and a single dish. The single-dish quasar flux is calibrated against observations of planets, whose fluxes are relatively well-established. Flux calibrators should be observed as often as possible, to account for changes in the sensitivity over time, and at the same frequency and sideband as the observations to be calibrated. There is no need to determine the difference in gain for each sideband (the $g_{\rm u}$ and $g_{\rm l}$ of the previous section) as they are included automatically.

[1]This includes terms to account for the non-ideal mirror surface, scattering losses and the correlator efficiency, etc.

The uncertainty associated with flux calibration is large (±20% is typical) and this should be kept in mind whenever fluxes are quoted.

We have yet to achieve a sensitivity approaching 50 Jy K^{-1}. One reason for this was the problem with the phase-switching (§2.8) that effectively lost one third of the signal. On the fourth run, the sensitivity was determined to be 140 Jy K^{-1}; fixing the phase-switching then gave about 90 Jy K^{-1}, still well short of the target. Filters and alignment problems at the JCMT are suspected. This whole issue will have to be addressed at the start of the next run by carefully measuring the efficiencies of the single dishes.

3.4 Passband calibration

To ensure that the spectrum generated from the data is as faithful as possible to the spectrum of the source, the effects of the instrument on the signals must be removed. Our interferometer measures a complex spectrum; the response of the system is also complex, with an amplitude and a phase that vary as a function of frequency. All components through which the signals pass will make some contribution to this passband response, but the dominant source of variation is the input stage of the correlator, where the signals are split into eight subbands using mixers and filters (§2.2.2). We would expect that the other parts of the system have amplitude and phase responses that only vary slowly as a function of frequency. There are several different ways to measure the passband response.

3.4.1 Using a quasar

The flux and phase of a quasar should both be independent of frequency over the 1-GHz bandwidth that we are interested in, and quasars are therefore ideal for determining the passband. The raw output spectrum, once corrected for any delay errors, should be the passband that we require. The quasar should be observed at the same frequency and sideband as the main object under observation.

The complex spectrum of the main source is divided by the passband (§4.1.5). Having carefully integrated on the main source over a long period to achieve the required signal-to-noise ratio, it is important not to degrade it by dividing by a noisy passband. The ratio of the signal-to-noise of the passband to that of the main observation should be as high as possible, but depends in practice on the relative strengths of the two sources. This is not a problem for faint sources, where we only need spend a small fraction of the total time measuring the passband, but it is restrictive for stronger sources, especially very bright line sources where the flux in a narrow range of frequency can easily exceed the brightest quasars. In this case we should spend more time observing the quasar than we do for the spectral line source. Masers are a case in point, as illustrated by the 321-GHz H_2O maser

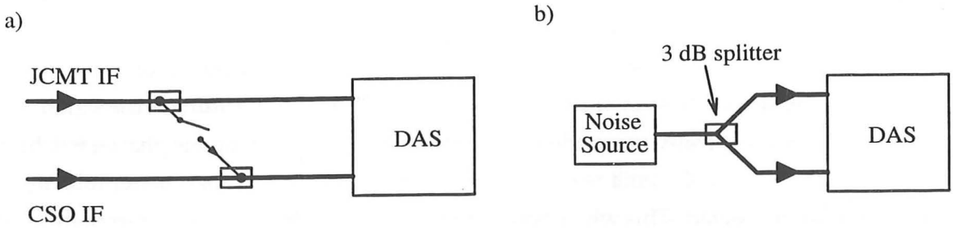


Figure 3.11: Two methods for measuring the passband of the correlator, using (a) cross-coupling between the IFs and (b) a common noise source for both IFs.

in VY CMa (chapter 5) and the hydrogen recombination-line maser in MWC 349 (§4.2). We can overcome the need for long periods of integration on a quasar by using a technique that measures the passband in two stages.

3.4.2 Using a quasar and cross-coupling the IFs

A quasar is used as above to determine the broad shape of the passband, but the finer detail is found by cross-coupling one of the IFs into the other just before the correlator. The spectrum of the fine detail is then added to the broad shape to give the overall passband.

The fundamental assumption is that only the correlator introduces fine structure into the passband, on scales smaller than, say, $n_{\rm smooth}$ channels. I denote the correlator passband by $\mathbf{C}(\nu)+\mathbf{c}(\nu)$; both terms are complex, with $\mathbf{C}(\nu)$ representing structure larger than $n_{\rm smooth}$ channels and $\mathbf{c}(\nu)$ representing anything finer. Remaining contributions to the passband (i.e. from all the components in the system before the correlator) are given by $\mathbf{G}(\nu)$. We wish to find the overall passband, given by the product $\mathbf{G}(\nu)[\mathbf{C}(\nu)+\mathbf{c}(\nu)]$. The quasar spectrum, once smoothed over $n_{\rm smooth}$ channels by convolution with a Gaussian (§4.1.8), is a good approximation of $\mathbf{G}(\nu)\mathbf{C}(\nu)$. The smoothing increases the signal-to-noise ratio of the quasar spectrum by a factor of $n_{\rm smooth}^{0.5}$; the integration time required to achieve a given signal-to-noise ratio for the passband is therefore reduced by the same factor.

The IFs are cross-coupled using the set-up shown in Figure 3.11a. It should also be possible to use the arrangement in Figure 3.11b, although we have not yet tried this. The spectrum obtained from cross-coupling, which does not have to be corrected for delay error, is given by $\mathbf{F}(\nu)[\mathbf{C}(\nu)+\mathbf{c}(\nu)]$, where $\mathbf{F}(\nu)$ is the cross-spectrum of the noise before it enters the correlator. Again, we must include the proviso that there is no fine structure in $\mathbf{F}(\nu)$, so that smoothing over the same number of channels as we did for the quasar gives $\mathbf{F}(\nu)\mathbf{C}(\nu)$. Subtracting this from the unsmoothed version leaves $\mathbf{F}(\nu)\mathbf{c}(\nu)$. Dividing the last two results eliminates $\mathbf{F}(\nu)$, and leaves the ratio $\mathbf{c}(\nu)/\mathbf{C}(\nu)$. Multiplying this by

$\mathbf{G}(\nu)\mathbf{C}(\nu)$ (the smoothed quasar spectrum) generates $\mathbf{G}(\nu)\mathbf{c}(\nu)$, which, when added to $\mathbf{G}(\nu)\mathbf{C}(\nu)$, gives $\mathbf{G}(\nu)[\mathbf{C}(\nu)+\mathbf{c}(\nu)]$. This is the overall passband.

This procedure automatically ensures that the spectra of the broad shape and the fine structure have the correct relative scaling before they are added together. Very little integration is needed in the cross-coupled configuration, since the two signals are highly correlated. Also, the fine structure—assuming that it does all come from the correlator—is independent of the observing frequency and sideband, and only has to be measured once for each of the correlator configurations.

I have not yet employed this technique for observations that we have made, but it will become the method of choice in the future, particularly for brighter sources. The assumption about the correlator fine structure should be tested.

3.4.3 The UKIRT test transmitter

We built a test transmitter, primarily for the purpose of testing and debugging the interferometer, and located it on the roof of the UKIRT observatory, approximately 800 m to the east of the JCMT. The design is very similar to that of a local oscillator, with the output of a phase-locked Gunn diode multiplied up to a frequency in the submillimetre wave-band. This signal, which is essentially monochromatic, is propagated in the general direction of the JCMT and the CSO in the valley below, using a small horn. If desired, a broad-band signal can be produced by either rapidly sweeping the frequency of the Gunn, or by injecting noise into the bias port of the multiplier.

This is not a flat spectrum source and is of little use for measuring the amplitude passband. But as a point source, it might be suitable for determining the phase passband of the interferometer. Unfortunately, there are saturation problems with the high instantaneous signal strength, and it is also inconvenient to to go up onto the roof of the UKIRT every time that a re-tune in the frequency is needed. The phase also depends on frequency to a small extent, because the source is in the near-field of the antennas; by going through the relevant Fresnel diffraction calculations, I estimated that the phase across a 1-GHz passband varies by up to 10° when observing at 345 GHz.

3.4.4 A new technique

Whilst trying to devise accurate ways of measuring the passband of the interferometer, I came up with an attractive technique for providing each antenna with a common, strong, flat-frequency source in the far-field. Figure 3.12 explains the idea. Both dishes are made to point exactly at one another. Energy from a thermal load at antenna 1 is propagated towards a partially reflecting plane mirror, M1, that divides the signal into two correlated

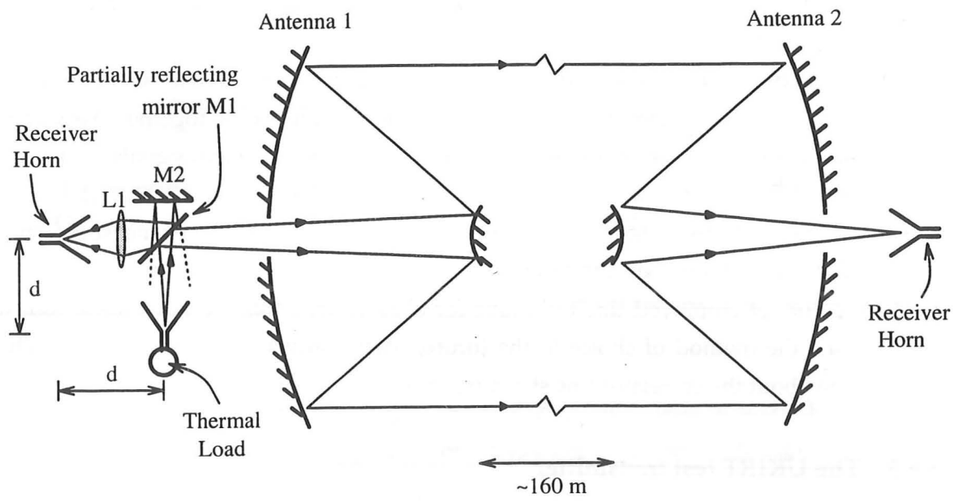


Figure 3.12: A new technique that can be used for passband calibration and debugging. The thermal load is imaged into the far-field of both antennas, presenting a bright, point source, common to both receivers.

components. One of these is reflected, via the primary and secondary mirrors of each antenna, to the receiver horn of antenna 2. If the two horns at antenna 1 are equidistant from the partially reflecting mirror, then the thermal load should be imaged onto the receiver horn of antenna 2, just as though antenna 2 is observing a source in the far-field.

The other component is transmitted through M1 on the first pass, whereupon it is reflected by plane mirror M2 back onto M1. Again the signal is split; one part is transmitted and lost and the other is reflected onto lens L1 which is chosen to focus the remaining energy onto the receiver horn of antenna 1. Again, the thermal load appears to be in the far-field. If the lengths of the two possible signal paths to the correlator are equalized (using delay lines and software correction) then the output of the correlator will be a very good approximation to the passband; it can be corrected for the shape of the black-body spectrum if desired. The one assumption is that M1, M2 and L1 do not introduce any significant structure to the passband.

It is possible to show that the optimum reflectivity for M1 is 67%, and if the thermal load horn is well-matched to the two receiver horns then the effective antenna temperature from a thermal load at temperature T_L will be $0.39T_L$. For $T_L = 300$ K, this will be almost 120 K. Any spurious reflections in the system will be at very much lower strength than the main signal and should not be a problem; the lens L1 helps to isolate the receiver horn of

antenna 1 from potentially troublesome reflections from the horn of antenna 2. M1 and L1 would be switched into the beam path when needed.

Although a nice idea in theory, there is a major problem in practice: it is not possible to point the JCMT at the CSO because of the large difference in altitude between the two antennas. It is not even possible to get a sizable overlap between the areas of the two dishes (we only require that the optical axes are parallel). It may, however, prove useful for other interferometers.

3.5 Summary

In this chapter I have covered a wide range of topics related to the calibration of the JCMT–CSO Interferometer.

The phase response as a function of time is now determined well enough to allow long, coherent integrations on faint sources, using a phase calibrator as a reference. In reaching this stage, I have determined the lengths of the delay lines and cabling, and measured the effects of azimuth changes and thermal drifts on the local oscillator reference fibre. I have also compensated for the focusing action of each antenna, the difference in altitude of the two dishes and the effect of differential refraction by the atmosphere. From observations of astronomical sources I have been able to determine accurately the L_x and L_y components of the baseline, and the combined offset between the azimuth and elevation axes of each dish, which turned out to be larger than expected.

I then showed how we determine the system temperature using a two-load method, and allow for different sideband sensitivities and atmospheric absorption. A flux calibrator must be observed to convert the antenna temperatures into units of flux.

Finally, I discussed techniques for measuring the passband of the interferometer. These included a promising new method that determines the broad shape and fine structure in the passband separately, before combining them to produce the overall passband. This reduces the amount of integration time needed on the passband calibrator source. I also presented a novel configuration of the dishes that imitates a thermal source in the far-field of both antennas, allowing the passband to be measured directly in a short time. Although this set-up cannot be used on our interferometer, it may be relevant to other instruments.

Chapter 4

DATA REDUCTION

This chapter describes the steps by which raw data are processed and added together to produce the final spectra. I explain the format of the raw data and the different stages in the software reduction, and conclude with observations that we made of the hydrogen recombination-line source MWC 349, the Galactic Centre object Sgr A* and the nearby radio-galaxy, Centaurus A. These three examples were chosen both to illustrate different data-reduction techniques and for their intrinsic scientific interest.

4.1 The data-reduction process

4.1.1 The raw data

The usual output of the correlator consists of two sets of 2048 measurements of the cross-correlation function, divided into as many as 8 different subbands. There is one set of values for each of the two phase-switchings (0° and 180°). The JCMT VAX computer performs a Fourier transform to give two sets of 1024 complex numbers, each corresponding to a particular frequency in a subband. The number of subbands, and the frequency range that each occupies, depend on the correlator configuration (§2.2.3).

The scaling of the raw data deserves an explanation. The correlator output is normalized for each subband, such that if two identical, fully-correlated waveforms enter the correlator, then the sum of the amplitudes over each subband is the same as the number of channels in that subband. The average amplitude of a fully-correlated signal is therefore unity. Almost all of the IF power in the system will be due to noise, with only a small fraction from the astronomical source being observed[1]. This is important when we come to scale the data to units of effective temperature.

The complex data are stored as sine and cosine components in a file with the JCMT

[1] the sun, moon and major planets are exceptions to this

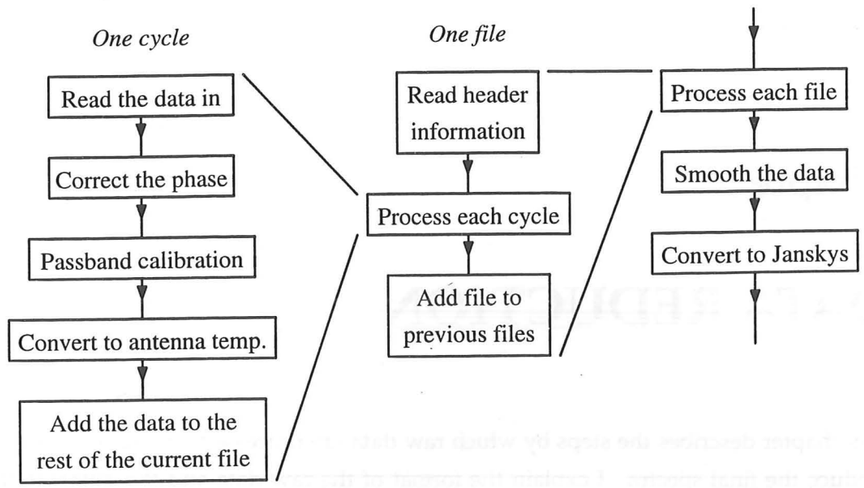


Figure 4.1: The basic steps in the data-reduction process

standard GSD format. Each file holds the results of ten integrations with a common header that contains a large number of parameters describing the observation. The date, source location, local oscillator frequency and details of the correlator configurations are just some examples of the wealth of information included by default, much of which is not directly useful for interferometry. A communications link between the CSO and the JCMT was established (C. Mayer, K. Young and F. Oliveira were responsible) so that the status of the CSO, and particularly the local oscillator system, could also be recorded in the GSD files.

4.1.2 INTEREDUCE

The software to reduce the data for the first interferometry run in May 1992 was written by Chris Mayer. This program was useful initially but more flexibility was soon needed to accommodate the ever-expanding number of routines. I therefore spent some time writing INTEREDUCE, a new data-reduction program incorporating some of the original code. It is written in FORTRAN and makes use of Rachael Padman's GENLIB library of routines to provide a flexible, command-driven structure and the ability to execute macros. There is a number of commands available, each of which performs a specific operation on the data. They may be typed in interactively by the user, or read from macro files containing lists of commands to be executed sequentially. There are currently 78 commands, and new ones are easily added when necessary. Figure 4.1 shows the main stages of data reduction, each of which is briefly described below.

4.1.3 Reading in the data

The raw sine and cosine data for the 180° phase-switch are subtracted from the 0° phase-switch data and then divided by two to preserve the normalization of the correlation. The components are then converted to the equivalent amplitude and phase (in the range −180° to +180°) for each of the 1024 channels. Some operations carried out on the data—the vector addition of two spectra, for example—require the sine and cosine components, whereas others need the amplitudes and phases. Arrays for each of these four quantities are maintained at all times, each with 1024 elements. INTEREDUCE can handle up to five different spectra at a time; to do this it has five registers, each consisting of four arrays holding the sines, cosines, amplitudes and phases of a spectrum. There are other arrays, including ones to describe the frequency of each channel (as an offset from the central observing frequency), the relative weighting of each channel (used for smoothing, since some channels have a higher signal-to-noise than others) and the blanking (channels at the edges of the subbands can be blanked out, for example). The frequency array is assigned on the basis of the information in the header of each GSD file.

4.1.4 Phase correction

The phases for each 10 s integration must be corrected for three effects: firstly the phase-gradient across the band due to the delay error, secondly the errors arising from things like the incorrect baseline length, the non-intersection of the azimuth and elevation axes and the effects of focusing the two dishes as the elevation changes (chapter 3), and thirdly the phase structure of the instrument passband. The latter is discussed in the next section.

Recall from the first chapter (equation 1.39) that the delay error, $\tau_{\rm err}$, resulting from the inexact compensation of the geometrical delay, generates a phase term of

$$\phi_{\rm delay} = \pm 2\pi \Delta\nu \tau_{\rm err}, \tag{4.1}$$

where the plus sign is appropriate for the upper sideband and the minus sign for the lower. $\Delta\nu$ is the offset from the centre frequency on the sky ($\nu_{\rm sky0}$ in the notation of chapter 1), so that a signal at 345.6 GHz has $\Delta\nu = +100$ MHz when the centre frequency is 345.5 GHz. INTEREDUCE calculates $\phi_{\rm delay}$ for each channel, as well as the contributions from the other errors described in chapter 3. It then adjusts the phases of the data accordingly, using simple addition or subtraction.

4.1.5 Passband calibration

The need for calibration of the passband was discussed in §3.4, which also described practical ways to measure it. This complex passband must be divided out of the data.

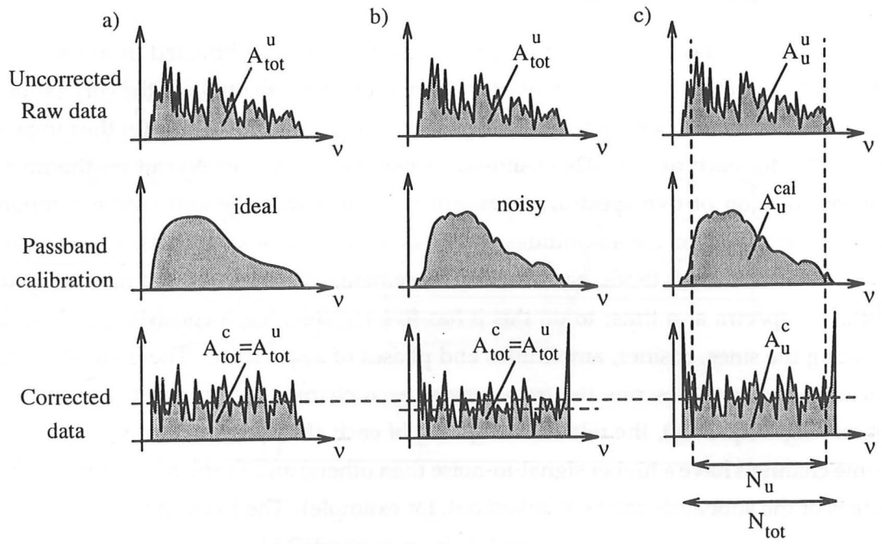


Figure 4.2: Getting the amplitude normalization right: (a) The ideal case with no noise on the passband. (b) Noise on the passband causes spikes at the subband edges, depressing the average level in the central region. (c) The correct level is restored by blanking out the edges and accounting for the shape of the passband (see main text). Superscripts: u = uncorrected, cal = passband calibration, c = corrected; subscripts: tot = total width, u = unblanked width.

The phase is simpler to deal with than the amplitude. When two complex numbers are divided, the phases are subtracted; INTEREDUCE subtracts the passband phase from the phase of the data, channel by channel. The phases must then be brought back into the range $-180°$ to $+180°$.

Figure 4.2a shows how calibration of the amplitude variation works for an ideal case, where the passband is accurately known. The data depicted represent a bright continuum source, whose flux is independent of frequency, but I have also included some noise. After division by the passband, the corrected amplitudes are scaled so that the area under the curve is the same as that for the uncorrected data.

In practice (Figure 4.2b), the passband measurement includes some noise. This is a problem at the edges of each subband; large spikes are generated in the corrected spectrum where noise has taken the passband curve down to very near the zero level. These bad channels can be discarded, but care must be taken to maintain the normalization. We want the same level of amplitude as for the ideal case (denoted by the long dashed line). If,

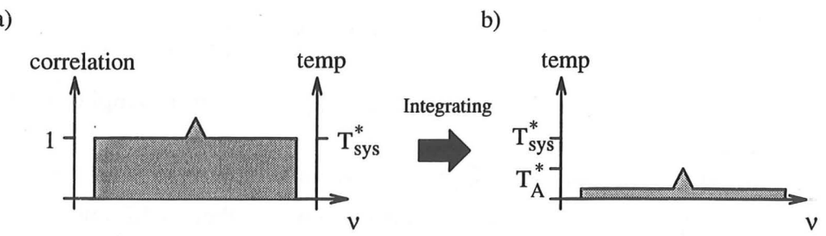


Figure 4.3: The conversion of correlation level to antenna temperature. (a) Fully-correlated noise has an average correlation of 1. This corresponds to the system temperature, T^*_{sys}. An astronomical signal (the line at the centre of the band) makes only a very small contribution. (b) The uncorrelated noise averages towards zero with integration, leaving the signal with a peak effective antenna temperature of T^*_A.

however, we simply scale to equate the areas as before, the spurious spikes depress the amplitude of the flatter section (short dashed line). To avoid this problem, I adopted the technique illustrated in Figure 4.2c. The edges of each subband are excluded by blanking off the specified number of channels. When the correct normalization has been applied, the area A^c_u of the unblanked region for the corrected data is different to that for the corresponding region of the uncorrected data, A^u_u. The area A^c_u is approximately a fraction N_u/N_{tot} of the total area for the ideal case, A^c_{tot} (Figure 4.2a), which in turn is equal to the total area under the uncorrected curve, A^u_{tot}. N_{tot} and N_u are the total and unblanked number of channels in the subband. Hence the ratio of A^u_u to A^c_u is given by

$$\frac{A^u_u}{A^c_u} \simeq \left(\frac{A^u_u}{A^u_{tot}} A^u_{tot}\right)\left(\frac{N_u}{N_{tot}} A^c_{tot}\right)^{-1} \tag{4.2}$$

$$\simeq \left(\frac{A^{cal}_u}{A^{cal}_{tot}}\right)\left(\frac{N_{tot}}{N_u}\right). \tag{4.3}$$

Owing to variations in the receiver sensitivities across the band, the noise level does not necessarily reflect the shape of the passband. Since the passband calibration curve should have a higher signal-to-noise than the data, it is this that is actually used for evaluating the ratio of the areas. The amplitudes after division are scaled appropriately to ensure the correct ratio of the unblanked areas.

4.1.6 Scaling the data to effective antenna temperature

If the system noise in the two IFs were fully correlated, then by this stage, on average, we would have the rectangular passband shown in Figure 4.3a, with a correlation level of

unity. The signal for normal astronomical sources would make a negligible contribution to the area under this curve—on the diagram I have included a spectral line at the centre of the subband. The system noise can be characterized by a temperature, $T^*_{\rm sys}$, defined in §3.3.1. INTEREDUCE multiplies the correlation by $T^*_{\rm sys}$ to convert the amplitudes to a scale of effective antenna temperature.

The noise should of course be completely uncorrelated, and its contribution to a spectrum can be reduced both by adding many integrations together (§4.1.7) and by smoothing over a number of channels (§4.1.8). In Figure 4.3b the average level of the noise has been lowered by integration. The spectral line, which is correlated in the two IFs, is the same size as before, with a peak effective antenna temperature of $T^*_{\rm A}$.

What is the meaning of $T^*_{\rm A}$? It is easiest to approach this problem by first considering a single-dish telescope. If a uniform thermal source of temperature $T_{\rm s}$ fills the beam of the antenna, then thermodynamic arguments can be used to show that $T^*_{\rm A} = T_{\rm s}$ [see, for example, Rohlfs (1986)]. This is true generally, but the remaining arguments only hold in the Rayleigh–Jeans limit, where flux is proportional to temperature. If the source subtends solid angle $\Omega_{\rm s}$ that is insufficient to fill the solid angle $\Omega_{\rm b}$ of the beam, then $T^*_{\rm A} = \frac{\Omega_{\rm s}}{\Omega_{\rm b}} T_{\rm s}$. The factor $\Omega_{\rm s}/\Omega_{\rm b}$ accounts for the beam dilution.

There are two complications for an interferometer. Firstly, some of the flux may be resolved out by the interferometer (Figure 1.4), so that only a fraction f of the total flux is detected. Secondly, the primary beam solid angle of the interferometer is a combination of those for the constituent dishes. In fact, the resultant beam size is given by $\frac{1}{\Omega_{\rm b}} = \frac{1}{2}\left(\frac{1}{\Omega_{\rm b1}} + \frac{1}{\Omega_{\rm b2}}\right)$. The effective antenna temperature for an interferometer is therefore

$$T^*_{\rm A} = \frac{f\,\Omega_{\rm s}}{2}\left(\frac{1}{\Omega_{\rm b1}} + \frac{1}{\Omega_{\rm b2}}\right) T_{\rm s}. \qquad (4.4)$$

4.1.7 Summing the data

The aim here is to ensure that the signal is added coherently, with all the vectors pointing in the same direction, as shown in Figures 4.4a and 4.4b. The resultant vector can then be used as an unbiased estimate of the signal. In this case,

$$\mathcal{R} \propto (\Delta\nu t)^{0.5}, \qquad (4.5)$$

where $\mathcal{R}$ is the signal-to-noise ratio after integrating for time t over a bandwidth $\Delta\nu$. To achieve this ideal, the phase vs. time response of the system needs to be known accurately at all times. In practice, thermal drifts, the changing atmosphere and other uncalibrated aspects of the instrument ensure that the phase of the signal vector from a point source varies with time (Figure 4.4c).

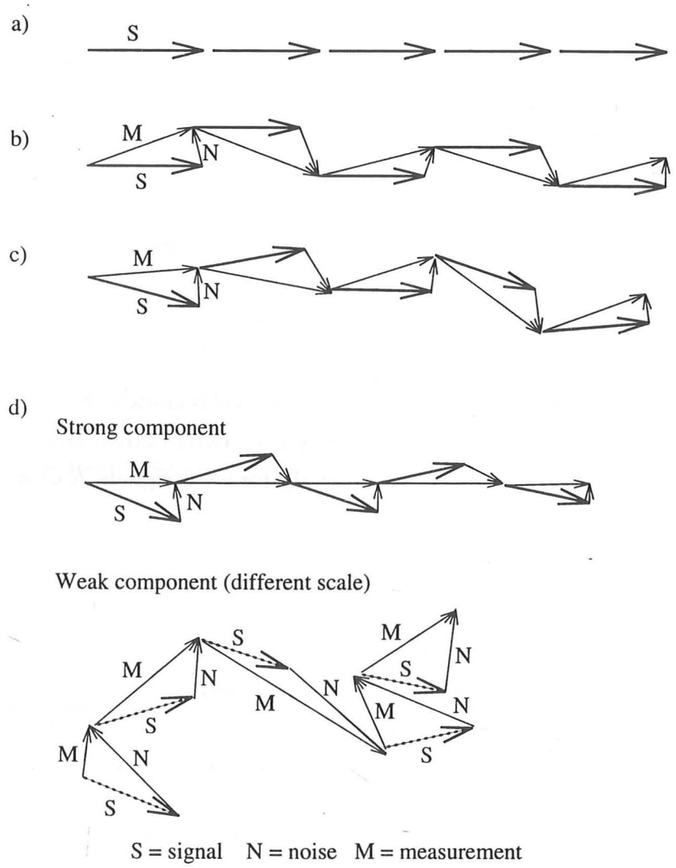


Figure 4.4: A schematic representation of different methods for summing the data. (a) Vector addition without noise or atmospheric phase fluctuations. (b) Vector addition with noise, but without atmospheric phase fluctuations. (c) Vector addition with noise and atmospheric phase fluctuations. (d) Phase referencing by adding the resultant vectors of a strong signal in a straight line, head-to-tail. The vectors of a weaker component are shown in the lower half, with signal vectors in the same orientation as those of the strong signal.

There are three different techniques available to combat this: incoherent addition of the data, use of a phase reference feature in the source being observed, or regular observation of a separate phase calibrator source.

Incoherent addition

Incoherent addition uses only the amplitude of the data and ignores the phase, and is therefore insensitive to phase fluctuations in the atmosphere or drifts of the instrument. An observation generates a vector M, comprising the signal S and noise N. We have

$$\mathrm{M} = \mathrm{S} + \mathrm{N} \tag{4.6}$$

$$\Longleftrightarrow \mathrm{M}\cdot\mathrm{M} = \mathrm{S}\cdot\mathrm{S} + \mathrm{N}\cdot\mathrm{N} + 2\mathrm{S}\cdot\mathrm{N} \tag{4.7}$$

$$\Longleftrightarrow \langle M^2\rangle = S^2 + \langle N^2\rangle \tag{4.8}$$

$$\Longleftrightarrow S^2 = \langle M^2\rangle - \langle N^2\rangle. \tag{4.9}$$

The angled brackets denote averaging. If the noise level is known, then we can calculate an unbiased estimate of the signal amplitude. There are two disadvantages to this approach: firstly, all the phase information is lost. Secondly, for weak signals, $\mathcal{R}$ does not improve as rapidly as it does for coherent addition.

Internal phase referencing

When the source has a bright component, either a strong spectral line or continuum, then the data may be added semi-coherently as shown in Figure 4.4d. When $\mathcal{R}$ is high, the phase fluctuations of the measurement vector are dominated by the atmosphere, rather than by the effect of the noise vectors. We therefore assume that the phase of a measurement M is a good estimate of the phase of the signal S. In doing so, we lose the component of S that is normal to M; it is simple to demonstrate that the fractional loss is $1/2\mathcal{R}^2$.

On its own, this method is little different from the incoherent addition described above. However, if there is also a second, weaker component in the signal at a different frequency, then this component will be summed with the same level of coherence as the stronger signal.

Using an external phase calibrator

The phase response of a well-calibrated interferometer is a slowly varying function of time, and can be estimated by fitting a smooth curve through the phases of a bright calibrator source observed at regular intervals (§3.2.2). Observations made on a faint source in the intervening periods can be added coherently once the drift in the phase is subtracted out.

If the atmosphere introduces a small, random phase variation with rms amplitude $\Delta\phi_{\mathrm{atm}}$, then a fraction $1/2\Delta\phi^2_{\mathrm{atm}}$ of the signal is lost.

The optimum strategy

In practice, the best method for integrating the data is a compromise dependent on the rms level of atmospheric phase fluctuations, $\Delta\phi_{\mathrm{atm}}$, the instrumental drift in time t, $\Delta\phi_{\mathrm{inst}}$, and the signal-to-noise ratio $\mathcal{R}$ after time t.

Simple vector addition should be used over short timescales until $1/\mathcal{R}$ becomes less than either $\Delta\phi_{\mathrm{atm}}$ or $\Delta\phi_{\mathrm{inst}}$. If $\mathcal{R} \gg 1$ at this stage then internal phase referencing should be used for further addition. Otherwise the data should be added incoherently. A phase calibrator is used to decrease the effective $\Delta\phi_{\mathrm{inst}}$.

For most of the observations made so far, the interferometer has not been particularly well-calibrated, but was stable over short time intervals of 100 s. For almost all of the data that I have reduced, I have applied simple vector addition over 100 s periods, followed by phase referencing over longer intervals. We have not so far made use of phase calibrator sources, but this will become more important for future observations of faint sources.

4.1.8 Smoothing the data

Smoothing is used to increase the signal-to-noise ratio of a spectrum at the expense of the frequency resolution. Smoothing over four channels, for example, improves $\mathcal{R}$ by a factor of two (equation 4.5). INTEREDUCE has three different smoothing routines: simple binning, boxcar average and Gaussian average. The effect of each is shown on the sample spectrum of Figure 4.5. The channels can also be weighted to account for the varying sensitivity over each subband. There is no smoothing across subband boundaries, and the first and last $(N-1)/2$ channels within each subband should be discarded (for these channels there are fewer than N contributions to the average).

There is also a routine to find the vector average over a specified range of channels, which can span more than one subband. Weighting by a factor of $(T^*_{\mathrm{sys}})^{-2}$ allows for the different system temperatures of each subband.

4.1.9 Conversion from antenna temperature to Jansky

The antenna temperatures are multiplied by the factor discussed in §3.3.2 to convert to units of flux that are independent of the sensitivity of the interferometer.

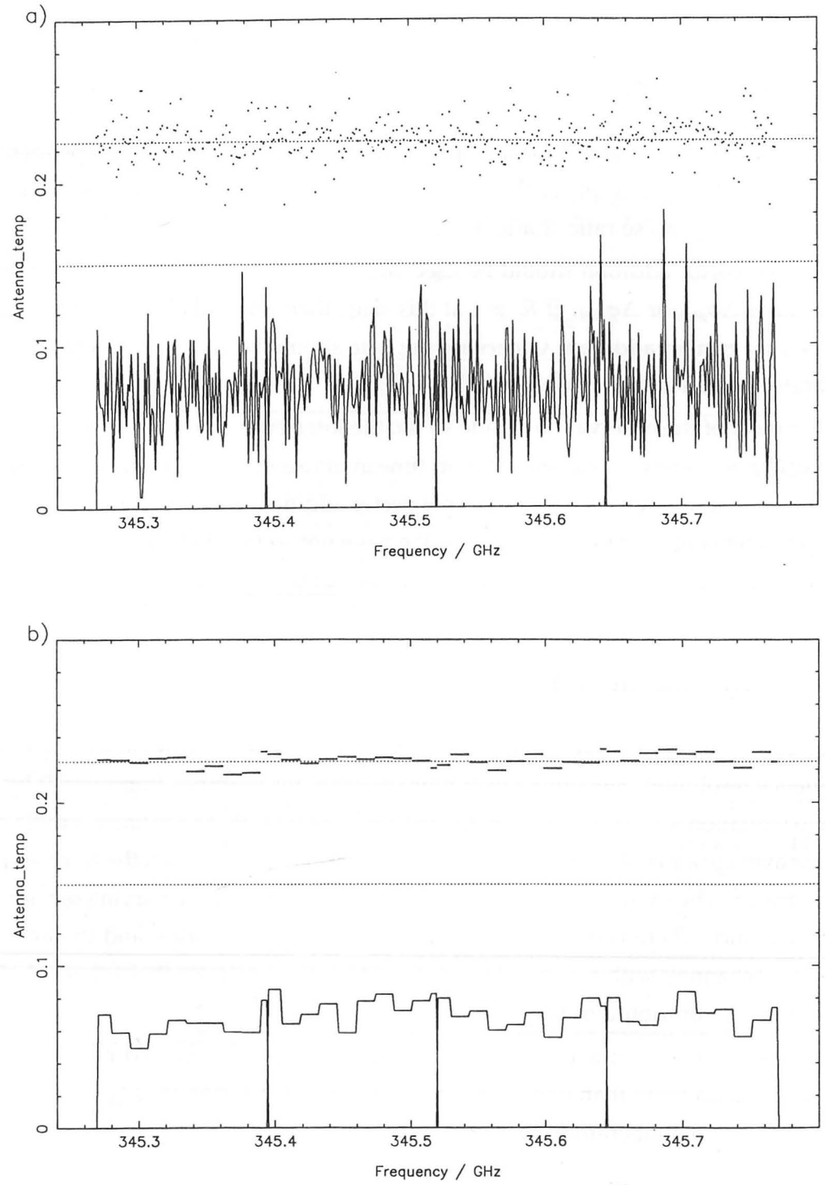


Figure 4.5: Different ways to smooth a spectrum: (a) Unsmoothed, complex spectrum of the quasar 3C273. (b) Binning the data into 11 channels. Continued on next page. . .

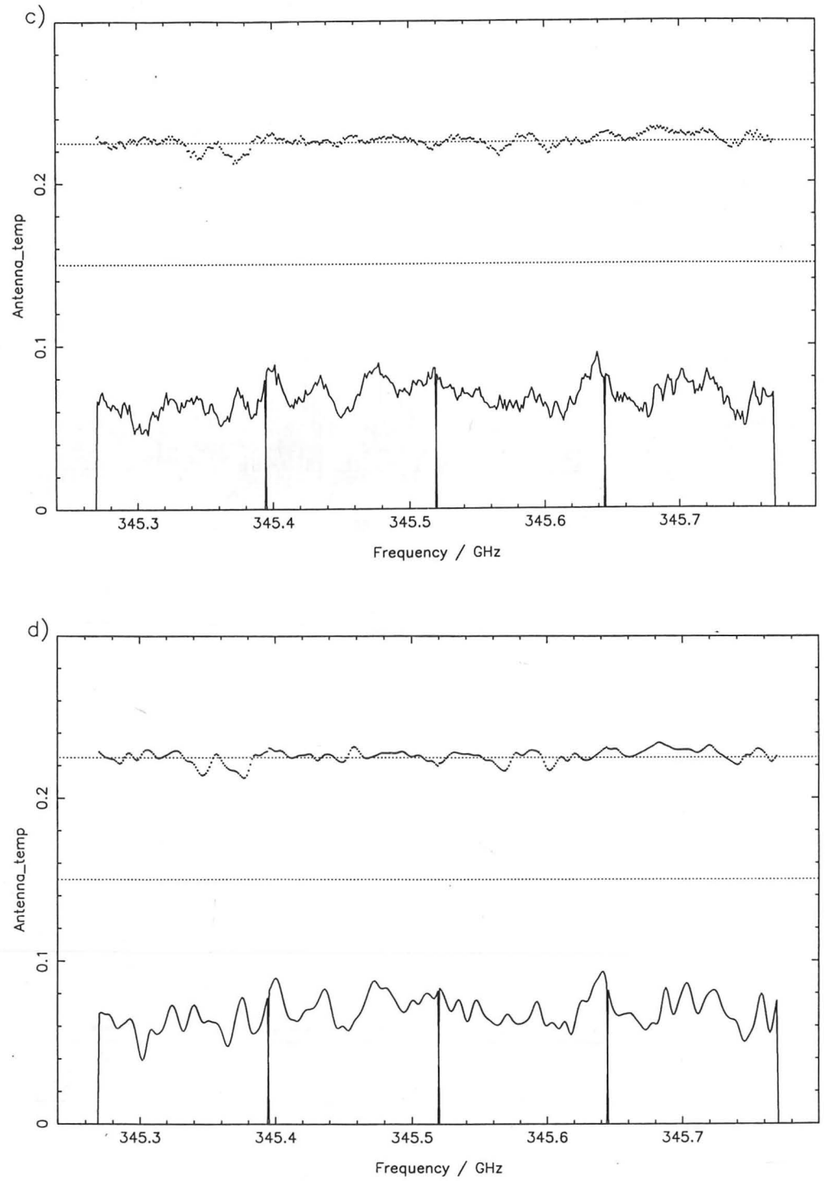
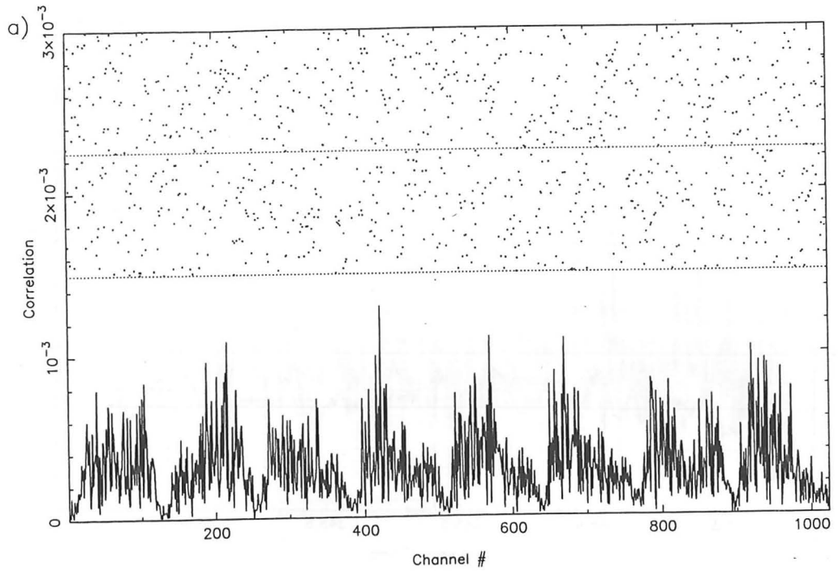


Figure 4.5: . . . continued from previous page. (c) Boxcar average over 11 channels and (d) Gaussian smoothing over an effective width of 11 channels.



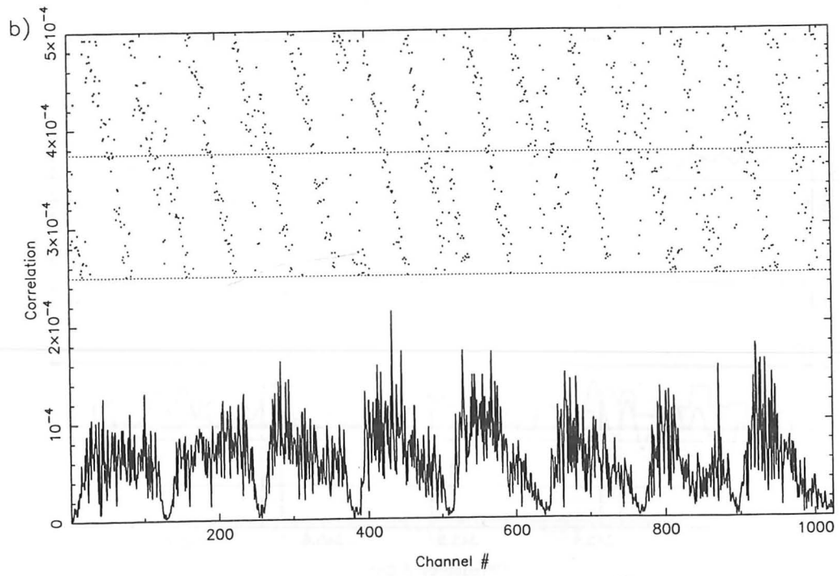
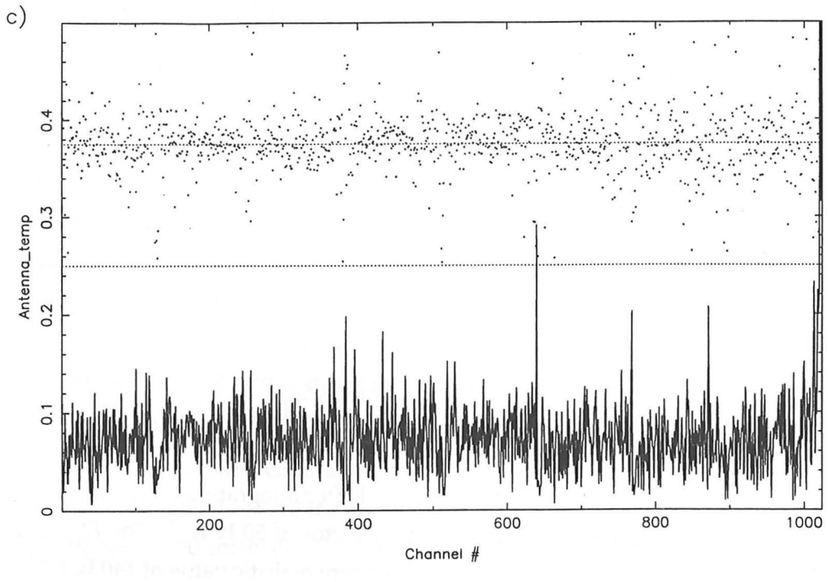


Figure 4.6: Reducing data for the quasar 3C273. (a) Raw amplitude and phase data from one integration of 10 s. (b) Effectively the same as (a), but more data have been used to emphasize the passband shape and the phase-gradient due to the delay error. Continued on next page...



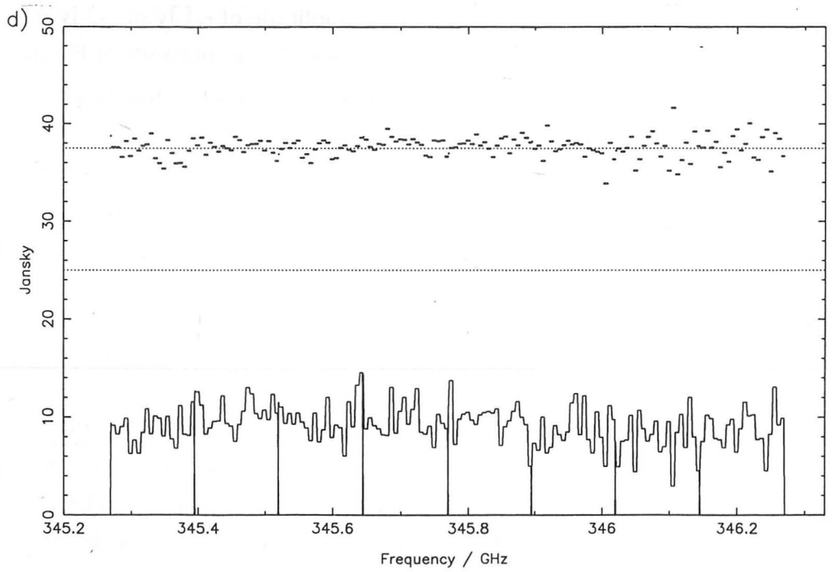


Figure 4.6: ...continued from previous page. (c) Spectrum corrected for the delay error and passband. Amplitudes have been converted to effective antenna temperature. Note the spikes at the edges of some of the subbands. (d) The spectrum has been binned over 5 channels, with the x-axis scaled for sky frequency and the amplitudes converted to Jansky (140 Jy K^{-1}).

4.1.10 3C273: an example of the reduction process

Figure 4.6 shows some of the stages in the reduction process[2] for the bright quasar 3C273, observed in November 1993. In Figure 4.6a, the correlation level for each of the 1024 channels is shown in the lower half and the corresponding phase in the top half, spanning a range of $-180°$ to $+180°$. Because the signal-to-noise ratio is low, the phase points appear to be randomly scattered. The 8 different subbands should also be apparent—this is the 10030 configuration in Figure 2.2.

In Figure 4.6b, both the shape of the amplitude passband and the phase-gradient due to the delay error are clear. In Figure 4.6c the data have been corrected for the delay error and the passband, and the amplitudes have been scaled to units of antenna temperature. Note the spikes at the edges of several of the subbands that were discussed in §4.1.5. In the final spectrum, the data have been binned, the x-axis converted to frequency and the amplitudes converted to Janskys. The phase and amplitude are both flat, as expected for a continuum point source.

Equation 1.41 gives an estimate of the noise level after integrating for duration $t_{\rm int}$ over bandwidth $\Delta\nu$, assuming a temperature-to-jansky factor of 50 Jy K^{-1}. For $T^*_{\rm sys} \simeq 800$ K, $\Delta\nu = 6.25$ MHz, $t_{\rm int} = 1000$ s and allowing for the more realistic value of 140 Jy K^{-1} at the time of observation, the noise has a predicted rms amplitude of $\sim$1 Jy or $\sim$3 Jy peak-to-peak. This compares well with the fluctuations in the first five subbands of Figure 4.6d; the last three are substantially more noisy than this, as receiver B3i at the JCMT has poor sensitivity in this part of the IF band.

The next three examples highlight different techniques for reducing the data and illustrate the wide range of science that is possible with the JCMT–CSO Interferometer.

4.2 The recombination line maser, MWC 349

4.2.1 Background

MWC 349 is a unique star: it is the only object towards which the recombination lines of hydrogen show maser action (see chapter 5 for a short description of masers). The spectra of the masing lines show double-peaked profiles, with velocities of –16 and +32 km s^{-1} with respect to the local standard of rest. The lines must originate in fairly dense, ionized gas that is close to the star, but it is unclear whether they come from the opposite lobes of an expanding, bipolar wind, or are linked to a rotating disc. Observations by Planesas and

[2]The signal-to-noise ratio in a single channel after 10 s of integration is quite low, so, for the sake of illustration, I have fully reduced 1000 s of data and then reversed the reduction process. This has been used for Figures 4.6b, c and d, but does not apply to Figure 4.6a which shows the raw data from one 10 s integration.

Martín-Pintado (1992) using the Owens Valley Millimeter Interferometer at 231.9 GHz seem to support the latter interpretation. They conclude that the emission originates from two maser spots, separated by 65 mas, each associated with a clump of higher density ionized gas located in a disc that is seen nearly edge-on. The nature of the source is disputed; Hamann and Simon (1988) believe it to be a young stellar object, whereas others believe it to be a post–main-sequence star (Harvey *et al.*, 1979). A similar difficulty in classification applies to the star VY Canis Majoris, to which I devote the whole of chapter 5.

4.2.2 Observations and reduction

We observed the H26α transition (353.6 GHz) of MWC 349 on three different nights during November 1993. The lines are very bright (they are easily visible in one 10 s integration) and are therefore ideal for debugging problems with the interferometer. We were interested primarily in the instrumental phase as a function of time, and therefore did not spend time obtaining a good passband calibration or an estimate of the sensitivity at this frequency.

The data presented here are the result of 4400 s of integration taken on 13th November 1993, using the 10050 correlator configuration (Figure 2.2). With a signal-to-noise ratio of about 30 after 10 s of integration, the line is bright enough to be used as a phase reference after every 10 s (§4.1.7).

4.2.3 Final spectra

Two spectra are shown in Figure 4.7. The first is at high resolution showing the two prominent maser peaks and the associated phase. It is likely that the phase variations, for example the slight gradient across the blue-shifted peak, are due to inaccurate passband calibration. I assumed a value of 140 Jy K^{-1} to convert from temperature to Janskys: this value was correct for observations made at 345 GHz, so there may be a small discrepancy at the higher frequency used here. The second spectrum comprises the four lower resolution subbands, spanning a much wider range of velocities. The data have been binned over 5 channels to show the continuum emission at a level of about 3 Jy. With the level of 1.67 Jy deduced by Planesas and Martín-Pintado (1992) at 231.9 GHz, our flux at 354 GHz implies a spectral index of $\sim$1.4 (i.e. $S \propto \nu^{1.4}$), midway between the value of 0.7 expected for an ionized outflow and the value of 2.0 expected for optically-thick dust emission. The lower phase for the most red-shifted subband is again likely to be due to the passband and is probably not inherent to the source.

This example illustrates the importance of accurate calibration of the passband, particularly the phase, when observing bright sources. In this case, there is no good passband measurement, so it is not clear whether the small variations of phase across the spectrum

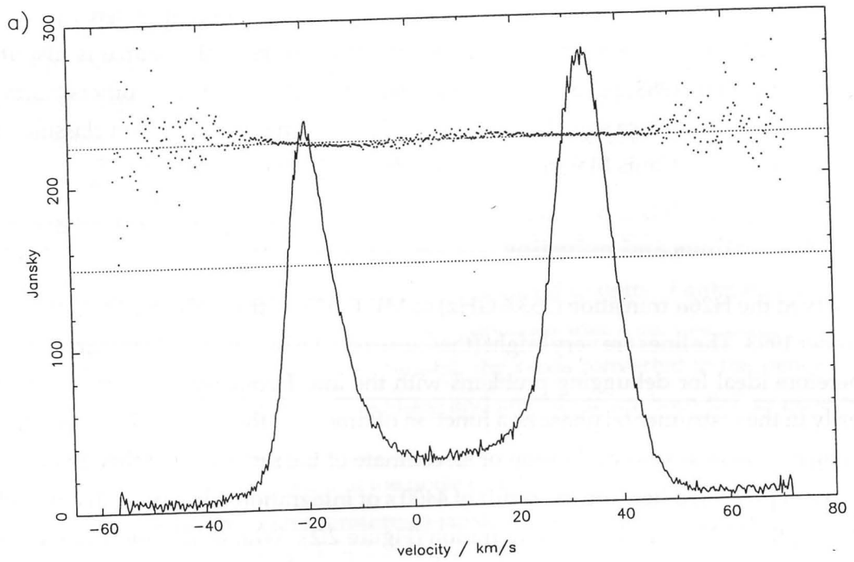


Figure 4.7: Spectra of the hydrogen recombination maser source, MWC 349, at both (a) high resolution and (b) lower resolution, showing the continuum at a level of about 3 Jy.

are real or not.

4.3 The nature of the source at the centre of our galaxy

4.3.1 Background

A black hole of mass $\approx 10^6$ $M_\odot$ is widely believed to exist at the centre of our galaxy, but as yet there is little direct evidence. The central source, known as Sgr A*, is certainly bright and very compact, but its mass is difficult to constrain. It is the origin of a strong continuum flux at millimetre and submillimetre wavelengths, most probably arising from some combination of thermal dust and non-thermal synchrotron emission. By establishing the dominant emission mechanism, we can hope to gain a better understanding of the source as a whole.

The flux of Sgr A* has been measured using single dishes, but the measurement is complicated by extended sources of emission that are present in the beam. Zylka, Mezger and Lesch (1992) (I will refer to this paper as ZML92) used the IRAM 30-m telescope to determine fluxes of $S_{1.3\ \mathrm{mm}} = 2.6 \pm 0.6$ Jy and $S_{0.87\ \mathrm{mm}} = 4.8 \pm 1.2$ Jy. They also obtained $S_{0.35\ \mathrm{mm}} = 18.5 \pm 9$ Jy using the NASA Infrared Telescope Facility (IRTF). The measurements needed a careful subtraction of the contribution from confusing sources.

The JCMT–CSO Interferometer provides a much better way of measuring the flux of the central source, as the extended emission is resolved out and we only detect the flux of compact sources in the beam.

4.3.2 Observations and reduction

We measured the flux of Sgr A* at frequencies of 230 GHz, 270 GHz and 354 GHz during our third run in March 1993. Total integration times were respectively 5200 s, 6300 s and 3600 s. The data were vector-averaged over 100 s and then added incoherently, using the method described in §4.1.7. The prominent absorption lines at 230 GHz were blanked out to prevent an adverse effect on the flux that we determined. The level of the noise was found by reducing the data with a large, artificial phase-gradient to ensure that the signal made only a negligible contribution to the continuum amplitudes after 100 s of integration.

Observations at the same frequencies were also made of the quasar 1730–130, interleaved with those of the Galactic Centre. These were reduced in the same way, so that we obtained effective antenna temperatures for each source at all three frequencies. These are listed in Table 4.1. The quasar fluxes are single-dish measurements, and are used to obtain the fluxes for Sgr A*.

	Freq /GHz	T_A^* /mK	S_ν /Jy
Sgr A*	230	23	2.9
	270	23	3.0
	354	32	3.2
1730–130	230	39	4.9
	270	32	4.2
	354	36	3.6

Table 4.1: Fluxes obtained using incoherent addition for Sgr A* and the quasar 1730–130

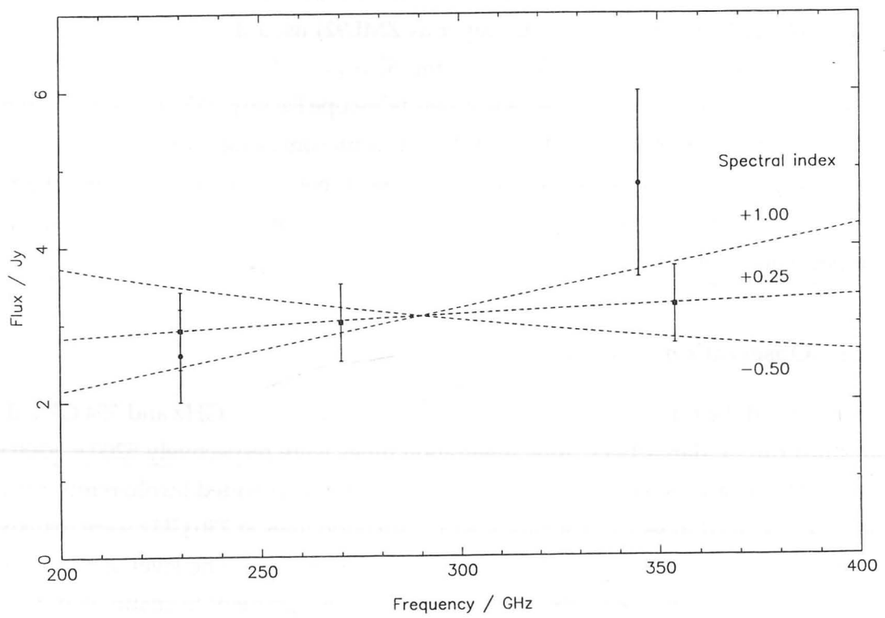


Figure 4.8: A plot of flux vs. frequency for the Galactic Centre source Sgr A*. The squares show the JCMT–CSO Interferometer fluxes at 230, 270 and 354 GHz. Circles show the fluxes obtained by Zylka, Mezger and Lesch (1992). Fits of the spectral index to our data are indicated.

The three fluxes for Sgr A* are plotted with the ZML92 values in Figure 4.8. The error bars correspond to the ±20% uncertainty in the flux calibration.

4.3.3 The spectral index at submillimetre wavelengths

We can use our fluxes to determine the spectral index of the emission, α. The lines shown on Figure 4.8 indicate that $-0.5 < \alpha < +1.0$. This is flatter than the value of 1.5 suggested by the ZML92 results. Emission from a compact ($\ll$1 arcsecond) dust source is ruled out by our fluxes, as we would expect $\alpha \geq 2$.

If the emission is from a source whose size is comparable in extent to our fringe-spacing, however, then we will have resolved out some fraction of the total flux of the central object. The amount of undetected flux is likely to increase with frequency as the fringe-spacing is reduced, leading to an underestimate of α. Therefore we are unable to rule out the possibility that the majority of the flux originates in a $\sim$1 arcsecond region of dust.

If this is the case, we would expect a flux in excess of 15 Jy at 350 μm, and indeed this appears to be borne out by the 18 Jy flux of ZML92 at 350 μm. Dent *et al.* (1993) used the JCMT to map the continuum emission in the Galactic Centre region at 1100, 800 and 450 μm. The fluxes at 1100 μm and 800 μm are both just over 3 Jy, in good agreement with our results. At 450 μm, however, they do not detect Sgr A* at all, and set an upper limit of 1.5 Jy. It is very unlikely that they are in error by a factor of ten, as the rest of the map at this wavelength was consistent with those at the longer wavelengths. We conclude that most of the emission does not come from dust, but instead from a highly variable, compact synchrotron source.

4.4 Observations of the radio-galaxy Centaurus A

4.4.1 Background

At a distance of just 5 Mpc, Centaurus A (NGC 5128) is the closest and most well-studied active galaxy. A good description of the complex radio structure of Cen A is given by Burns, Feigelson and Schreier (1983), from which Figure 4.9a is reproduced. The full extent of the radio emission is staggering: the lobes span almost 10° on the sky. On the smallest scale there is a nucleus and a jet. Figure 4.9b shows the radio emission superimposed on an optical image; the host galaxy is bisected by a dark lane, believed to be a disc, rich in dust and molecular material. Careful modelling of the CO (2–1) emission shows that the disc is probably warped, but with most of the matter still moving in circular orbits (Quillen *et al.*, 1992). There is also evidence to suggest that Cen A is the result of a merger some 10^8 years ago between a small spiral and a larger elliptical galaxy (Baade and Minkowski,

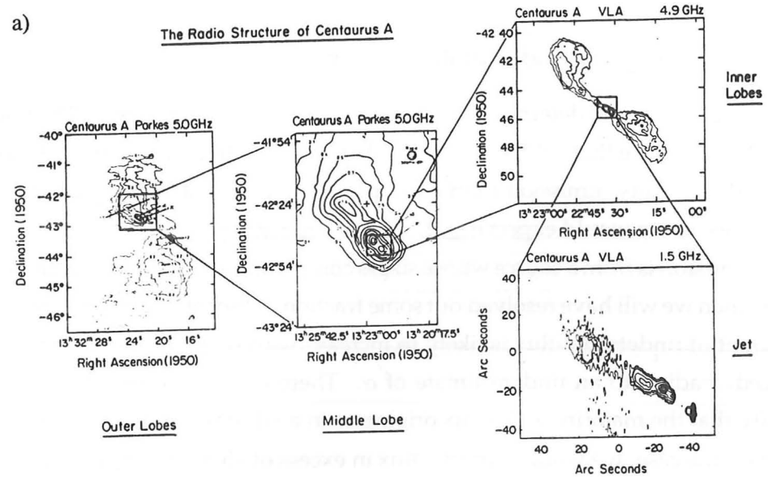


Figure 4.9: Centaurus A. (a) Maps showing the radio structure over a wide range of scales. (b) Radio emission superimposed on an optical image, clearly showing the dark dust lane. Both are reproduced from Burns, Feigelson and Schreier (1983).

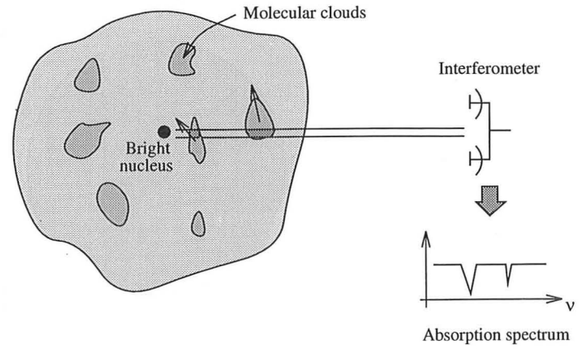


Figure 4.10: A schematic illustration of how absorption lines are caused by molecular clouds along the line of sight to the bright nucleus.

1954; Malin *et al.*, 1983).

4.4.2 Absorption experiments

Radiation from the bright, non-thermal nucleus of Cen A is absorbed by the cool molecular clouds in front of it, and observations have revealed some interesting results. Figure 4.10 illustrates this process.

Absorption lines were first observed in HI by van der Hulst, Golisch and Hashick (1983) using the VLA (see Figure 4.11a). Deep lines at the systemic velocity of +550 km s^{-1} are seen towards both the jet (components 1 and 2) and the nucleus, as expected for material in circular orbits. The surprise was the presence of the red-shifted lines at +576 and +596 km s^{-1}, seen only towards the nucleus. Seaquist and Bell (1990), again using the VLA, found absorption at the systemic velocity for a range of other molecules, allowing them to estimate temperatures and abundance ratios for the absorbing clouds. The red-shifted features were detected only for the OH 1667-GHz line.

Transitions in the millimetre region were probed by Eckart *et al.* (1990) using the SEST. Israël *et al.* (1991) used the single-dish CSO to go still higher in frequency, and Figure 4.11b shows the spectra obtained. The features are clearest for the HCO^+ (1–0) transition. For most of these single-dish observations, the absorption lines are superimposed on a pedestal of emission from the more extended gas in the galaxy, making it difficult to determine the amount of absorption of the nuclear continuum emission. An interferometer resolves out this extended emission and is ideal for these experiments. Cen A has a declination of $-43°$, making it inaccessible to the main millimetre interferometers, which are all situated further north than Hawaii, so our instrument is well-placed to make the observations.

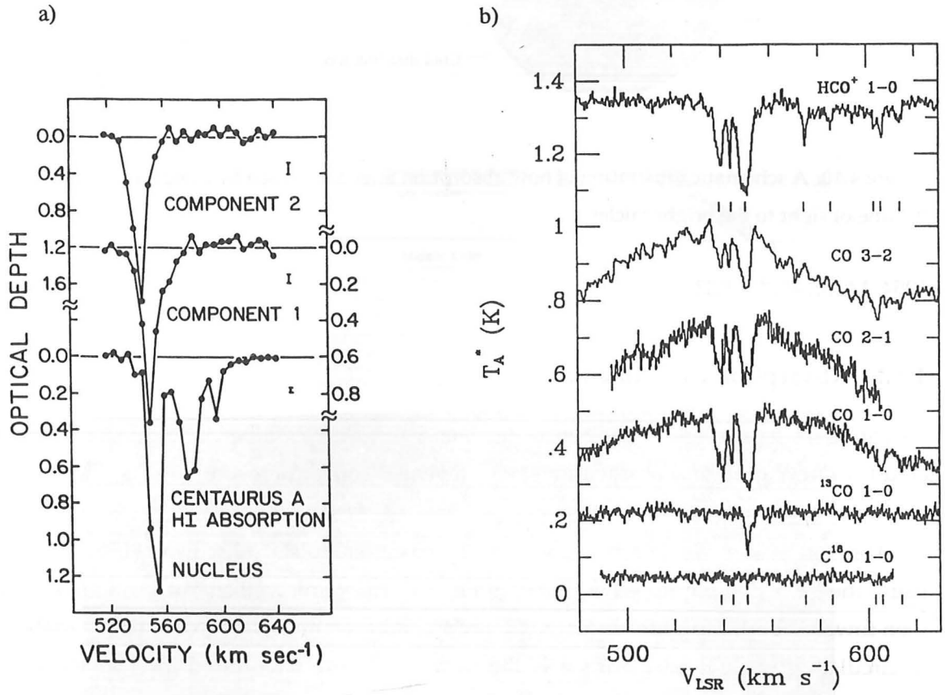


Figure 4.11: (a) Absorption lines in HI, seen against both the nucleus and jet (components 1 and 2) of Centaurus A by van der Hulst, Golisch and Haschick (1983). (b) Absorption lines towards the nucleus of Centaurus A for a number of molecules at millimetre wavelengths, from Israel *et al.* (1991).

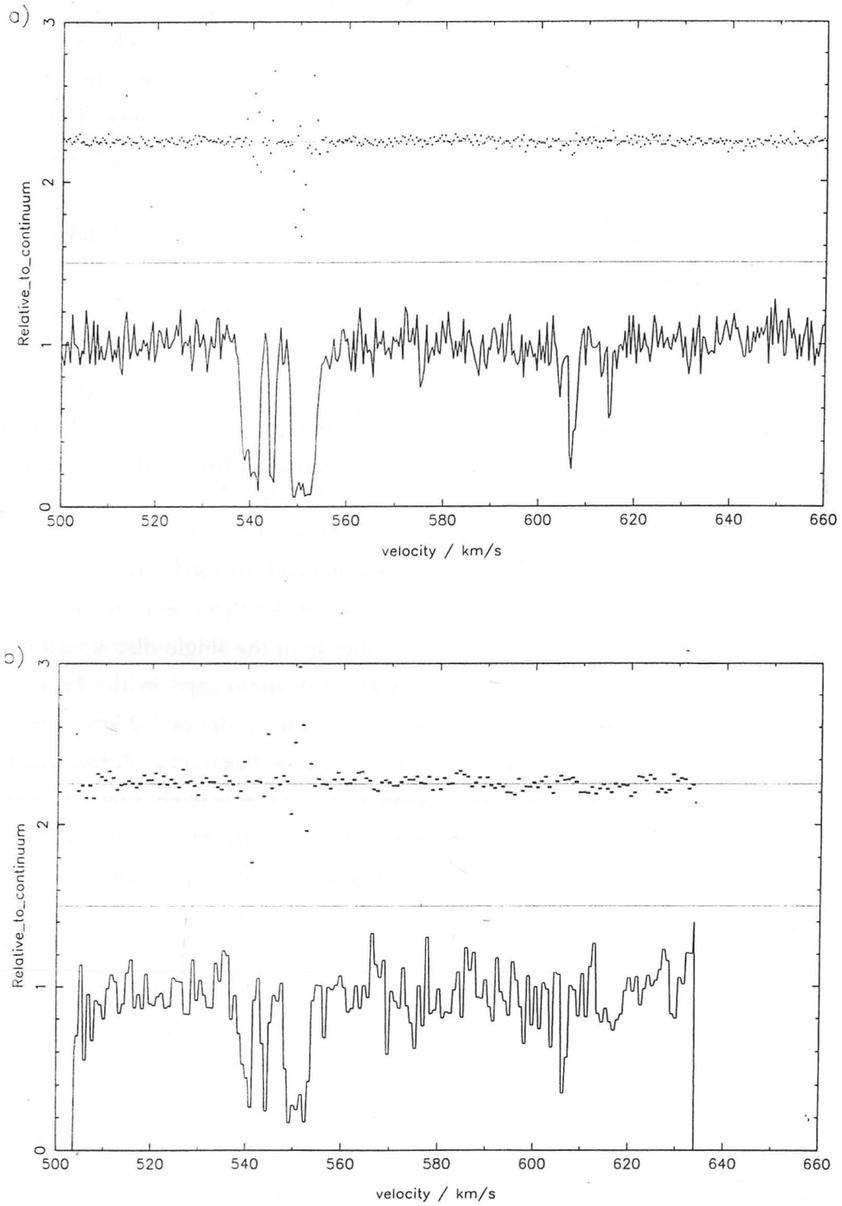


Figure 4.12: Absorption line spectra obtained with the JCMT–CSO Interferometer towards the nucleus of Centaurus A for the transitions (a) CO (2–1) and (b) CO (3–2).

4.4.3 Observations and reduction

We observed Cen A at both the 230-GHz CO (2–1) and 345-GHz CO (3–2) transitions during the third observing run in March 1993. The correlator was used in the hybrid 10050 configuration. In each case the data were vector-averaged over 100 s, after which the nuclear continuum emission over the full 500 MHz was used as a phase reference. The spectra for 230 GHz and 345 GHz, after integration times of 4700 s and 7800 s respectively, are shown in Figure 4.12. Despite the much longer integration time, the CO (2–1) spectrum has a much better signal-to-noise ratio, as the system temperature was much lower. Velocity resolutions are 0.4 km s^{-1} for the CO (2–1) and 0.8 km s^{-1} for the CO (3–2).

In the CO (2–1) spectrum, there are several overlapping, optically-thick lines at around the systemic velocity of +550 km s^{-1}, very similar to those in Figure 4.11b. A number of red-shifted lines are also seen, with velocities of +575, +604, +607 and +614 km s^{-1}. The features at the systemic velocity are also apparent in the CO (3–2) spectrum, as is the red-shifted line at +607 km s^{-1}, but the spectrum is too noisy to be certain about any of the others.

4.4.4 Discussion

Our CO (2–1) spectrum is of much higher quality than the single-dish spectra. The velocities of the absorption lines agree very well with those seen in the HCO^+ data of Figure 4.11b. The lines appear to be very narrow—of the order of 1–2 km s^{-1}—implying that the absorbing clouds have a low velocity dispersion. Eckart *et al.* (1990) quote widths of ~14 km s^{-1} for the red-shifted components in HCO^+; it is likely that this applies to a blend of much narrower lines. The features that we observe are certainly much narrower than those seen in HI by van der Hulst, Golisch and Haschick (1983), and it would appear that the +596 km s^{-1} line in this spectrum does not correspond to the lines we see at velocities in excess of +600 km s^{-1}.

After an analysis of the widths and depths of the single-dish absorption lines, Eckart *et al.* (1990) concluded that the lines result from discrete molecular clouds along the line of sight that block the continuum radiation from the nucleus, rather than self-absorption of the emitting clouds by a nearer, cooler region of gas. Our interferometer is not sensitive to extended regions of absorption, so we have confirmed their result by a more direct method.

How close are these clouds to the nucleus? Are they falling into a central black hole? The Quillen *et al.* (1992) warped disc model with its circular orbits fully accounts for the emission line profiles of the material further than 1 kpc away from the nucleus, so it is unlikely that the red-shifted emission arises this far out. Furthermore, van der Hulst,

Golisch and Haschick (1983) believe that the absorbing HI clouds are within 450 pc of the centre. By combining our data with the HCO^+, OH and HI data, it should be possible to infer temperatures, densities and abundances for these red-shifted clouds, but this has not yet been attempted. Tidal disruption will increase the velocity dispersion of clouds close to the nucleus, and we expect the very narrow line widths observed to put strong constraints on the proximity of the clouds to the centre. At this stage, it seems most likely that the red-shifted clouds are part of an inner bar in Centaurus A, where the non-axisymmetric potential results in highly non-circular orbits. Shlosman, Frank and Begelman (1989) advocate such bars as being necessary to transport material from large distances out in the disc into the heart of active galactic nuclei. There is certainly more work to be done here.

4.5 Summary

The main stages in the reduction of data from the JCMT–CSO Interferometer were discussed in the first half of this chapter. Inevitably, the software will evolve with time and improvements will be made. In particular, I would like to compress the amount of space taken up by the data. The raw data files are large and unwieldy, but have proved useful for finding and correcting the problems that were discovered after the first attempt at reduction. As the operation of the interferometer becomes more routine, I intend to change the software so that the data are saved in a more compact, semi-processed format, probably as 100 s vector-averaged integrations.

Three examples of observations for which the interferometer is ideally suited made up the second half of the chapter, with brief discussions of the scientific background and interpretation of the data. Two further illustrations of the useful science that is possible with a single-baseline submillimetre interferometer are developed in the next two chapters, where I present in-depth observations and analysis for the evolved star VY Canis Majoris and the protostars L1551-IRS 5 and HL Tau.

Chapter 5

VY CANIS MAJORIS

5.1 Introduction

In this chapter I describe how observations of the evolved star VY CMa with the JCMT–CSO Interferometer have shed new light on the nature of this strange source. I also summarize the wide range of existing observational data, including the suite of bright maser lines for which VY CMa is well-known. I begin with a simple introduction to evolved stars [see Tayler (1972) for a more detailed account] and their associated maser emission, followed by a description of the features of VY CMa that mark it out as such an odd source.

5.1.1 Evolved stars

A star on the main sequence of the Hertzsprung–Russell (HR) diagram burns hydrogen in its core to form helium. When this fuel is exhausted the core starts to collapse, slowly at first but then much more rapidly as the Schönberg–Chrandrasekhar limit of pressure and density is exceeded. If the stellar mass exceeds $\sim$3 $M_{\odot}$—as is the case for the high-mass stars that I am considering here—then the collapse is arrested by the onset of helium burning to form carbon. Meanwhile, hydrogen burning proceeds in a shell around the core and the outer parts of the star (the convective envelope) expand and cool. Once the core helium is used up, the reaction to form carbon continues in a shell, in the same way as for the hydrogen to helium process. The core again contracts and heats up until the temperature is sufficient to allow carbon to burn. This process continues, building up a layered structure with hydrogen on the outside and heavier elements nearer the centre.

The radius of the star is by now very large and the surface relatively cool. The shell-

burning phases trigger periods of convective circulation in the envelope, during which the star moves almost vertically up a Hayashi track in the HR diagram. Stars at this stage of evolution, where they return repeatedly to the Hayashi track, are said to be on the Asymptotic Giant Branch (AGB) of the HR diagram. With a surface temperature of 2000–3000 K and a radius of 10^{11}–10^{13} m, such stars are of type M, with luminosities as high as 10^4 to 10^5 $L_\odot$. Variability in this luminosity is the norm: it is believed to result from radial pulsations of the convective envelope, the cause of which is still under debate [see for example Bowen (1988) and Wood (1979)].

Mass loss is a characteristic of these evolved stars; the convective envelope material is lost as a wind of gas and dust. When the temperature is low enough for dust to form, it is accelerated away from the star by radiation pressure. Collisions between the dust particles and gas transfer momentum to the gas, forming a wind and a circumstellar envelope (Chapman and Cohen, 1986). Eventually, the entire convective envelope is blown off to form a detached, expanding shell, leaving the underlying degenerate core. The object is now classified as a planetary nebula (PN).

Most of the mass loss occurs at a stage intermediate between the AGB and PN phases; post-AGB or proto-planetary nebula (PPN) are commonly used as labels to describe such objects. An interesting feature is that the mass loss in many stars appears to be aspherical [see for example Bowers (1991) and the review of Podsiadlowski and Clegg (1992)]. A number of theories has evolved to explain this, based variously on rotation (Calvert and Peimbert, 1983), primordial discs (Matese *et al.*, 1989), non-spherical radial pulsations (Icke and Heske, 1992), magnetic fields (Pascoli, 1992) and binary systems (Morris, 1987). In each case the equatorial gas is denser than that towards the poles, and the general distribution of material around the star is in the form of a prolate ellipsoid.

VY CMa is commonly classified as a post-AGB object and as an OH/IR supergiant. The second term refers to the combination of strong infrared emission from the warm dust that surrounds the star, and the bright OH maser emission that is seen. Indeed, much of the information on the environment of evolved stars is deduced from observations of their associated circumstellar masers.

5.1.2 Masers

MASERs (Microwave Amplification by the Stimulated Emission of Radiation) were first discovered in space in 1965 (Weaver *et al.*, 1965). If a molecule has energy levels A and B ($E_B > E_A$), then a photon with frequency ν_{AB}, such that $h\nu_{AB} = E_B - E_A$, can interact with a molecule in the excited state B and stimulate the emission of a second, coherent photon. This travels in the same direction, and with the same frequency and phase as the first. Maser amplification requires that there is a population inversion (more molecules in state B than in state A). If a photon enters the end of a column of such molecules it will stimulate the emission of an identical photon; these two can then generate two more photons, and so on. The number of photons travelling along the column grows exponentially (the emission is said to be unsaturated) until there are no longer enough molecules in the excited state to maintain the growth. The maser is then saturated and the gain becomes proportional to the path length.

Seen on the sky by interferometers, masers appear as small ($\sim$1 mas) spots with very high, non-thermal, brightness temperatures ($\sim 10^{10}$ K). The emission has a frequency corresponding to the doppler velocity of the material in the column of gas behind the spot, so masers are therefore useful probes of the circumstellar velocity field. There has been much work towards understanding maser action in a number of different molecules; in general the behaviour is complex and requires detailed modelling to include the nature of the pumping mechanism that generates the population inversion, together with geometry and kinematics. Most maser lines are sensitive to their local physical conditions—the gas temperature and the external radiation field, for example—and can tightly constrain the environment in which they occur. The envelopes of oxygen-rich evolved stars are a particularly rich source of maser activity, particularly for the SiO, H_2O and OH molecules. Table 5.1 summarizes the different characteristics of some of the important transitions. Further details on masers can be found in Elitzur (1992).

5.1.3 The trouble with VY CMa...

Interest in VY CMa was first aroused as long ago as 1900, when it was believed to be a multiple star. Some of the components originally observed were not found by later observers: it was increasingly suspected that VY CMa was not a multiple source at all, and that the apparent companions were merely condensations in the $8'' \times 12''$ nebula that

Table 5.1: Masers from evolved stars

Molecule	Frequency	T	Comments
H_2O	22 GHz	640 K$^{(a)}$	Mapped with VLA[1], MERLIN[2], VLBI[3]. Found in inner circumstellar envelope.
	321 GHz	~1860 K$^{(b)}$	Only single-dish observations until now.
	325 GHz	~470 K$^{(b)}$	Difficult to observe because of absorption by our atmosphere.
SiO	43 GHz	>1200 K$^{(c)}$	Found in stellar atmosphere, where the temperature is too hot for dust to form.
	many more		
OH	1612 MHz	~50 K$^{(d)}$	Produced in the wind relatively distant from the star where H_2O is photo-dissociated. Pumped by IR photons.
	1665/7 MHz	150–280 K$^{(e)}$	Less well-studied than 1612 MHz.

[1] Very Large Array, New Mexico.

[2] Multi-Element Radio-Linked Interferometer, England.

[3] Very Long Baseline Interferometry.

(a) Temperature above ground-state (p.226, Elitzur 1992).

(b) Temperature above ground-state (Yates 1993).

(c) No dust above this temperature (p.291, Elitzur 1992).

(d) Typical gas temperature (p.241, Elitzur 1992).

(e) Typical dust temperature required (p.247, Elitzur 1992).

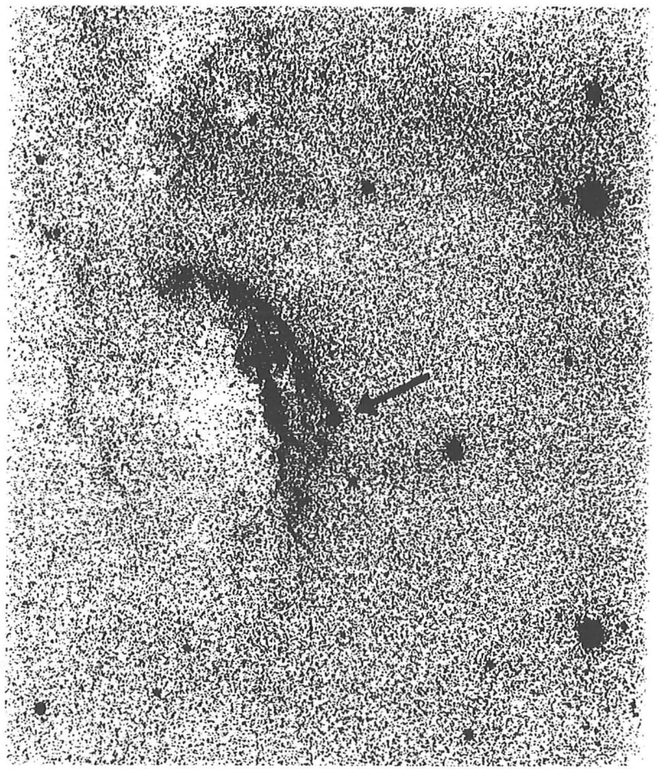


Figure 5.1: Optical plate showing VY CMa at the apex of an arrow-shaped bright rim. The rim spans ~1° from top to bottom.

surrounds the star. In 1951, van den Bos commented that "the entire image is so nebulous that it is difficult to measure anything at all".

Herbig has devoted a lot of time to this object; for him, VY CMa is "one of the most fascinating and investigated objects." By studying the spectral energy distribution, he concluded that VY CMa is a protostar surrounded by large quantities of dust in the form of an equatorial disc (Herbig, 1970). A further paper (Herbig, 1972) examined the polarization of the nebula and established that the bright optical sources seen in it were simply concentrations of dust that were scattering the light from the central star.

The general environment of VY CMa constrains the nature of the star. VY CMa is located towards the edge of Sharpless 310, a large HII region some 5° in diameter. This region is probably excited by two bright O stars, τ CMa and UW CMa, located near the

centre, and by members of the young cluster NGC 2362. Closer to the edge of the HII region is a very prominent, arrow-shaped, bright rim, where the edge of a molecular cloud is being excited (Figure 5.1). At the tip of this arrow we find VY CMa. The obvious conclusion is that VY CMa is a source of ionizing radiation that excites the cloud; it would be quite a coincidence if such a peculiar source was not directly associated with the rim and lay fortuitously along the line of sight. Lada and Reid (1978) argued that the rim was the edge of a molecular cloud complex whose velocity was similar to that of the NGC 2362 cluster. Therefore it is probable that VY CMa was located at the same distance as these objects, namely 1500 pc. However: at this distance VY CMa has the rather spectacular luminosity of 5×10^5 $L_\odot$ and consequently must be very massive. Lada and Reid suggest a mass in excess of 15–20 $M_\odot$; a mass of 50 $M_\odot$ is inferred by Knapp, Sandell and Robson (1993) from Cox's (1980) relationship between luminosity and pulsation period. If VY CMa is now an evolved star, then previously, whilst on the main sequence, it would have been an O star, and would have ionized and blown away the surrounding molecular material, just as τ CMa and UW CMa have done out to distances of 40 pc or more. Lada and Reid (1978) reasoned that VY CMa must be a massive protostar and that the rim must be excited by the two O stars. They could not explain why VY CMa should be projected onto the tip of the arrow.

Furthermore, Eliasson and Bartlett (1969) discovered that VY CMa is a bright source of OH maser emission at 1612 MHz, and in the same year Knowles (1969) detected H_2O emission. The double-peaked OH spectrum has only been seen towards evolved stars, and the SiO maser emission also indicates strongly that VY CMa is much more likely to be an evolved star than a protostar. The generally accepted view at the present time is that VY CMa is a very massive, evolved, M-type OH/IR supergiant. There has been no attempt to resolve the issue of the nearby molecular clouds or the source of the radiation ionizing the rim. The new data presented here, and its interpretation, should help to solve the problem.

In the next section I will describe the data that we obtained for VY CMa using the JCMT–CSO Interferometer, comprising observations of both the 321-GHz H_2O line and the CO (3–2) transition at 345 GHz.

5.2 The JCMT–CSO Interferometer data

5.2.1 The 321-GHz H_2O maser data

Observations and reduction

The observations described here were made with the interferometer on March 27th, 1993. The very strong H_2O maser line at 321 GHz is one of the brightest objects in the sub-millimetre sky and was ideal for helping to calibrate the instrument. We observed the source for 8800 s in the lower sideband from hour angles –1.1 to +1.9, using the 10050 correlator configuration (Figure 2.2c). The main line (+14 to +24 km s^{-1}) was used as a phase-reference feature to correct every 100 s for drifts in the phase. Because the line is so strong (2.2 K), any decorrelation is solely due to atmospheric phase fluctuations. The antenna temperatures were scaled to units of flux using a factor of 115 Jy K^{-1}, which includes an extra 10% to account for the noise from the atmospheric H_2O line in the upper sideband (§3.3.1).

The spectra

Figure 5.2a shows the 321-GHz H_2O maser line with a resolution of 321.5 kHz, or 0.3 km s^{-1}. It is apparent that the line comprises at least five main components and spans +5 to +35 km s^{-1} in velocity. In the range +14 to +22 km s^{-1} the phase is constant to within $\pm 7°$, corresponding to an angular size on the sky of 50 mas. Small variations are visible, suggesting that the different maser components are spatially distinct. The noise contributes at a level of $\pm 1°$ for a maser component of 100 Jy, so that the main uncertainty comes from the passband calibration (§3.4). Phase errors in the passband calibration will result in an overestimate of the angular spread of the masers on the sky. There was, however, no such phase structure in the passband spectrum that was used for the calibration.

Spectra generated separately from the first and second halves of the data are almost identical, and this can be used to constrain the spatial extent of the maser emission in the direction parallel to the fringes. In the time between the first half and the second half of the data, the fringes have rotated through an angle of about 11°, but the phase variations across the line are the same to within ~6°, or 20 mas on the sky. Therefore the spread of the different maser spots cannot be extended parallel to the fringes by more than 20 mas/ tan 11° ≈ 100 mas. We have pinned down the centroids of the sources of emission

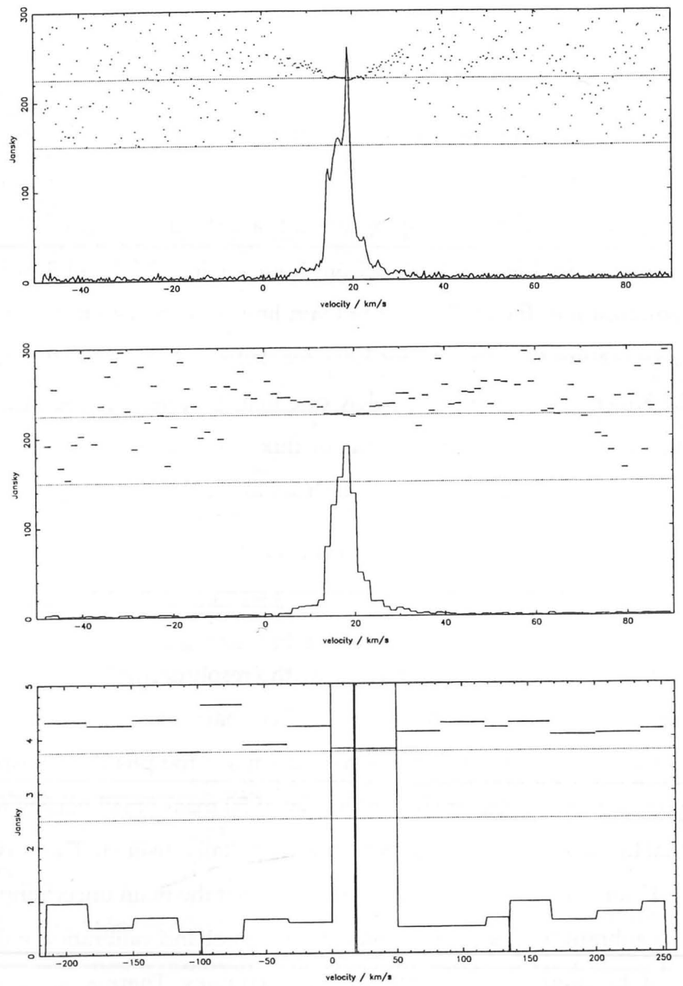


Figure 5.2: Complex spectra of the 321-GHz H_2O line in VY CMa taken with the JCMT–CSO Interferometer in March 1993. Integration time is 8800 s. The phase increases to the east. (a) High resolution showing the main line. Resolution is 0.3 km s^{-1}; noise is 2 Jy, rms. (b) Smoothed to show up the phase of the weaker, broader component. Bins are 1.5 km s^{-1} (1.6 MHz) wide; noise is 0.7 Jy, rms. (c) The full 500-MHz bandwidth, with smoothing over to show the continuum which is offset in phase from the main line. Bins are 34 km s^{-1} (36 MHz) wide; noise is 0.3 Jy, rms.

for the different bright-line components to an area not larger than 50 × 100 mas. If we assume that this also sets an upper limit to the size of an individual spot, then the inferred brightness temperature for a 100 Jy line is 2.6×10^5 K. This is clearly non-thermal. We have therefore demonstrated for the first time that maser activity must be responsible for the 321-GHz H_2O line, something that has always been assumed.

Smoothing Figure 5.2a over 1.6 MHz gives the spectrum shown in Figure 5.2b, and reveals that the higher velocity maser emission (+5 to +14 km s^{-1} and +22 to +35 km s^{-1}) is displaced to the east, by an amount that increases with the velocity difference from nominal stellar velocity of 19 km s^{-1}.

There seem to be two separate contributions to the 321-GHz H_2O maser emission: (1) the strong lines in the range +14 to +22 km s^{-1} coming from a region no larger than 50×100 mas and (2) a weaker, underlying component with velocities ranging from +5 to +35 km s^{-1} with an offset to the east dependent on the doppler velocity.

Applying smoothing over 9 MHz and looking at the entire range of available velocities gives the spectrum of Figure 5.2c. This shows continuum emission present at a level of ∼0.8 Jy, offset to the east by ∼200 mas from the main H_2O line.

5.2.2 The CO (3–2) data

Observations

After observing the strong 321-GHz H_2O maser emission in March 1993, we retuned the receiver to the CO (3–2) line and were surprised to find a strong detection, indicating bright and compact emission. In November 1993 we again observed VY CMa at this frequency. I describe these two sets of observations separately.

The March 1993 data

We observed for 2300 s on March 28th, 1993 with the source between hour angles 2.2 and 3.2. The signal-to-noise ratio over the main part of the line (0 to +50 km s^{-1}) was ∼3 after 100 s of integration. This interval was used as a phase reference. The correlator was in the hybrid 10050 configuration. The amplitudes were scaled to flux by a factor of 96 Jy K^{-1}.

Figure 5.3a shows the CO line reaching a peak effective antenna temperature of about 80 mK, with the line spanning the velocity range 0 to +70 km s^{-1}. There appears to be

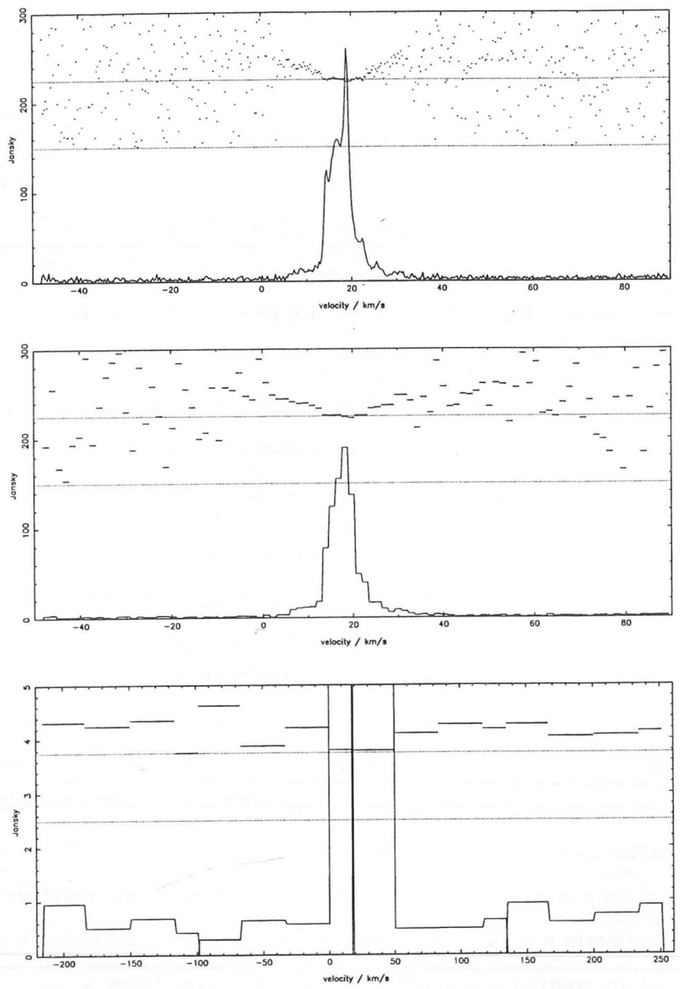


Figure 5.2: Complex spectra of the 321-GHz H_2O line in VY CMa taken with the JCMT–CSO Interferometer in March 1993. Integration time is 8800 s. The phase increases to the east. (a) High resolution showing the main line. Resolution is 0.3 km s^{-1}; noise is 2 Jy, rms. (b) Smoothed to show up the phase of the weaker, broader component. Bins are 1.5 km s^{-1} (1.6 MHz) wide; noise is 0.7 Jy, rms. (c) The full 500-MHz bandwidth, with smoothing over to show the continuum which is offset in phase from the main line. Bins are 34 km s^{-1} (36 MHz) wide; noise is 0.3 Jy, rms.

for the different bright-line components to an area not larger than 50 × 100 mas. If we assume that this also sets an upper limit to the size of an individual spot, then the inferred brightness temperature for a 100 Jy line is 2.6×10^5 K. This is clearly non-thermal. We have therefore demonstrated for the first time that maser activity must be responsible for the 321-GHz H_2O line, something that has always been assumed.

Smoothing Figure 5.2a over 1.6 MHz gives the spectrum shown in Figure 5.2b, and reveals that the higher velocity maser emission (+5 to +14 km s^{-1} and +22 to +35 km s^{-1}) is displaced to the east, by an amount that increases with the velocity difference from nominal stellar velocity of 19 km s^{-1}.

There seem to be two separate contributions to the 321-GHz H_2O maser emission: (1) the strong lines in the range +14 to +22 km s^{-1} coming from a region no larger than 50×100 mas and (2) a weaker, underlying component with velocities ranging from +5 to +35 km s^{-1} with an offset to the east dependent on the doppler velocity.

Applying smoothing over 9 MHz and looking at the entire range of available velocities gives the spectrum of Figure 5.2c. This shows continuum emission present at a level of ∼0.8 Jy, offset to the east by ∼200 mas from the main H_2O line.

5.2.2 The CO (3–2) data

Observations

After observing the strong 321-GHz H_2O maser emission in March 1993, we retuned the receiver to the CO (3–2) line and were surprised to find a strong detection, indicating bright and compact emission. In November 1993 we again observed VY CMa at this frequency. I describe these two sets of observations separately.

The March 1993 data

We observed for 2300 s on March 28th, 1993 with the source between hour angles 2.2 and 3.2. The signal-to-noise ratio over the main part of the line (0 to +50 km s^{-1}) was ∼3 after 100 s of integration. This interval was used as a phase reference. The correlator was in the hybrid 10050 configuration. The amplitudes were scaled to flux by a factor of 96 Jy K^{-1}.

Figure 5.3a shows the CO line reaching a peak effective antenna temperature of about 80 mK, with the line spanning the velocity range 0 to +70 km s^{-1}. There appears to be

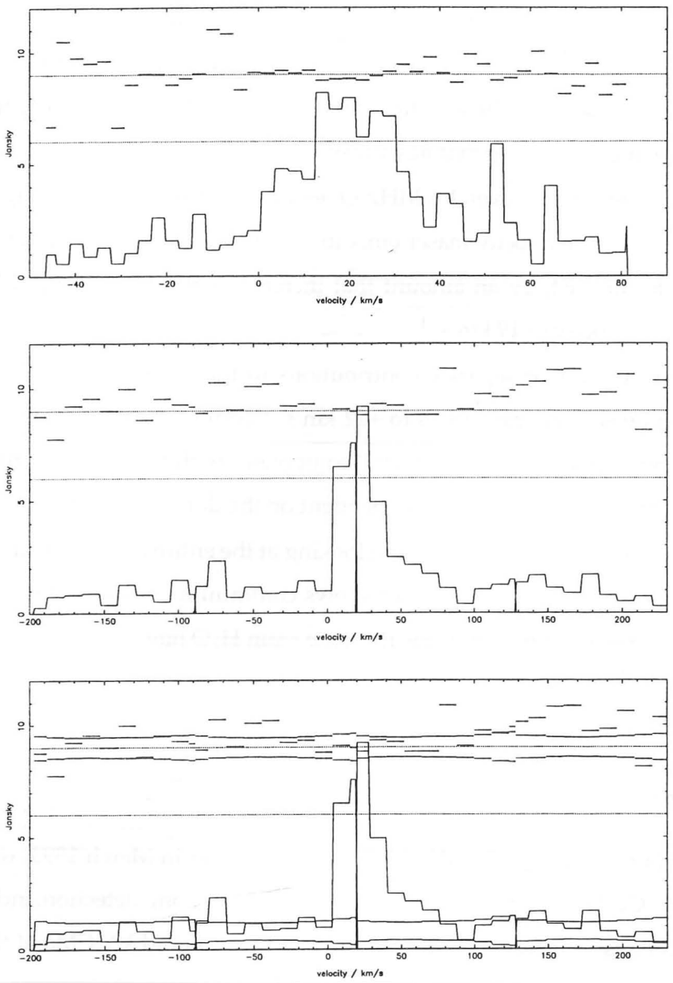


Figure 5.3: Complex spectra of the CO (3–2) line for VY CMa taken with the JCMT–CSO Interferometer in March 1993. Integration time is 2300 s. The phase increases to the east. (a) The main line. There appears to be a red wing composed of discrete emission peaks at +40, +53 and +65 km s^{-1}. Bins are 3 km s^{-1} (3.4 MHz) wide; noise is 0.9 Jy, rms. (b) The full 500-MHz bandwidth. In addition to the main line and continuum emission, there appears to be excess emission at around +130 km s^{-1} and –80 km s^{-1} with corresponding phase excursions. Bins are 12 km s^{-1} (14 MHz) wide; noise is 0.4 Jy, rms. (c) As for previous plot, but with lines to show the expected $\pm 1\sigma$ amplitude and phase variations from noise superimposed on a 0.8 Jy continuum source.

more emission on the red-shifted side of the line than the blue, with at least two distinct peaks at +52 and +65 km s^{-1}.

Figure 5.3b shows the full bandwidth that was observed. The continuum emission is now apparent at a level of about 1 Jy, consistent (to within the errors) with the 0.8 Jy continuum detected at 321 GHz. The phase shows a quite remarkable deviation from zero in the +70 to +170 km s^{-1} region, and to a lesser extent in the blue-shifted velocity range –30 to –90 km s^{-1}. There also appears to be an increase in the amplitude in these ranges.

Could these deviations be noise? Figure 5.3c indicates the $\pm 1\sigma$ range for the data in both amplitude and phase, assuming continuum emission at a level of 0.8 Jy, and with zero phase. The phase deviations are highly significant: seven points exceed the 2σ-level on the red-shifted side; three on the blue-shifted side. The statistical expectation is for only two points out of forty to lie outside $\pm\, 2\sigma$. If the continuum emission is actually less than 0.8 Jy, then although the phase excursions become less important, the excess amplitudes become more significant, and vice versa. The large deviations in both amplitude and phase cannot be made consistent with a chosen continuum level and phase. A spectrum taken on the quasar 3C273 at the same frequency just before the VY CMa observations was flat in both amplitude and phase, and there are no other molecular lines nearby in frequency that might be responsible.

The data therefore suggest the existence of high velocity molecular material. The phase offsets are both to the east; the red-shifted material is offset by up to 300 mas and the blue by 200 mas.

November 1993 data

On November 13th, 1993 we observed VY CMa again at the CO (3–2) transition in an attempt to confirm the March observations. This time the data were taken when the source was near zero hour angle. After 1400 s of integration using the 10050 correlator configuration, we switched to observing over the full 1-GHz bandwidth (the 10030 configuration). I again phase-referenced the data to the main line after every 100 s of integration. The antenna temperatures were scaled to flux by a factor of 140 Jy K^{-1}.

Figure 5.4a shows the main line as observed in the high-resolution portion of the hybrid correlator configuration. The peak brightness is the same as it was in March and again there are red-shifted emission peaks. The most prominent is at +42 km s^{-1} which has the

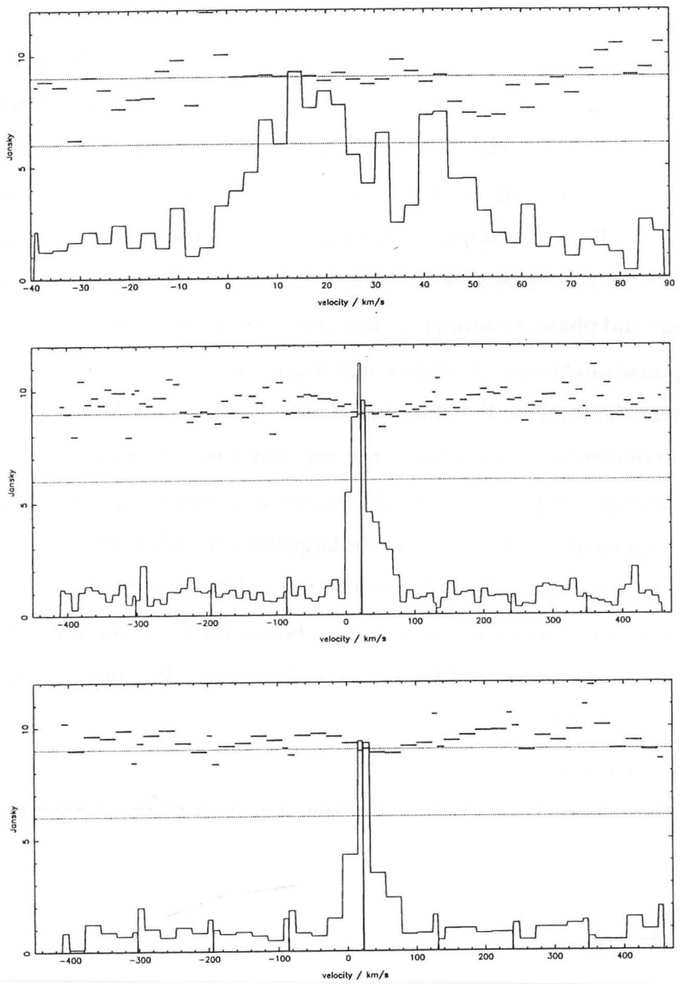


Figure 5.4: Complex spectra of the CO (3–2) line for VY CMa taken with the JCMT–CSO Interferometer in November 1993. Phase increases to the east. (a) The main line after 1400 s of integration. A red-shifted emission feature at +42 km s^{-1} is prominent. There is also a large phase excursion to the west between +45 and +55 km s^{-1}. Bins are 3 km s^{-1} (3.4 MHz) wide; noise is 1.6 Jy, rms. (b) 5000 s of integration over 1 GHz of bandwidth, showing the main line with the red wing. Bins are 10 km s^{-1} (11 MHz) wide; noise is 0.9 Jy, rms. (c) As for previous plot, but with more smoothing to emphasize the continuum at about 1 Jy and the phase offset between the line and continuum. The continuum is offset to the east of the line by about 80 mas. Bins are 23 km s^{-1} (26 MHz) wide; noise is 0.6 Jy, rms.

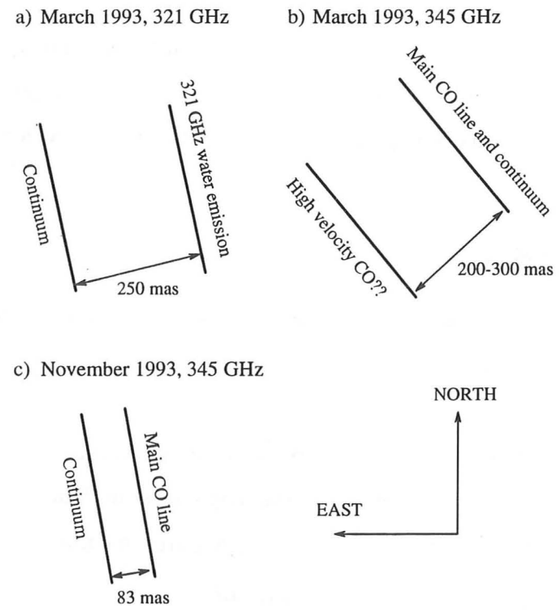


Figure 5.5: The orientations on the sky of the JCMT–CSO Interferometer fringes for the March and November 1993 data. Relevant offsets are shown.

same phase as the main part of the line. There is a significant phase offset to the west in the velocity range +45 to +55 km s^{-1}. The interferometer detects less than 15% of the peak single-dish flux for the CO (3–2) line (see §5.3.5). It is therefore likely that although the general profile of the line may remain the same, detailed features such as the red-shifted emission peaks will vary, depending on the exact spacing and orientation of the fringes.

Figure 5.4b shows the spectrum taken over the full 1-GHz bandwidth with a moderate amount of smoothing; more has been applied to generate Figure 5.4c. The amplitudes of the line and continuum are consistent with the March observations. The line again shows a distinct red wing of emission and there is also a small phase offset from the continuum; the continuum is approximately 80 mas eastward (in the direction perpendicular to the fringes) of the line emission. In these observations, however, there is no evidence for the supposed high velocity material seen in March.

Figure 5.5 shows the orientation of the fringes and the relevant offsets for the interferometer data from March and November 1993.

5.3 Data from other sources

There is a wealth of data available for VY CMa generated by many different instruments and observing techniques. The following sections describe the data for other maser lines, measurements of the dust flux and polarization seen at optical wavelengths, and the single-dish CO (3–2) spectrum.

5.3.1 The 22-GHz H_2O maser data

Both the Very Large Array (VLA) and the MERLIN array have been used to observe the 22-GHz H_2O maser line at high spatial resolution [VLA: Bowers, Claussen and Johnston (1993), MERLIN: Yates (1993)]. Figures 5.6a and 5.6b show the spot maps obtained with each instrument for comparison. Each spot represents an individual masing source with a well-defined location and velocity. Although the maps differ in detail and the MERLIN map spans a larger range of velocities, they have key features in common:

- The emission is clearly asymmetric.
- There are regions of extreme red- and blue-shifted emission very close to one another, but with the blue displaced by about 50 mas to the north-west.
- There is red-shifted material about 200 mas to the north-east of the map centres.

The map of the MERLIN integrated emission is instructive (Figure 5.7). The emission is extended with a position angle of about 150° and the material to the east shows up well. Peak emission coincides with the centre of the adjacent red- and blue-shifted emission.

5.3.2 OH Masers

The 1667-MHz line

Morris and Bowers (1980) find OH Maser emission at 1667 MHz that has similarities with the sources OH231.8+4.2 and M1-92, both of which are bipolar nebulae. Spikes of emission at +4 km s^{-1} and +40 km s^{-1} are separated by a low-level plateau of maser emission. The authors suggest that this is incompatible with a spherically-symmetric source and invoke an axially-symmetric model. They are reluctant to classify VY CMa as a bipolar nebula because of its assumed protostellar nature. They also propose a pumping mechanism that requires dense, cool (< 100 K) dust.

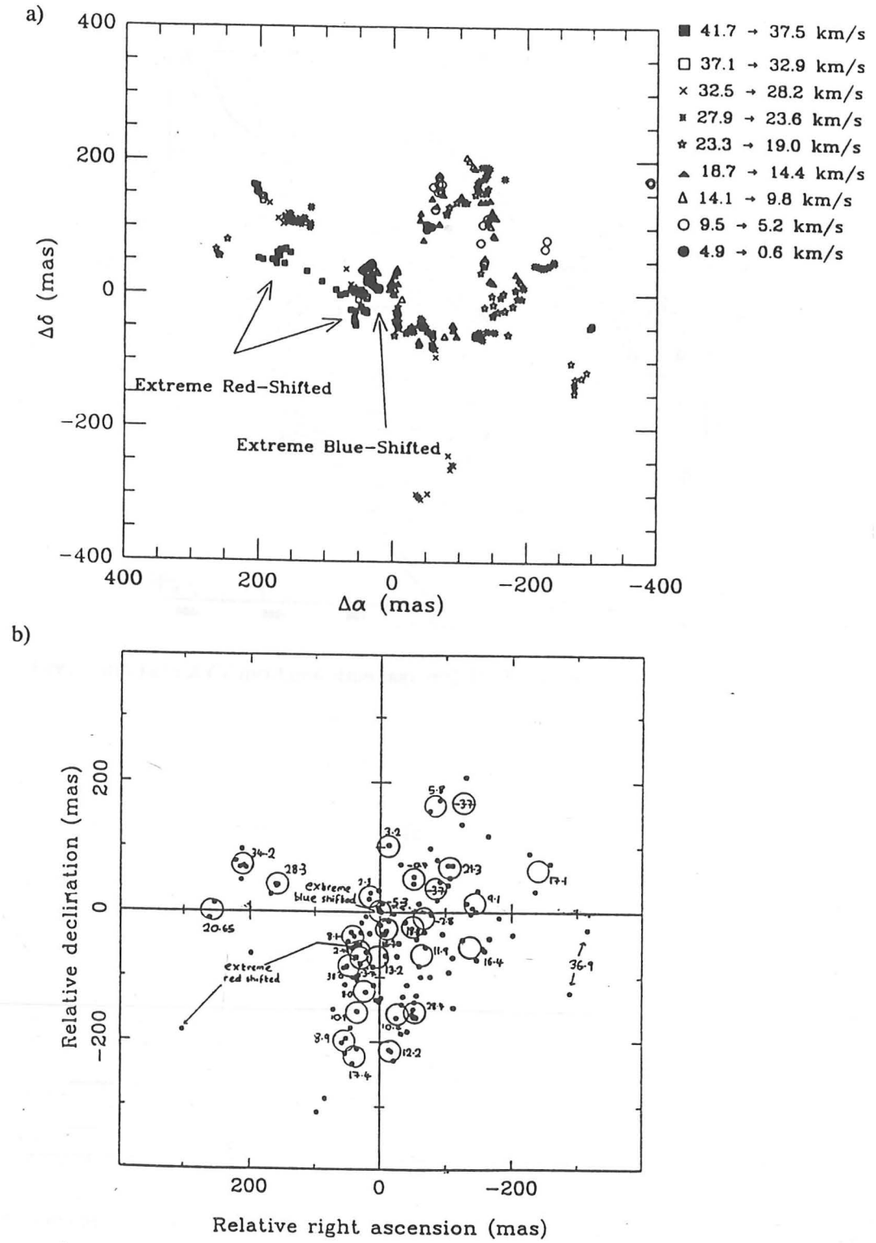


Figure 5.6: Spot maps of the 22-GHz H_2O maser emission from VY CMa. Each maser spot is labelled with its velocity. (a) Map made with the VLA by Bowers, Claussen and Johnston (1993). (b) Map made with MERLIN by Yates (1993).

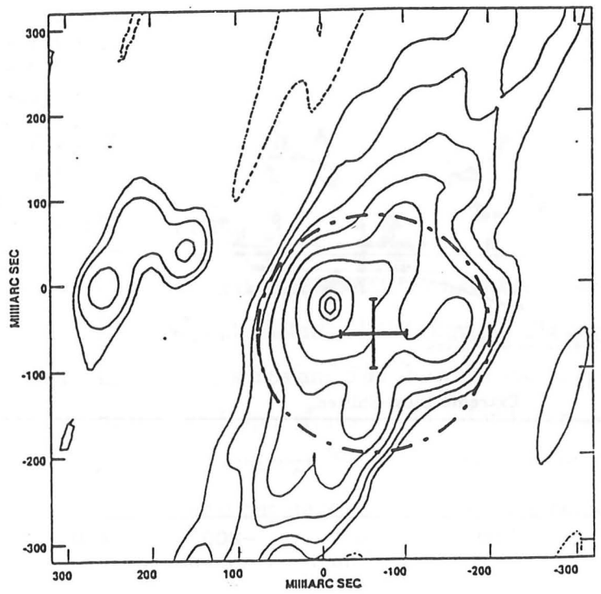


Figure 5.7: MERLIN integrated 22-GHz H_2O maser emission from VY CMa (Yates, 1993).

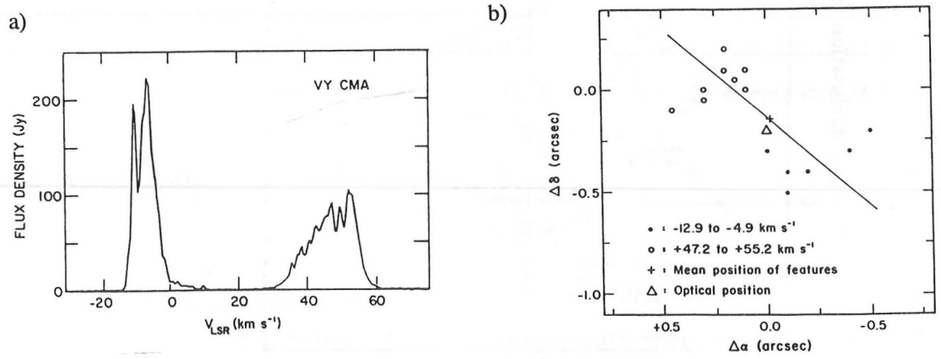


Figure 5.8: 1612-MHz OH maser emission from VY CMa (Bowers, Johnston and Spencer 1983). (a) Spectrum taken with the old Green Bank dish. (b) Map to show how the extreme red- and blue-shifted features are spatially offset from one another.

The 1612-MHz line

Figure 5.8a shows the spectrum of the 1612-MHz OH maser line from Bowers, Johnston and Spencer (1983). The stellar velocity of about +20 km s^{-1} is midway between the extremes of the OH emission. They also found that the extreme red- and blue-shifted velocities were displaced from one another by about 0.6 arcseconds, as depicted in Figure 5.8b. These results were interpreted as emission from an expanding disc, with its axis along the line indicated in the diagram. The radius in the north-south direction was about 2 arcseconds, suggesting that the disc is highly inclined to the plane of the sky. They also find that at intermediate velocities (–5 to +47 km s^{-1}) emission is mostly to the east of the star.

5.3.3 SiO Masers

Cernicharo, Bujarrabal and Santarén (1993) detected 14 different maser transitions of SiO, including the $\nu = 4$, $J = 5$–4 line which has an excitation temperature of about 7000 K. They suggest that a number of dense clumps moving at high velocity is required to excite this line. The observations were made with the 30-m IRAM millimeter telescope, so there is no spatial information available.

Observations of several SiO isotope lines by Deguchi *et al.* (1983) were explained by the authors in terms of a rotating, edge-on disc around the central star. This is based solely on spectral line data—again there is no spatial information.

Lucas *et al.* (1992) observed the thermal SiO $\nu = 0$ emission from VY CMa, but were unable to resolve it in the 3 arcsecond beam of the IRAM interferometer.

Wright *et al.* (1990) used the BIMA array to fix a position for the SiO 43-GHz masers. This, and the position of the 22-GHz H_2O emission and OH masers, is displaced about 1 arcsecond to the east of most determinations of the optical position. There is quite a spread in the various optical measurements because of the confusing nebulosity; it is entirely possible that the star itself cannot be seen at all, and that the optical astrometry is led astray by bright clumps.

5.3.4 Dust

Flux

Knapp, Sandell and Robson (1993) used the JCMT bolometer to measure the flux from several stars, including VY CMa, at eight different frequencies in the millimetre and submillimetre band. For VY CMA they determined a continuum flux of 1.7 Jy at a frequency of 351 GHz (after the flux from the CO (3–2) transition was subtracted) and a spectral index of 2.4 (i.e. $S \propto \nu^{2.4}$). They reasoned that the emission must be from warm dust, since the stellar photosphere only contributes ~0.05 Jy at 1.1 mm and there is probably no ionized wind present. They also compile a spectral energy distribution for wavelengths between 10 cm and 1 μm using previously measured fluxes, and show that it is consistent with dust emission. Recall that the 345-GHz continuum flux observed with the JCMT–CSO Interferometer is about 0.8 Jy—approximately 50% of the single-dish flux. I conclude that there must be a bright and compact dust component.

IR interferometry

McCarthy (1979) determined dimensions of 0.5×0.7 arcseconds for the emission at 10.2 μm, using infrared interferometry. The longer axis is oriented in a roughly east-west direction. These measurements reflect a product of the dust density and temperature distributions, and are sensitive to hotter dust than that detected by our interferometer at longer wavelengths.

Polarization

Measurements of the polarization seen at optical and infrared wavelengths are sensitive only to the spatial distribution of dust in the region, not the temperature; we therefore expect that they will reflect the dust density variation over much larger scales than the interferometric observations described above. Daniel (1982) finds that the position angle of the polarization varies, and indeed reverses, at a wavelength of 1 μm or so. The model he uses to explain this behaviour requires a relatively dense distribution of dust elongated in the north-south direction which dominates the scattering at longer wavelengths, and more diffuse dust in the east-west direction. This diffuse component becomes important at shorter wavelengths as the denser dust becomes optically thick. At an intermediate

wavelength, where neither of the two regions is dominant, the polarization will rotate through 90°, as observed.

5.3.5 Single-dish CO (3–2)

Spectra taken of the CO (3–2) transition (K. Young, private communication) with the CSO have a peak antenna temperature of about 1.5 K (≈60 Jy) with a central peak at +18 km s^{-1} and two side peaks of 80% strength at 0 and +40 km s^{-1}. The profile is roughly symmetric about the accepted stellar velocity of 19 km s^{-1}.

It is likely that the CO (3–2) emission from VY CMa is optically thick, for two reasons. Firstly, the lines that make up the single-dish CO (3–2) spectrum have parabolic profiles with relatively flat tops, characteristic of optically-thick emission. Secondly, Heske *et al.* found that the envelopes of all the OH/IR stars in their sample (which did not, however, include VY CMa) are optically thick in both the CO (1–0) and the CO (2–1) transitions, strongly suggesting that this is also true of the CO (3–2) line for such stars.

5.3.6 Summary of the data for VY CMa

Any model that hopes to describe the nature of VY CMa should provide an explanation for several key pieces of observational data:

- The bright component of the 321-GHz emission comes from a region not larger than 100×50 mas and is offset from the continuum emission by at least 200 mas.
- There is a compact continuum source (it must be of order 500×500 mas to be unresolved by the JCMT–CSO Interferometer) which emits approximately half of the 800 μm flux.
- Compact CO emission is seen with a distinct red wing. The peak flux is ~10% of the single-dish flux.
- The 22-GHz integrated H_2O maser emission is elongated at a position angle of about 150°. 1612-MHz OH maser emission shows elongation at a similar angle, but out to a radius of 2 arcseconds. Extreme red- and blue-shifted velocities are offset at a position angle of 50°, with the red to the north-east.

- VY CMa is probably the source of a cone of UV radiation that lights up the nearby molecular cloud to produce an excited rim. The cone is inferred to have a roughly east-west axis. An M-type star does not normally produce UV flux.

- Because there is a relatively intact molecular cloud nearby, VY CMa could not have been an O star. There must therefore be a flaw in the arguments that suggest a minimum mass of $\sim 15\ M_\odot$.

5.4 Models

It is clear that the circumstellar envelope of VY CMa is not spherically symmetric; much of the data suggests that there may be axial symmetry. The same preferred directions, namely position angles of $\sim 50°$ and $\sim 150°$, crop up repeatedly, and are best illustrated by the map of the integrated MERLIN emission (Figure 5.7). My first criterion for models is that they should conform to this orientation.

CO line profiles

My interpretation of the CO (3–2) data is common to all three models that I put forward and is depicted schematically in Figure 5.9. The idea is similar to the explanation of asymmetric line profiles seen towards protostars with infall (Zhou *et al.*, 1993), except that in this case there is outflow. It is likely that the CO (3–2) emission from VY CMa is optically thick (§5.3.5), so that there is absorption of the emission from the relatively hot and compact blue-shifted material just in front of the star by the cooler and more extended blue-shifted gas closer to us. The blue-shifted emission that finally escapes towards us is detected by a single dish, but is resolved out by the interferometer. For the hot, compact, red-shifted gas just on the far side of the star, however, there is very little cooler gas at the same velocity along its path to us and consequently very little absorption. This emission is detected by both a single dish and the interferometer. The CO line measured with the interferometer therefore appears to have a red wing.

It is possible that the distinct emission lines we detect with the interferometer in the red wing (Figures 5.3a and 5.4a) originate in clumps close to the star, similar to those proposed by Cernicharo, Bujarrabal and Santarén (1993) to explain the SiO $\nu = 4$, $J = 5$–4 maser line.

I now describe three models that attempt to account for the rest of the data.

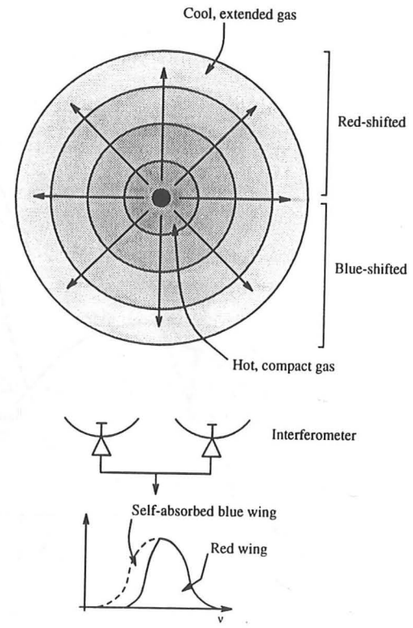


Figure 5.9: A simple explanation for the apparent red wing of the CO (3–2) line seen with the interferometer.

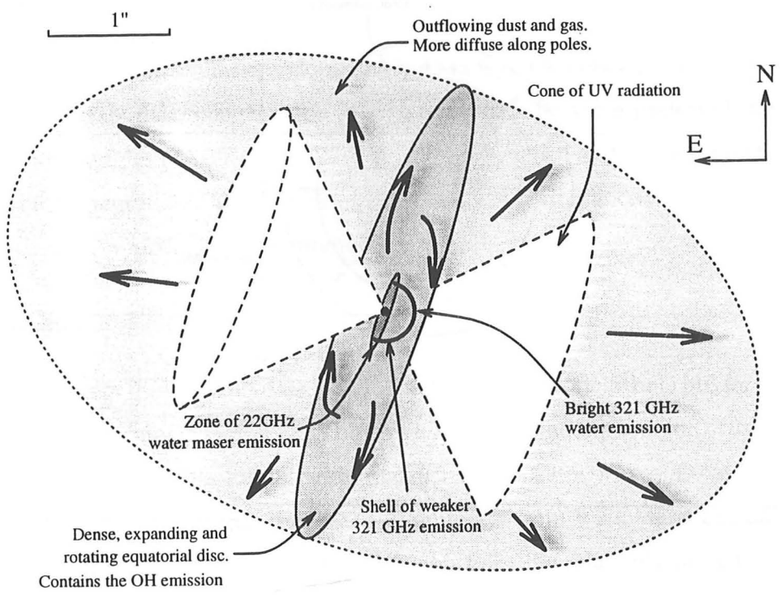


Figure 5.10: Model 1: Single-star model for VY CMa with the 321-GHz H_2O emission offset from the star (see main text for description).

5.4.1 Single-star models

Model 1: 321-GHz maser emission displaced from the star

Figure 5.10 shows the single-star scenario which is most consistent with the data and has the bright 321-GHz maser emission offset to the west of the star.

The star has an equatorial density enhancement in the form of an expanding and rotating disc. The plane of the disc is quite highly inclined with respect to the plane of the sky, with its axis pointing towards us in the north-east, in line with the 1612-MHz OH maser observations. The polar regions contain more diffuse material that is being accelerated away from the star. The continuum emission is centred on the star with the 321-GHz emission arising in a hemispherical shell to the west. UV radiation from the star escapes near the polar axis where the dust is more diffuse than at the equator. The pros and cons of this model are summarized below.

For:

- √ Simple model.
- √ Accounts reasonably well for the distribution of 22-GHz H_2O maser data.
- √ Fits the distribution of dust required by optical polarization.
- √ Strong 321-GHz H_2O emission where the shell is seen tangentially, i.e. where the velocity gradient along the line of sight is minimal. This will naturally tend to be near the stellar velocity.
- √ The higher doppler velocities seen in 321-GHz H_2O emission will be more along the line of sight to the star than the main emission and will therefore appear to be closer to the star.
- √ Fits the 1612-MHz OH maser data.

Against:

- × Why is there no 321-GHz H_2O maser emission to the east of the star?
- × The bright component of the 321-GHz H_2O emission is very compact over a relatively wide range of velocities. This model does not provide an obvious explanation.

× The 321-GHz line needs a high excitation temperature (~2000 K). It is strange that it occurs relatively distant from the star. Shock excitation would be needed.

× In this model it is hard to see how such a large fraction of the continuum emission seen from the extensive nebulosity surrounding VY CMa can be attributed to the inner regions near the star.

× How does VY CMa excite the nearby bright rim? There is no obvious mechanism to generate UV flux, so the projection of the star onto the tip of the rim must be a coincidence.

× No resolution of the O star problem.

I find this model unconvincing in its explanation of the 321-GHz emission offset from such a bright and compact continuum source. The remaining models that I consider have the 321-GHz H_2O maser emission centred on the star.

Model 2: 321-GHz emission centred on the star

Because the excitation temperature for H_2O maser emission is higher for the 321-GHz line than for the 22-GHz line, it is natural to expect the former to be found nearer the star. Here we require that the strong, compact continuum emission source is displaced to the east of the star. Figure 5.11 depicts the most likely candidate with this geometry.

This model has the same equatorial disc as before, in the same orientation. The 321-GHz H_2O maser emission is centred on the star. The continuum comes from free-free emission of shocked gas, resulting perhaps from the interaction of fast, outflowing material from the star with slower material along the polar axis. The emission to the west of the star is obscured by the optically-thick equatorial disc.

For:

√ Relatively simple model.

√ Good agreement with 22-GHz emission. The two distinct regions of emission along the polar axis (Figure 5.7) correspond to the location of shocked continuum emission.

√ Fits the 1612-MHz OH maser data.

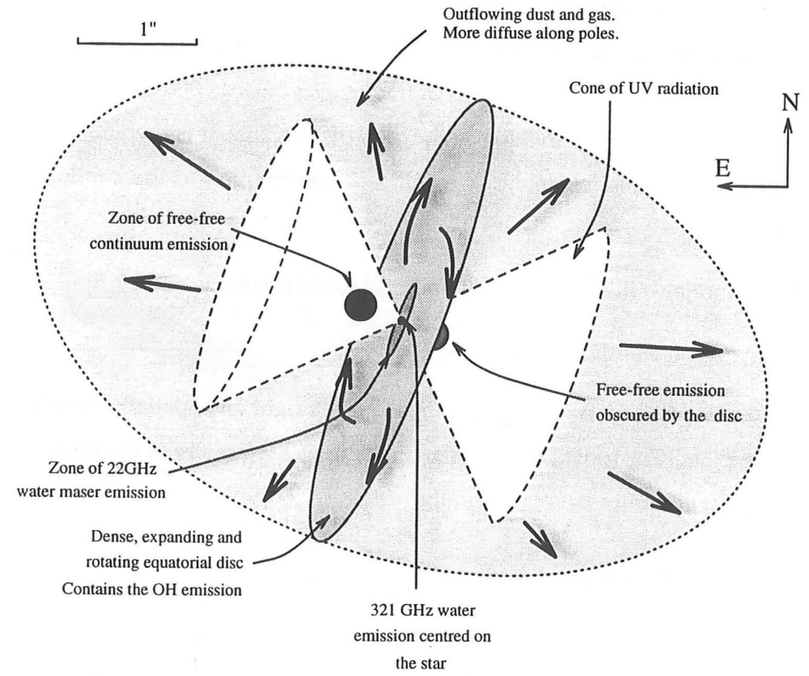


Figure 5.11: Model 2: Single-star model for VY CMa with the 321-GHz H_2O emission centred on the star (see main text for description).

× The 321-GHz line needs a high excitation temperature (~2000 K). It is strange that it occurs relatively distant from the star. Shock excitation would be needed.

× In this model it is hard to see how such a large fraction of the continuum emission seen from the extensive nebulosity surrounding VY CMa can be attributed to the inner regions near the star.

× How does VY CMa excite the nearby bright rim? There is no obvious mechanism to generate UV flux, so the projection of the star onto the tip of the rim must be a coincidence.

× No resolution of the O star problem.

I find this model unconvincing in its explanation of the 321-GHz emission offset from such a bright and compact continuum source. The remaining models that I consider have the 321-GHz H_2O maser emission centred on the star.

Model 2: 321-GHz emission centred on the star

Because the excitation temperature for H_2O maser emission is higher for the 321-GHz line than for the 22-GHz line, it is natural to expect the former to be found nearer the star. Here we require that the strong, compact continuum emission source is displaced to the east of the star. Figure 5.11 depicts the most likely candidate with this geometry.

This model has the same equatorial disc as before, in the same orientation. The 321-GHz H_2O maser emission is centred on the star. The continuum comes from free-free emission of shocked gas, resulting perhaps from the interaction of fast, outflowing material from the star with slower material along the polar axis. The emission to the west of the star is obscured by the optically-thick equatorial disc.

For:

√ Relatively simple model.

√ Good agreement with 22-GHz emission. The two distinct regions of emission along the polar axis (Figure 5.7) correspond to the location of shocked continuum emission.

√ Fits the 1612-MHz OH maser data.

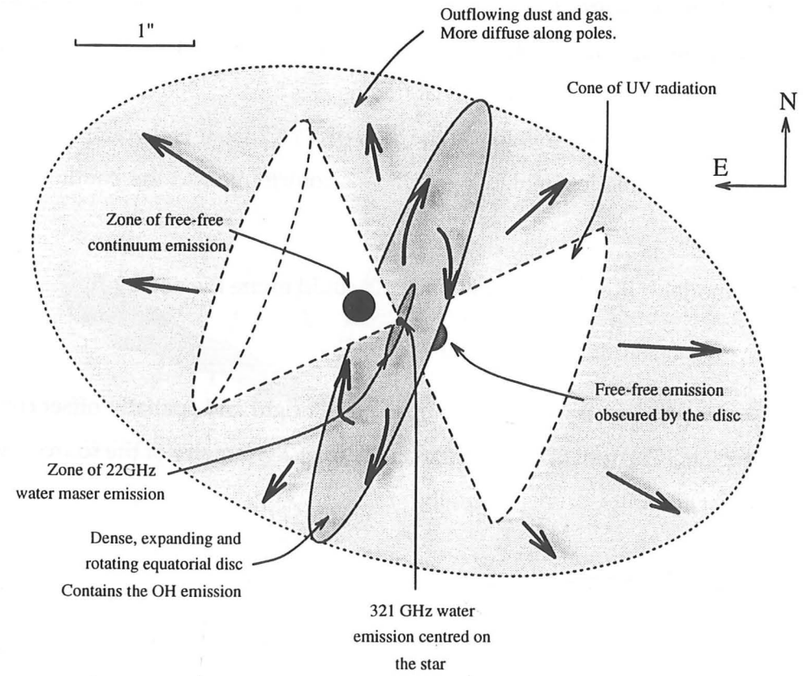


Figure 5.11: Model 2: Single-star model for VY CMa with the 321-GHz H_2O emission centred on the star (see main text for description).

√ Dust distribution fits the optical polarization data.

√ The weak 321-GHz component is further away from the star to the east. Corresponding emission to the west is hidden by the disc.

Against:

× The spectral energy distribution (SED) for VY CMa, presented by Knapp, Sandell and Robson (1993), is well-fitted by dust emission. In the same paper, they show the SED for the source CRL 618; in addition to a dust component similar to that of VY CMa, this bipolar reflection nebula has an HII region that emits ~1 Jy at 345 GHz. This contribution shows up clearly at frequencies below ~300 GHz. There is no such evidence for free-free emission in the SED of VY CMa. It is therefore very unlikely that free-free emission makes a significant contribution to the continuum flux of VY CMa.

× As for model 1, it is unclear how VY CMa could excite the nearby rim.

× There is no resolution of the O star problem.

I find this model less feasible than model 1; the bright and spatially offset continuum is hard to explain. The root of the problem is the basic asymmetry of the source—we need something that accounts for this naturally.

5.4.2 Model 3: A binary system

A binary companion of the main star is the obvious solution to the required asymmetry. Figure 5.12 illustrates the possibilities of this model for explaining much of the observational data.

A model of this type was first suggested by Morris (1987) to explain bipolar nebulae. The evolved star starts to lose mass in the form of a wind, part of which is intercepted by the secondary which forms an accretion disc. The secondary is unable to accrete matter as fast as it is gained from the primary, and the disc grows; once it fills the Roche lobe the material spills out to form a circum-binary disc and, possibly, a disc around the primary. The secondary resembles a young protostar; as well as the accretion disc it is likely to have a high velocity outflow along the polar axis. The inner edges of the disc may give rise to UV emission which is radiated outwards in a cone around the poles, directed by the

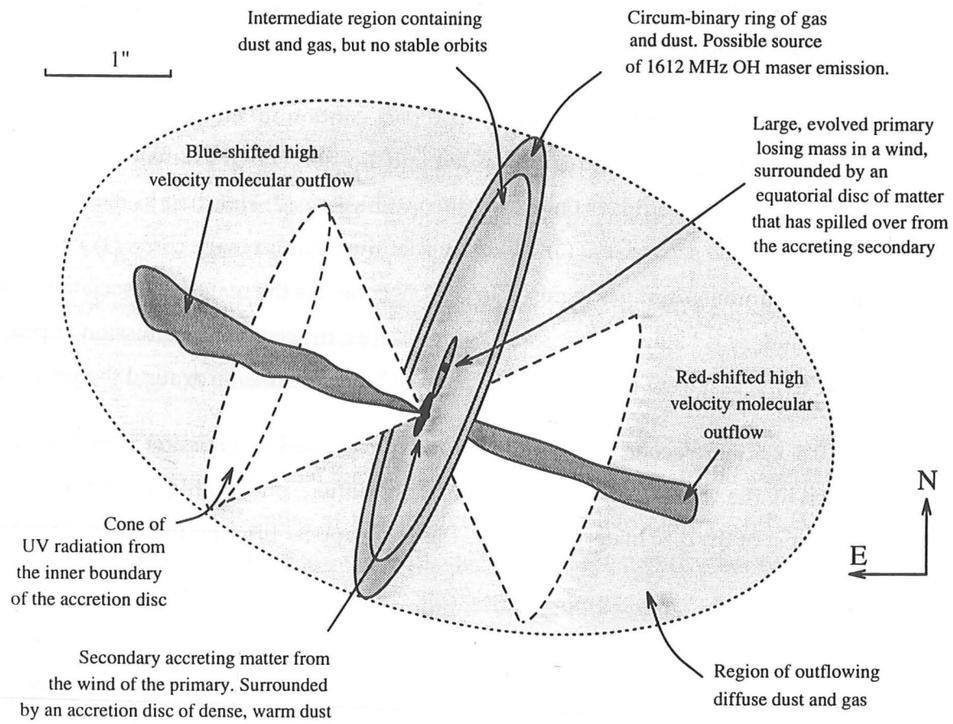


Figure 5.12: A binary model for VY CMa (see main text for description).

√ Dust distribution fits the optical polarization data.

√ The weak 321-GHz component is further away from the star to the east. Corresponding emission to the west is hidden by the disc.

Against:

× The spectral energy distribution (SED) for VY CMa, presented by Knapp, Sandell and Robson (1993), is well-fitted by dust emission. In the same paper, they show the SED for the source CRL 618; in addition to a dust component similar to that of VY CMa, this bipolar reflection nebula has an HII region that emits ~1 Jy at 345 GHz. This contribution shows up clearly at frequencies below ~300 GHz. There is no such evidence for free-free emission in the SED of VY CMa. It is therefore very unlikely that free-free emission makes a significant contribution to the continuum flux of VY CMa.

× As for model 1, it is unclear how VY CMa could excite the nearby rim.

× There is no resolution of the O star problem.

I find this model less feasible than model 1; the bright and spatially offset continuum is hard to explain. The root of the problem is the basic asymmetry of the source—we need something that accounts for this naturally.

5.4.2 Model 3: A binary system

A binary companion of the main star is the obvious solution to the required asymmetry. Figure 5.12 illustrates the possibilities of this model for explaining much of the observational data.

A model of this type was first suggested by Morris (1987) to explain bipolar nebulae. The evolved star starts to lose mass in the form of a wind, part of which is intercepted by the secondary which forms an accretion disc. The secondary is unable to accrete matter as fast as it is gained from the primary, and the disc grows; once it fills the Roche lobe the material spills out to form a circum-binary disc and, possibly, a disc around the primary. The secondary resembles a young protostar; as well as the accretion disc it is likely to have a high velocity outflow along the polar axis. The inner edges of the disc may give rise to UV emission which is radiated outwards in a cone around the poles, directed by the

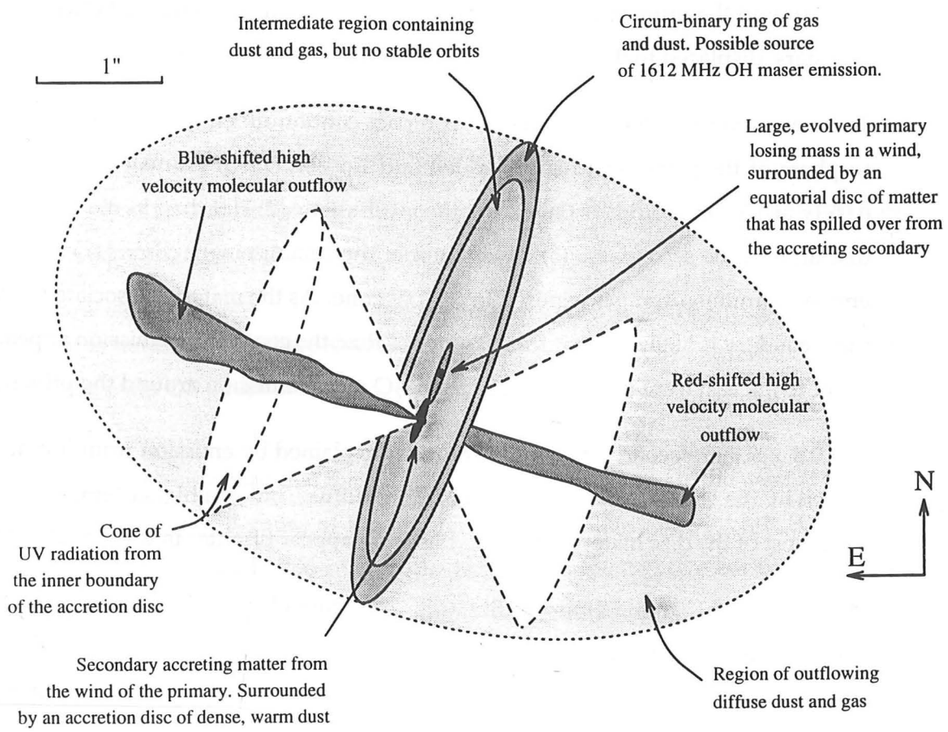


Figure 5.12: A binary model for VY CMa (see main text for description).

surrounding dust in the disc. A region of ionization moves out from the apex of the cone. The wind from the primary can move faster along the polar axis and is more diffuse there. Maser emission, particulary the 1612-MHz OH line, may be affected by the cone of UV flux from the accreting secondary (Lewis, 1992).

For:

√ This model explains the two components of the 321-GHz H_2O maser: strong emission from around the primary and a weaker component arising somewhere between the two stars, displaced to the east.

√ There are several potential sources of compact continuum emission: (1) The photosphere of the primary; Knapp, Sandell and Robson (1993) estimated a value of 0.05 Jy at 1.1 mm, so this is only a small contribution. (2) Hot dust in the accretion disc around the secondary. (3) Warm dust in the circum-binary disc. (4) Free-free emission from ionized gas exposed to the UV cone. As the material associated with the secondary is likely to be a strong source of flux, the continuum emission appears offset to the east from the strong 321-GHz H_2O maser emission around the primary.

√ The 0.5 × 0.7 arcsecond far-infrared source is explained by emission from hot dust heated by the UV radiation. This dust is more diffuse, but at a higher temperature, than most of the disc material, and can therefore appear brighter in the far-infrared.

√ The 22-GHz H_2O maser emission fits well. The main elongated region is emission from the disc of material around the primary. The blobs along the polar axis may arise where the wind from the primary is shocked as it encounters ionized material in the UV cone.

√ There are two possible ways to account for the spatial distribution of the 1612-MHz OH masers. One option is that they are located in the circum-binary disc. The alternative is that the emission would be spherically symmetric about the primary but for the UV radiation, which inhibits maser activity along the polar directions. The OH masers therefore appear to be distributed in a thick, expanding disc.

√ The optical polarization data are consistent with this model. The dust distribution has a bipolar morphology with the right orientation.

√ Dissipation in the inner regions of the accretion disc provide a natural source of UV radiation.

√ This model suggests an answer to the problem of the nearby, relatively intact molecular cloud. It seems likely that the accreting secondary is responsible for a large part of the luminosity attributed to VY CMa, so it need not have enjoyed O star status in its younger days. If it was not an O star, then there would have been no UV flux to wipe out the molecular cloud.

Against:

× The model is more complicated than single-star scenarios and it is therefore easier to accommodate a wide range of observations.

× It is not clear why there should be high velocity red-shifted gas to the south-east of the main CO emission.

So far I have assumed that the 321-GHz H_2O line is 250 mas to the west of the compact continuum source. Because of the ambiguity of the phase, however, it is possible that the H_2O line is actually 750 mas to the east. This is attractive because the accreting secondary might emit substantial amounts of visible radiation; the optical position would then appear an arcsecond or so to the west of the maser activity associated with the main star, as observed by Wright *et al.* (1990). The disadvantage is that this no longer accounts for the excess matter seen to the east of the primary (1612-MHz OH masers and weak component of 321-GHz H_2O masers). In the model that I set out, it is possible that Herbig-Haro–type objects in the fast outflow could generate bright optical emission that would confuse the astrometry. Observations over a wider range of hour angles may help to resolve this issue.

The model predicts that there is a fast outflow, probably oriented as shown in Figure 5.12, and that the spectral energy distribution can be modelled by the combination of an evolved star and an accreting secondary that provides a substantial fraction of the overall luminosity.

surrounding dust in the disc. A region of ionization moves out from the apex of the cone. The wind from the primary can move faster along the polar axis and is more diffuse there. Maser emission, particulary the 1612-MHz OH line, may be affected by the cone of UV flux from the accreting secondary (Lewis, 1992).

For:

√ This model explains the two components of the 321-GHz H_2O maser: strong emission from around the primary and a weaker component arising somewhere between the two stars, displaced to the east.

√ There are several potential sources of compact continuum emission: (1) The photosphere of the primary; Knapp, Sandell and Robson (1993) estimated a value of 0.05 Jy at 1.1 mm, so this is only a small contribution. (2) Hot dust in the accretion disc around the secondary. (3) Warm dust in the circum-binary disc. (4) Free-free emission from ionized gas exposed to the UV cone. As the material associated with the secondary is likely to be a strong source of flux, the continuum emission appears offset to the east from the strong 321-GHz H_2O maser emission around the primary.

√ The 0.5 × 0.7 arcsecond far-infrared source is explained by emission from hot dust heated by the UV radiation. This dust is more diffuse, but at a higher temperature, than most of the disc material, and can therefore appear brighter in the far-infrared.

√ The 22-GHz H_2O maser emission fits well. The main elongated region is emission from the disc of material around the primary. The blobs along the polar axis may arise where the wind from the primary is shocked as it encounters ionized material in the UV cone.

√ There are two possible ways to account for the spatial distribution of the 1612-MHz OH masers. One option is that they are located in the circum-binary disc. The alternative is that the emission would be spherically symmetric about the primary but for the UV radiation, which inhibits maser activity along the polar directions. The OH masers therefore appear to be distributed in a thick, expanding disc.

√ The optical polarization data are consistent with this model. The dust distribution has a bipolar morphology with the right orientation.

√ Dissipation in the inner regions of the accretion disc provide a natural source of UV radiation.

√ This model suggests an answer to the problem of the nearby, relatively intact molecular cloud. It seems likely that the accreting secondary is responsible for a large part of the luminosity attributed to VY CMa, so it need not have enjoyed O star status in its younger days. If it was not an O star, then there would have been no UV flux to wipe out the molecular cloud.

Against:

× The model is more complicated than single-star scenarios and it is therefore easier to accommodate a wide range of observations.

× It is not clear why there should be high velocity red-shifted gas to the south-east of the main CO emission.

So far I have assumed that the 321-GHz H_2O line is 250 mas to the west of the compact continuum source. Because of the ambiguity of the phase, however, it is possible that the H_2O line is actually 750 mas to the east. This is attractive because the accreting secondary might emit substantial amounts of visible radiation; the optical position would then appear an arcsecond or so to the west of the maser activity associated with the main star, as observed by Wright *et al.* (1990). The disadvantage is that this no longer accounts for the excess matter seen to the east of the primary (1612-MHz OH masers and weak component of 321-GHz H_2O masers). In the model that I set out, it is possible that Herbig-Haro–type objects in the fast outflow could generate bright optical emission that would confuse the astrometry. Observations over a wider range of hour angles may help to resolve this issue.

The model predicts that there is a fast outflow, probably oriented as shown in Figure 5.12, and that the spectral energy distribution can be modelled by the combination of an evolved star and an accreting secondary that provides a substantial fraction of the overall luminosity.

5.5 Similar sources

I believe that this is convincing evidence in support of a binary star scenario for VY CMa. Other evolved stars are also suspected of having companions. I have already mentioned that Morris and Bowers (1980) found the 1667-MHz OH maser emission from VY CMa to have features in common with those observed towards the bipolar nebulae OH 231.8+4.2 and M1-92, both of which may well be binaries; M1-16 also appears to be of a similar nature:

OH 231.8+4.2: Outflow velocities of up to $\approx$200 km s^{-1} are seen along the polar directions for OH231.8 (Bowers, 1991). Reipurth (1987) found Herbig-Haro–type objects indicative of a highly collimated outflow which, according to Podsiadlowski and Clegg (1992), "can only be understood if the central system is a binary consisting of a red giant and a more compact companion which provides the source of the fast outflow". Excess blue emission in the reflection spectrum of the nebula provides additional evidence for the presence of a companion.

M1-92 (Minkowski's Footprint): This is also a bipolar nebula, with two lobes separated by a gap. The northern lobe is brighter than the southern one and is believed to be moving towards us, unobscured by a dense equatorial disc. There is also a fast wind which, together with the disc, has led to confusion over the evolutionary state of this object—protostar or evolved star? (Seaquist and Plume, 1991). On the basis of observations from the IUE satellite, Feibelmann and Bruhweiler (1990) "strongly favor that the central object of M1-92 is an evolved binary system . . . in the process of forming a planetary nebula."

M1-16: Schwarz (1992), using long-slit optical spectroscopy, finds radially opposed sets of bubbles or lobes in this evolved object with projected shock velocities of $\approx$230 km s^{-1}. He concludes that M1-16 is a binary system and that bipolar transition objects all arise from binaries and go on to form bipolar planetary nebulae.

5.6 Future observations and analysis

Several observations would further test the binary model for VY CMa. The data from the JCMT–CSO Interferometer for the high velocity CO outflow are not as clear as I would like, and this should be confirmed by further observations at hour angles around +3. A map of the CO (2–1) emission using one of the millimetre arrays should clearly show the fast outflow, if it is present, and I have applied for time at OVRO to do this. This outflow may also be the source of shocked emission as it runs into the slower wind from the primary. Optical observations looking for shocked Hα and/or [SII] emission would be a good test, although I am not sure how easy this would be in the presence of such bright and extended nebulosity. If the excited rim of the nearby cloud is due to UV radiation from the secondary, then it may be possible to detect free-free continuum emission from within the cone of radiation. With the VLA it should be feasible to show that this continuum emission is displaced eastwards of the bulk of the 22-GHz maser emission, which is centred on the primary, using the phase-referencing technique of Reid and Menten (1990).

It would also be valuable to see if the spectral energy distribution for VY CMa can be adequately modelled in terms of a binary system with both accretion and circumbinary discs. This might give some indication of the mass and luminosity of the secondary. The mass of the primary can then be re-assessed to demonstrate that it need not have been an O star.

5.7 Summary and conclusions

The JCMT–CSO Interferometer observations of VY CMa have provided evidence for a new interpretation of this puzzling source. Both a summary of our data and my conclusions are listed below:

- On the basis of the brightness temperature inferred for the emission, we can confirm that the 321-GHz H_2O emission from VY CMa is the result of maser activity.
- The 321-GHz H_2O masers are spatially offset from a compact $\sim$1 Jy continuum source.
- The CO (3–2) line at 345 GHz was also detected with the interferometer. It has a pronounced red wing of emission. The continuum was again present at $\sim$1 Jy.

- Single-star models have trouble explaining the offset of the 321-GHz masers from the bright and compact continuum. Our data point to a binary model for VY CMa along the lines suggested by Morris (1987). This is corroborated by much of the other observational evidence that is available.
- The binary model suggests that much of the luminosity of the system is due to an accreting secondary, so that the primary does not have to be as spectacularly luminous as has previously been thought.
- VY CMa is probably a very young bipolar nebula that will develop into an object resembling OH 231.8+4.2 and M1-92.
- These results lend support to the theory that some, if not all, bipolar nebulae are the result of binary systems, and that this is an important cause of non-spherical symmetry in general. The high velocity outflows observed for many of these sources arise from the secondary, in a similar way to the outflows from accreting protostars.

Chapter 6

OBSERVATIONS OF PROTOSTARS

6.1 Introduction

6.1.1 A simple picture of star formation

The basic process of star formation is now fairly well-understood. In the depths of molecular clouds, where the external radiation field is greatly attenuated by the large optical depth, gas cools and condenses under gravity into dense clumps. The in-fall of gas onto a central core starts at the middle and proceeds outwards; this is dubbed 'inside-out collapse'. It is likely that magnetic fields exert a braking influence, since ions in the gas are effectively tied to the field lines, and are also coupled to the neutrals through collisions. For low-mass stars, the initial rate of collapse is largely controlled by the speed at which the neutral particles slip past the bound ions—a process known as ambipolar diffusion.

All molecular clouds have some net rotation, and conservation of angular momentum ensures that the material cannot simply fall directly onto the central point. Hence there must always be some form of accretion disc, where the dust and gas move in roughly circular, keplerian orbits about the centre. Matter continues to accrete onto the central core, losing its angular momentum in a mechanism that is not yet fully comprehended. As gas falls into the centre, the loss of gravitational potential energy causes heating of the core. Nuclear fusion of hydrogen into helium, which will fuel the star for the major part of its life, commences once a critical temperature and pressure have been reached. Eventually a stellar wind is formed from particles that 'boil off' the tenuous photosphere and get propelled away from the star by the photon radiation field. This wind clears away most of the material that is still falling inwards from the parent cloud, finally unveiling the underlying star.

Observations have shown that outflowing 'jets' are often associated with young protostars. The detailed mechanics of this phenomenon, where material is ejected at high

speeds along the polar axis, are still poorly understood. A similar process is seen in active galactic nuclei. It seems probable that directed outflow is a natural result of the accretion of material that has net angular momentum, and that magnetic fields play an important part.

In this chapter I concentrate on the accretion discs expected to occur around young, low-mass stars.

6.1.2 Protostellar accretion discs: theory and previous observations

In 1966, Mendoza discovered that some young, visible stars appear to have an excess of infrared radiation compared to normal stars, and this was soon attributed to the presence of warm dust. The amount of dust required, if distributed with spherical symmetry, would easily obscure the star at optical wavelengths, and so flattened discs of dust were invoked as an explanation. Although there is thick dust in equatorial regions, the polar directions are relatively clear and allow us to see the central star. The solar system provides further indirect evidence for the existence of discs around young stars; the planets are believed to have condensed out of a disc-like structure of gas and dust soon after the sun was formed. It is estimated that a mass of at least 0.02 $M_\odot$ was required to form the planetary bodies of the solar system, and comets can be found out to distances of $\sim$100 AU. Theory also predicts that accretion discs will have radii of about 100 AU (Adams *et al.*, 1987; Galli and Shu, 1993). The nearest star-forming clouds are about 150 pc away, at which distance accretion discs will have an angular diameter of an arcsecond or so.

The study of the spectral energy distributions (SEDs) of young stars, from millimetre wavelengths to the near-infrared, has provided a large amount of indirect evidence for discs. Modelling the variation of single-dish flux with frequency has put constraints on the distribution of temperature and mass within the disc (Adams and Shu, 1986; Adams *et al.*, 1987; Kenyon and Hartmann, 1987; Butner *et al.*, 1994). The survey of pre-main-sequence stars by Beckwith *et al.* (1990) found indications of discs in 42% of the objects observed.

Infrared speckle imaging has been used to observe T Tauri stars at high angular resolution. These are young stars that have blown away most of the surrounding material, but still have the remnants of a dusty accretion disc. HL Tau has probably received the most attention from the speckle observers; Grasdalen (1984) found spatially extended emission at 2 μm, elongated in the same direction as the outflow. It is probable that most of the light detected does not come from the disc at all, but is radiation from the star that is scattered off the more diffuse dust above and below the disc (Yamashita *et al.*, 1994). This highlights the need for observations at longer wavelengths that detect the thermal dust emission, rather than scattered light.

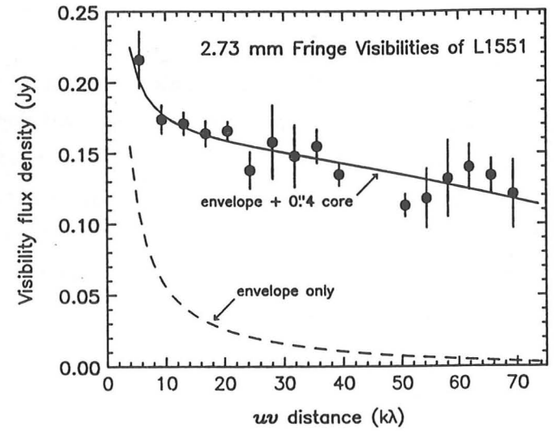


Figure 6.1: Visibility curve for the protostar L1551-IRS 5, from the observations of Keene and Masson (1990) using the Owens Valley Millimeter Interferometer

Sargent and Beckwith (1987) used the Millimeter Interferometer at the Owens Valley Radio Observatory (OVRO) to look for molecular line emission from HL Tau. They found evidence for a large disc of gas, some 2000 AU in radius, and claimed that there was rotation present. The argument for rotation of this 'molecular disc' has since been thrown into doubt, and further observations (Hayashi *et al.*, 1993) show that this gas is more likely to be the remains of the in-falling envelope—the 'pseudo-disc' predicted by Galli and Shu (1993). The continuum emission from HL Tau was unresolved in the observations at 3 mm by Ohashi *et al.* (1991), at 2.65 mm by Sargent and Beckwith (1991), and at 1.4 mm by Woody *et al.* (1990). Molecular line observations of the quadruple source GG Tau also point to a large ($\sim$500 AU), rotating disc-like structure (Koerner *et al.*, 1993; Kawabe *et al.*, 1993; Dutrey *et al.*, 1994).

OVRO was also used by Keene and Masson (1990) to measure the continuum flux from the core of the well-known protostar L1551-IRS 5. The visibility curve that they obtained is shown in Figure 6.1. They attributed the shape of the curve to two components: a spatially extended envelope which gives the sharp peak on short baselines, and a barely resolved, central, compact source which gives the much flatter contribution. They believed that the compact core, with a radius $\leq$ 64 AU, is the accretion disc.

In summary, there is a large amount of convincing, but as yet fairly indirect, evidence to support the existence of accretion discs around young, low-mass stars.

6.1.3 The JCMT–CSO Interferometer as an instrument for observing discs

With a maximum baseline of ~200 kλ and a minimum fringe-spacing of about one arcsecond for operation at $\lambda = 0.8$ mm, our interferometer is ideal for observing accretion discs. Many protostars are obscured by their surrounding envelopes at wavelengths shorter than about 0.1 mm, so observations of their cores must be made at longer wavelengths. However, dust emission rises rapidly with frequency, and the sources become faint at long wavelengths. In addition, observations become confused by free-free emission, and it is therefore at submillimetre wavelengths that we stand the best chance of detecting warm, compact dust in the protostellar environment.

6.1.4 Binary protostellar systems

A recent study of T Tauri stars using infrared speckle imaging (Ghez *et al.*, 1993) revealed that at least 60% of them exist in multiple systems (binaries, triples, quadruples, etc.). This suggests that multiplicity is a common by-product of star-formation, or even a necessary ingredient, and it is important to be able to look for binaries at early stages of their formation.

Most of this chapter describes our observations of the two protostars L1551-IRS 5 and HL Tau (both of which appear to be single stars), and my subsequent analysis and interpretation of the data. The last section briefly describes my findings for the binary source NGC1333-IRAS4, and discusses the importance of submillimetre interferometry in the search for young, embedded binary systems.

6.2 Observations

6.2.1 Sources observed

We decided to concentrate our efforts on two well-known protostars, L1551-IRS 5 and HL Tau, both of which have prominent outflows and are strongly suspected of having discs. They are located in the Taurus–Auriga star-forming region, at an estimated distance of 140 pc (Elias, 1978) and a declination of +18°.

6.2.2 Observing technique and conditions

The observations were made over several nights, during the November 1993 interferometry run. Receiver B3i was used at the JCMT with a typical system temperature of ~1000 K; at CSO the 800 μm heterodyne receiver was used with a system temperature of ~500 K (single sideband). The DAS was used in the 10050 configuration (see Figure 2.2) generating

Night	Date	Sources	Frequency /GHz	Sensitivity /Jy K^{-1}	Data quality
3	4/11/93	HL Tau	345.785	?	Poor S/N
4	5/11/93	L1551-IRS 5	356.727	?	Poor S/N
5	6/11/93	L1551-IRS 5	356.726	110	Good data, more scatter than expected
6	12/11/93	L1551-IRS 5	345.785	140	Good data
		HL Tau	345.785	140	Good data
9	15/11/93	L1551-IRS 5	356.721	80	More sensitive by 170% Jump in flux?

Table 6.1: Observation details for L1551-IRS 5 and HL Tau

512 channels spanning 500 MHz, with a resolution of 1.25 MHz. Table 6.1 summarizes for each night the protostellar sources observed, the frequencies used, the sensitivity achieved and the quality of the data that was obtained.

The time-dependent phase response of the interferometer was not properly understood when the data were analysed, and there were systematic drifts in the phase of a source during a night's observing that could not be accounted for. These drifts could reach 360° per hour when the sources were close to the horizon, but were not significant over the 100-s periods used for coherent averaging. The sensitivity was determined from observations of the quasars 3C273 and 0528+134, with both the interferometer and a single dish. They are uncertain for Nights 3 and 4 because of the unreliable state of the system at that time. The quoted values have an associated error of about ±20%.

On Nights 3 and 6 the CO (3–2) transition fell within the passband. Examination of the spectra showed that the line contributes at most 5% of the overall flux and is not significant. As a check, the line was removed from the data and the results were not affected. In §6.6.2, I show that this line can give us valuable extra information.

6.3 Processing and analysis

6.3.1 Data processing

The data were processed using the INTEREDUCE data-reduction package (see chapter 4). A vector average was performed over the 500-MHz bandwidth, weighted to account for the system temperature, to give an amplitude and phase for every 10-s integration. These 10-s integrations were then coherently added to produce an amplitude and phase for every

100 s, which were recorded with the appropriate hour angle of the source and the noise level. This last figure was calculated from the prevailing system temperature. There was no significant decorrelation over this interval but, because of drifts in the instrumental phase over longer timescales, the phases were discarded at this stage and the analysis continued using the visibility amplitudes only.

6.3.2 Model fitting

For each 100-s data point, the corresponding values of u and v were calculated from the baseline components and the hour angle of the source. The data are a series of effective antenna temperatures (a measure of the visibility modulus) measured at certain points in the (u, v) plane. With a single-baseline interferometer the (u, v) coverage is fairly sparse (Figure 1.6) and to attempt to construct an image would be futile. Instead we must make a model of the brightness distribution and compare the visibilities predicted by the model with the actual data.

A χ^2 fit to a simple model

For any model of the brightness distribution, we must first Fourier transform it to the (u, v) plane before it can be tested against the data. In the first instance we compared some of the data with a circular Gaussian model for the brightness distribution, $B(\theta)$. This is readily transformed to the (u, v) plane:

$$B(\theta) \propto \exp\left(-\frac{D^2 \ln 2}{r_{1/2}^2}\,\theta^2\right) \overset{F-T}{\Longleftrightarrow} \mathcal{V}(u,v) = S_{\mathrm{T}} \exp\left(-\frac{\pi^2 r_{1/2}^2}{D^2 \ln 2}\,(u^2+v^2)\right), \qquad (6.1)$$

where S_{T} is the total flux, D is the distance to the star, and $r_{1/2}$ is the radius to the contour of half-maximium brightness. Simple, least-squares fitting was used to determine the best fit. An example for each source is shown in Figure 6.2, with the corresponding values of the S_{T} and $r_{1/2}$. These were useful for some quick, provisional estimates of the source size and flux, and clearly demonstrate that we have resolved the source of the submillimetre continuum flux in both cases. There are two problems with this simple fit, however. Firstly, the noise on the visibility measurements is not Gaussian; visibility is always positive, so that when the signal-to-noise ratio is very low, the data points tend to lie above the actual value. Instead we must use what is known as the Rice distribution. The second problem concerns the nature of the model, and I will tackle this issue after a description of the Rice distribution.

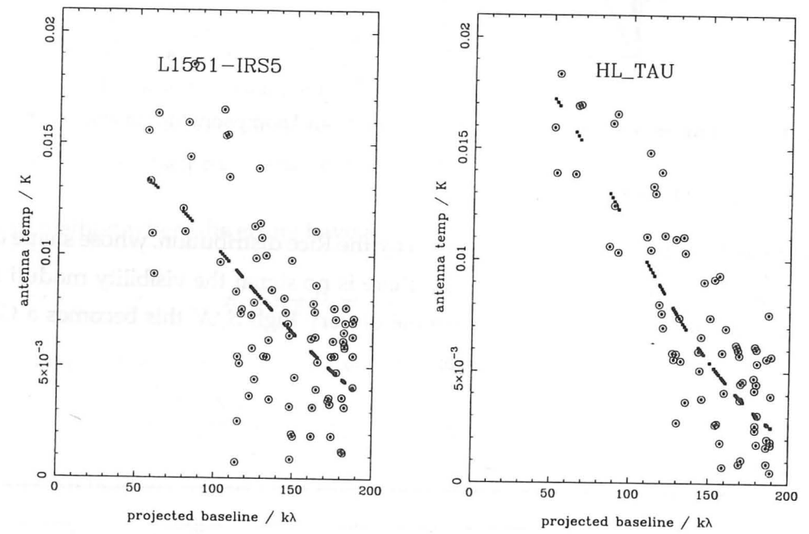


Figure 6.2: Visibility data and best-fitting curves for L1551-IRS 5 and HL Tau from the data of Night 6. The parameters for the best-fitting, circularly-symmetric models are: $S_{\mathrm{T}} = 2.1$ Jy (15 mK), $r_{1/2} = 50$ AU (0.″33) for L1551-IRS 5; $S_{\mathrm{T}} = 2.8$ Jy (20 mK), $r_{1/2} = 60$ AU (0.″42) for HL Tau.

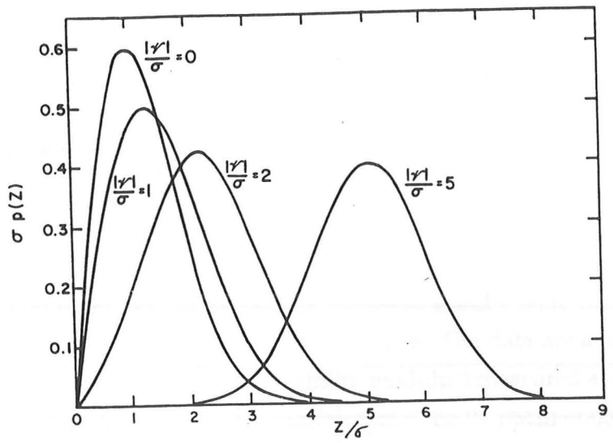


Figure 6.3: Examples of the Rice Distribution, taken from Thompson, Moran and Swenson (1986).

Using the Rice Distribution

The expected scatter of data points is given by the Rice distribution, whose shape depends on the signal-to-noise ratio, $\mathcal{S}/\mathcal{N}$. When there is no signal the visibility moduli follow a Rayleigh Distribution; in the other extreme of very high $\mathcal{S}/\mathcal{N}$ this becomes a Gaussian. The full probability distribution is given by

$$\mathcal{P}_{\rm Rice}(Z) = \frac{Z}{\sigma^2}\, exp\left(-\frac{Z^2+|\mathcal{V}|^2}{2\sigma^2}\right) I_0\left(\frac{Z\,|\,\mathcal{V}\,|}{\sigma^2}\right), \tag{6.2}$$

where Z is the modulus of the visibility that results from adding real and imaginary noise components of standard deviation σ to the true complex visibility $\mathcal{V}$. I_0 is the modified Bessel function of zero order. Some example curves are shown in Figure 6.3. Given a set of predicted visibilities, the actual measured values and the rms noise in the data, we can use the Rice distribution to evaluate the relative probability of obtaining the data, assuming the model to be correct. The probability of a model $\mathcal{M}$, given the dataset $\{\mathcal{D}_{\rm i}\}$ is given by

$$\mathcal{P} = Prob(\mathcal{M}|\{\mathcal{D}_{\rm i}\}) \tag{6.3}$$

$$\propto \prod_{\rm i} \mathcal{P}_{\rm i}, \tag{6.4}$$

where $\mathcal{P}_{\rm i}$ is the probability of an individual data point given the model:

$$\mathcal{P}_{\rm i} = Prob(\mathcal{D}_{\rm i}|\mathcal{M}). \tag{6.5}$$

The most probable model can be found by maximising $\log \mathcal{P}$;

$$\log \mathcal{P} = \sum_{\rm i} \log \mathcal{P}_{\rm i}. \tag{6.6}$$

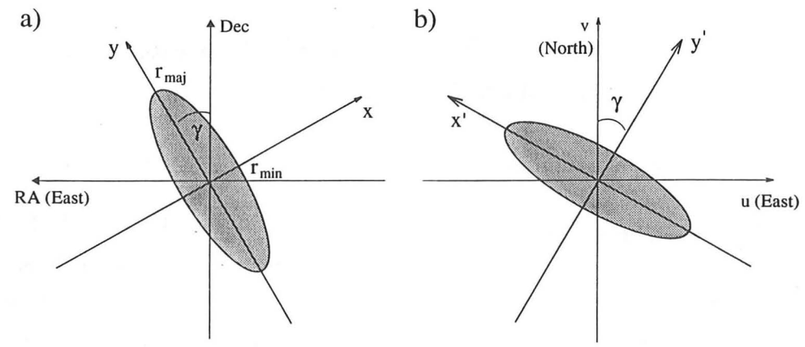


Figure 6.4: The elliptical model adopted for the brightness distribution. a) The model as seen on the sky and b) Fourier transformed into the (u, v) plane. The half-maximum contour is shown. Note the different east-west sense in the two domains.

In the χ^2 method where the errors have a Gaussian distribution,

$$K + \log \mathcal{P}_{\rm i} = -\left(\frac{\mathcal{D}_{\rm i} - \mathcal{M}}{\sigma}\right)^2 = -\chi^2_{\rm i}, \tag{6.7}$$

where K is a normalization constant, whereas with the Rice distribution

$$\log \mathcal{P}_{\rm i} = \log[\mathcal{P}_{\rm Rice}(\mathcal{D}_{\rm i})]. \tag{6.8}$$

Instead of minimizing χ^2 $(= \Sigma\chi^2_{\rm i})$, we must maximize the statistic $\log \mathcal{P}$ $(= \Sigma \log \mathcal{P}_{\rm i})$.

The Elliptical model

A circular disc gives an elliptical brightness distribution when seen in projection, and the model should allow for this. This means using a more complex model than the simple circularly-symmetric one used previously. Figure 6.4 shows the parameters adopted and their meaning in both the plane of the sky and the (u, v) plane. In the (u, v) plane, the model is given by

$$\mathcal{V}(u,v) = S_{\rm T} \exp\left(\frac{-\pi^2}{D^2 \ln 2}\left[r^2_{\rm min}(-u\cos\gamma + v\sin\gamma)^2 + r^2_{\rm maj}(u\sin\gamma + v\cos\gamma)^2\right]\right). \tag{6.9}$$

There are four free parameters; the total disc flux $S_{\rm T}$, the position angle of the major axis γ and the two radii to half-maximum brightness, $r_{\rm min}$ and $r_{\rm maj}$. A search was made over the 4-dimensional parameter space to find the most probable model. To do this I divided a given search range for each parameter into 11 values to be tried. All 11^4 (~20 000) possible

models were stepped through and the probability of obtaining the data assessed for each one.

Once I had found the best-fit combination of parameters, it was necessary to estimate the error in each. I did this by Monte Carlo simulation.

6.3.3 Monte Carlo error analysis

This powerful technique is well-described in *Numerical Recipes* (Press *et al.*, 1992) and I will merely outline it here. Basically, there is an ideal set of our four parameters, $\{I_j\}$, that most closely describe the brightness distribution on the sky. Because the data are noisy, the actual best-fit set of parameters to the data, $\{A_j\}$, is only an estimate of $\{I_j\}$ to within some set of errors, $\{\Delta A_j\}$. Here, the aim is to determine $\{\Delta A_j\}$. We assume that the noise level N_{sys}, derived from the system temperature, accounts for all the noise in the data (we can check this—see next section). The best-fit model gives the visibility $\mathcal{V}_m(u,v)$ in the absence of noise, but we can simulate artificial datasets with

$$\mathcal{V}_s(u,v) = \sqrt{(\mathcal{V}_m(u,v) + N_x)^2 + N_y^2}, \tag{6.10}$$

where N_x and N_y are both random numbers drawn from a Gaussian distribution with variance $N_{sys}^2/2$ (the complex noise vector with components N_x and N_y will have an rms length of N_{sys}). A simulated dataset $\{\mathcal{D}_i^{sim}\}$ contains visibility values for the same (u,v) points as the real dataset $\{\mathcal{D}_i\}$. For each simulated dataset we can determine the best-fit parameters $\{A_j^{sim}\}$; repeating the process N times will generate N best-fit values for each parameter, forming a distribution centred roughly on the best fit for the actual data, $\{\mathcal{D}_i\}$. The set of errors $\{\Delta A_j\}$ can be determined from the width of the distribution for each parameter. The following examples should clarify this brief outline.

6.3.4 The fits made to the data

I present all of our data in Figures 6.5 to 6.18. The first plot on each page shows the visibility modulus as a function of hour angle, with the best-fit model and ten simulated datasets. The corresponding coverage in the (u,v) plane is shown adjacent to this. Note that when the model has a low visibility, most of the data points have a higher flux, as expected from the Rice distribution; conversely, when the model visibility is high, the data are distributed more evenly on either side of the model value.

Histograms showing the error distributions of the four parameters are shown in the lower half of each page. They are derived from fits to 50 simulated datasets. The values fitted to the actual data are highlighted. In addition, the distribution of $\log \mathcal{P}$ (which is analogous to $-\chi^2$) is shown, but I have made no attempt to normalize it to the number of

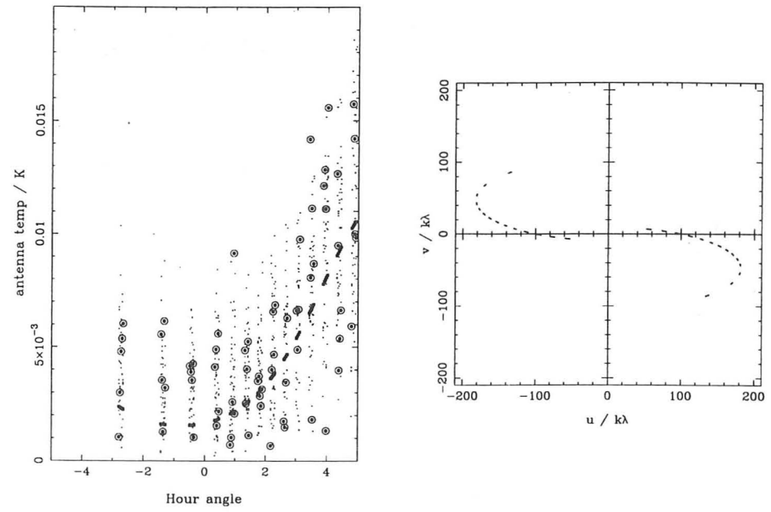


Figure 6.5: Data (circled dots), model (dashed line), simulated data (dots) and (u,v) track for HL Tau, taken on Night 3.

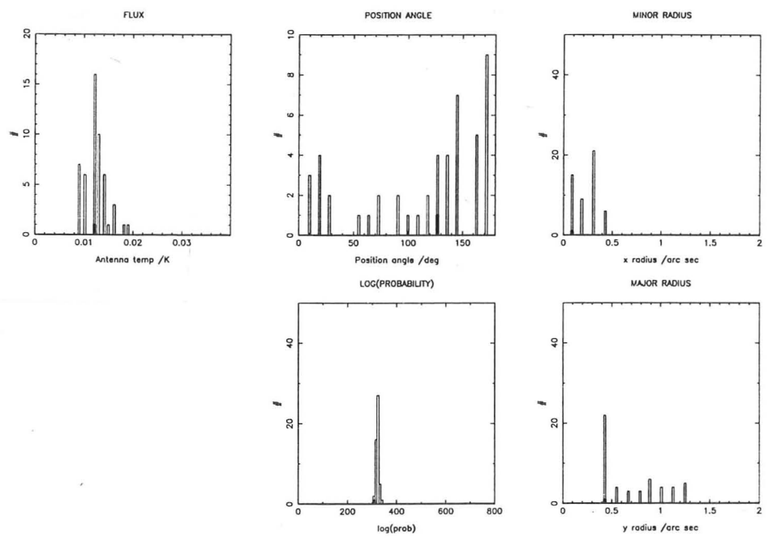


Figure 6.6: Histograms showing the results of fitting to the real dataset (highlighted) and the simulated datasets for HL Tau on Night 3.

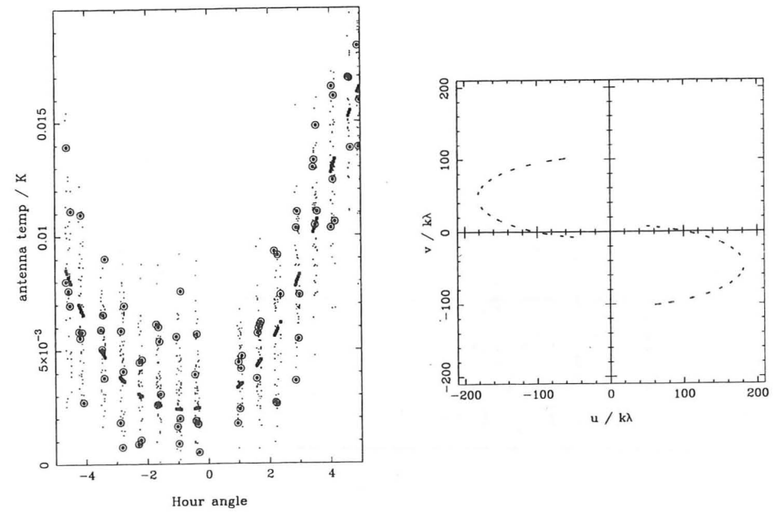


Figure 6.7: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for HL Tau, taken on Night 6.

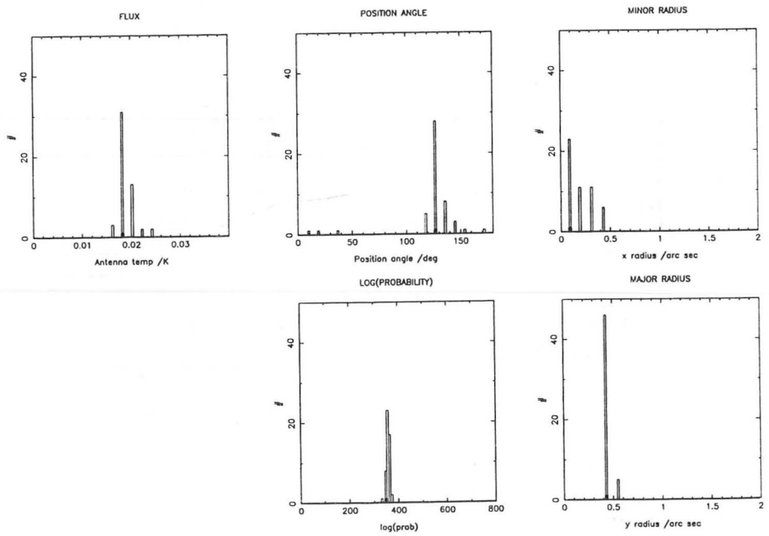


Figure 6.8: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for HL Tau on Night 6.

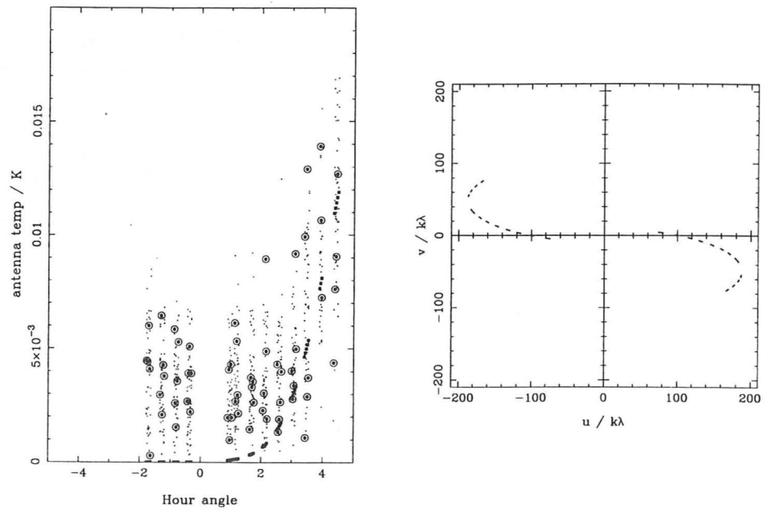


Figure 6.9: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for L1551-IRS 5, taken on Night 4.

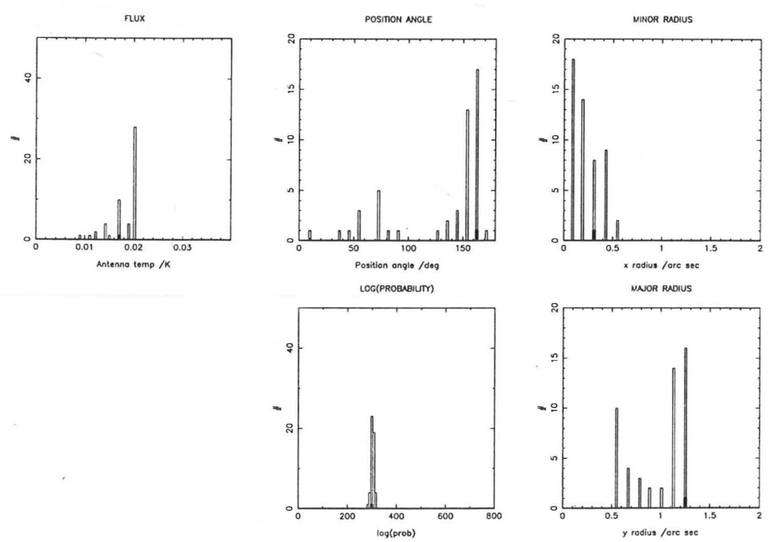


Figure 6.10: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for L1551-IRS 5 on Night 4.

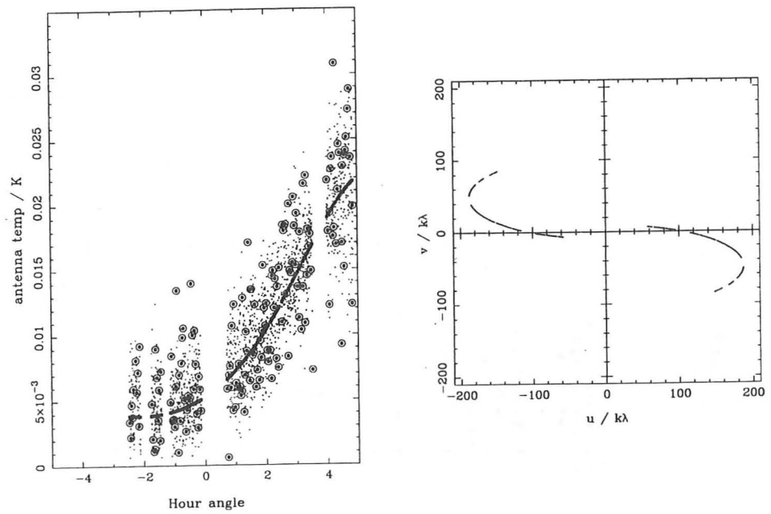


Figure 6.11: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for L1551-IRS 5, taken on Night 5.

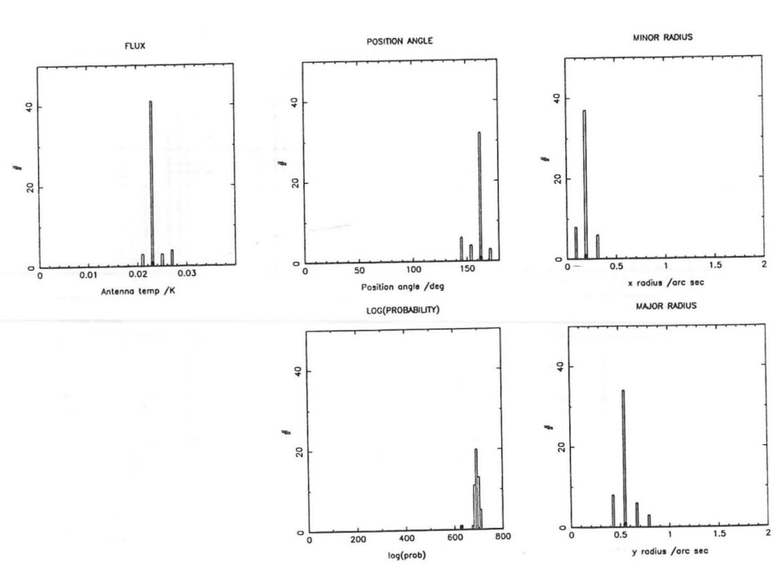


Figure 6.12: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for L1551-IRS 5 on Night 5.

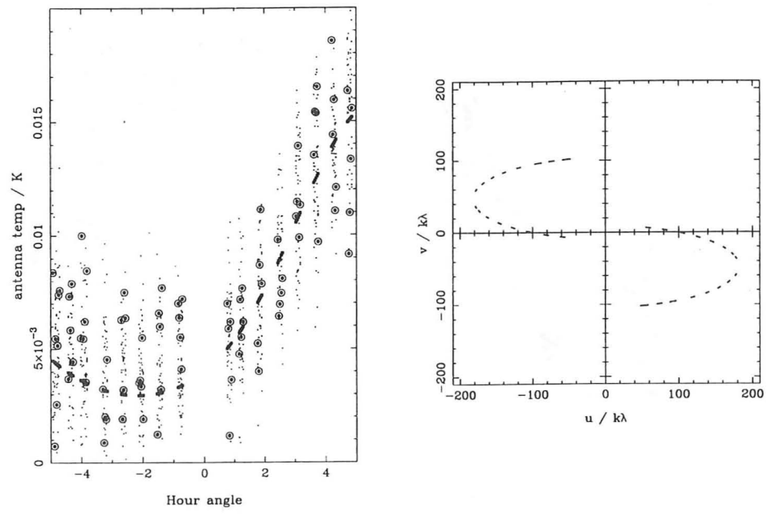


Figure 6.13: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for L1551-IRS 5, taken on Night 6.

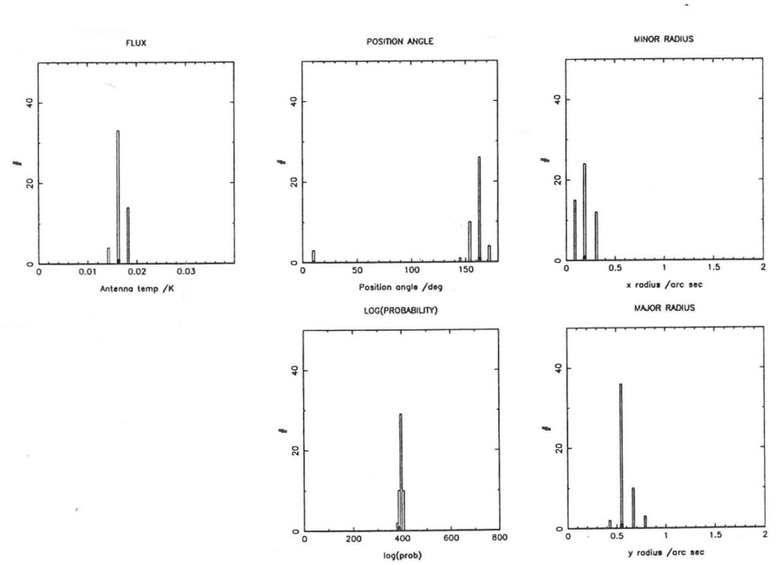


Figure 6.14: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for L1551-IRS 5 on Night 6.

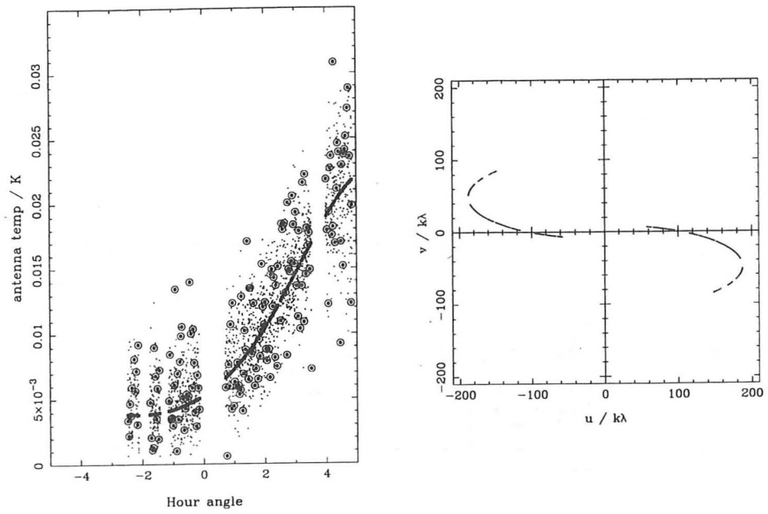


Figure 6.11: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for L1551-IRS 5, taken on Night 5.

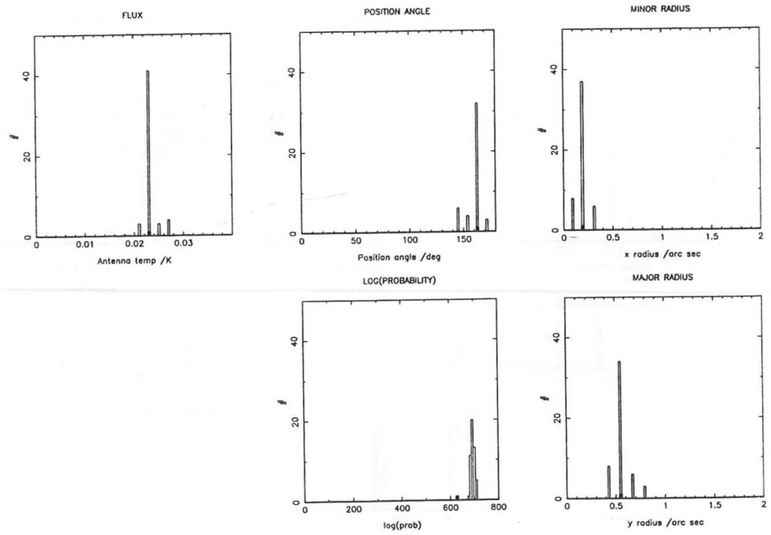


Figure 6.12: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for L1551-IRS 5 on Night 5.

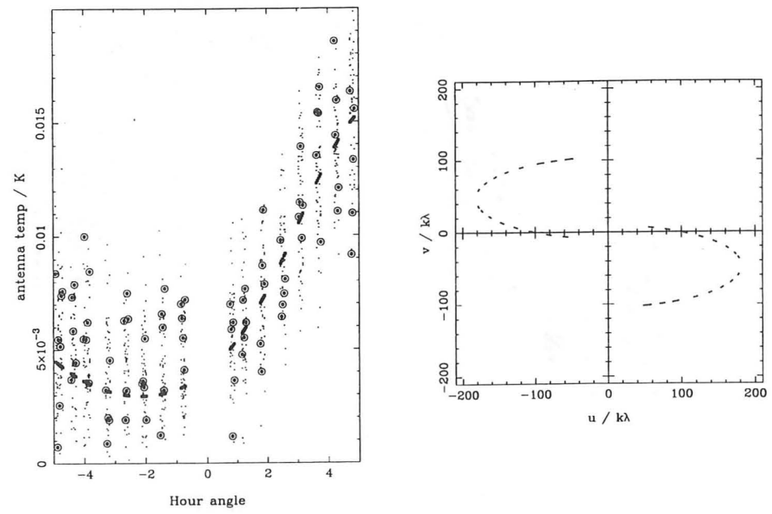


Figure 6.13: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for L1551-IRS 5, taken on Night 6.

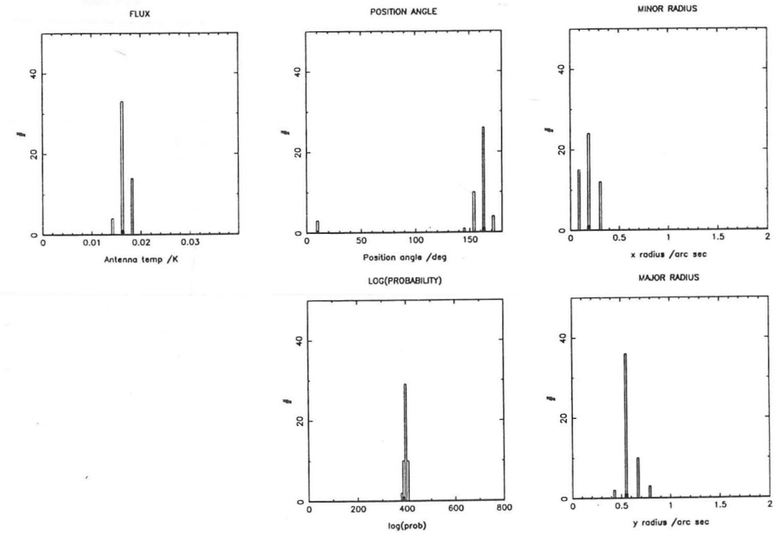


Figure 6.14: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for L1551-IRS 5 on Night 6.

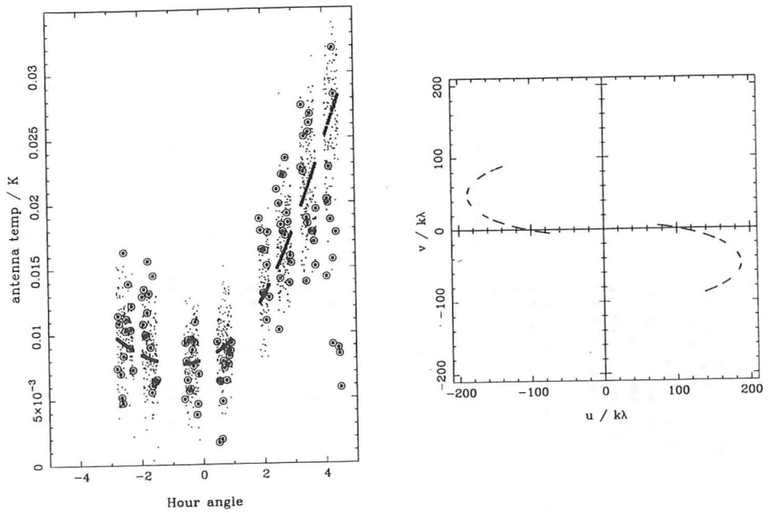


Figure 6.15: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for L1551-IRS 5, taken on Night 9.

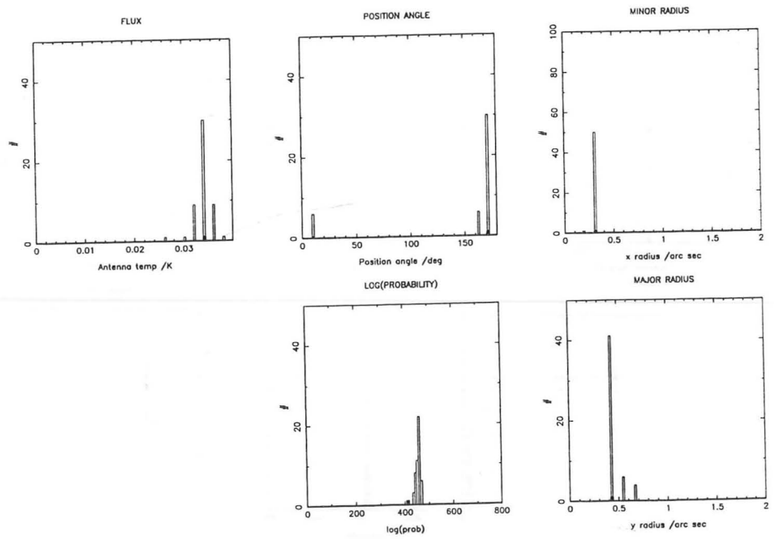


Figure 6.16: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for L1551-IRS 5 on Night 9.

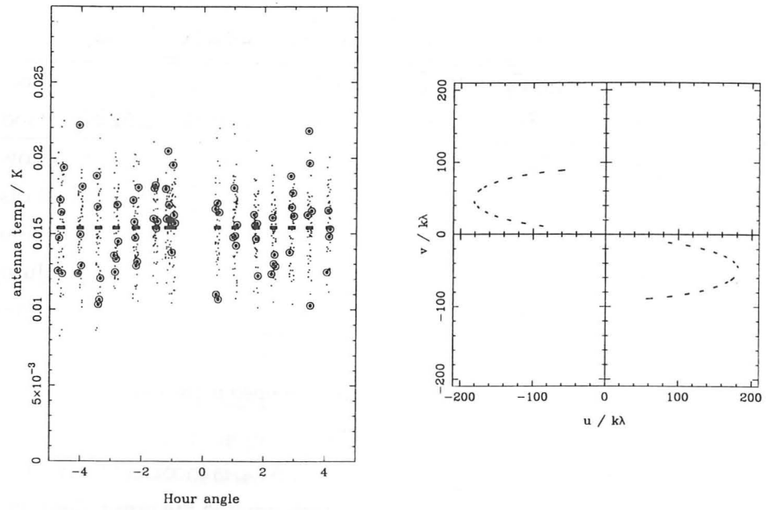


Figure 6.17: Data (circled dots), model (dashed line), simulated data (dots) and (u, v) track for 0528+134, taken on Night 6.

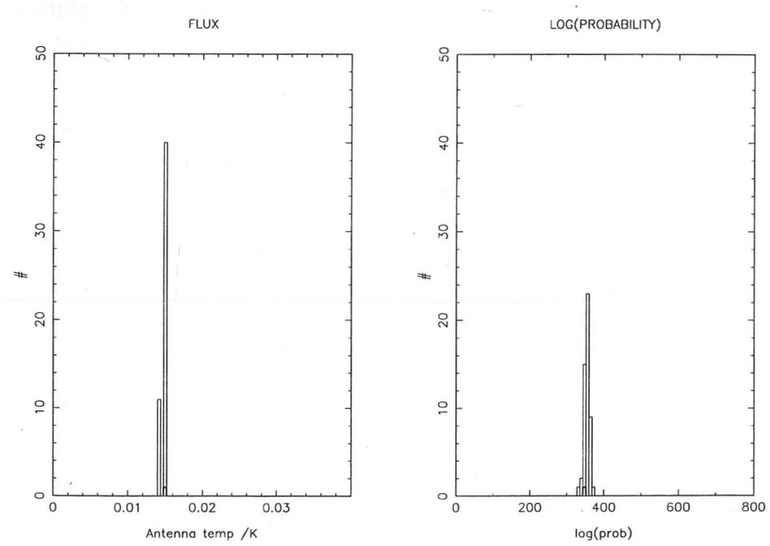


Figure 6.18: Histograms showing the results of fitting to the real dataset (highlighted) and simulated datasets for 0528+134 on Night 6.

Source	Night	Flux /mK	γ /deg.	r_{min}/D /arcsecs	r_{maj}/D /arcsecs	Flux /Jy	Comments
HL Tau	3	$11.6\,^{17}_{8.5}$	$126\,^{?}_{?}$	< 0.5	> 0.42	?	Low S/N
HL Tau	6	$18\,^{23}_{17}$	$126\,^{157}_{112}$	< 0.37	$0.43\,^{0.49}_{0.37}$	$2.52\,^{3.22}_{2.38}$	Good data
L1551-IRS 5	4	$17\,^{?}_{13}$	$162\,^{167}_{123}$	< 0.62	$1.25\,^{?}_{0}$	?	Low S/N
L1551-IRS 5	5	$23\,^{26.5}_{21.5}$	$162\,^{180}_{144}$	< 0.33	$0.57\,^{0.75}_{0.43}$	$2.53\,^{2.92}_{2.37}$	Lots of data
L1551-IRS 5	6	$16\,^{18.5}_{14}$	$162\,^{180}_{150}$	< 0.37	$0.57\,^{0.71}_{0.50}$	$2.24\,^{2.59}_{1.96}$	Good data
L1551-IRS 5	9	$34\,^{37}_{31}$	$171\,^{190}_{158}$	$0.32\,^{0.37}_{0.26}$	$0.43\,^{0.60}_{0.40}$	$2.72\,^{2.96}_{2.48}$	Jump?
0528+134	6	$15.4\,^{15.6}_{14.4}$				2.2	Point source

Table 6.2: Summary of the fits obtained to the data

data points. If our estimate of the noise in the data, N_{sys}, is incorrect then it will show up in this distribution. For example, if we have underestimated the noise, then the simulated datasets will have less scatter than the real one, and the value of $\log \mathcal{P}$ for the real dataset will be significantly less than that for the simulated ones. If the noise estimate was good, however, and the model is a good fit to the data, then the actual $\log \mathcal{P}$ should be within the scatter of the values derived for the simulated sets. This is the case for data from Nights 3, 4 and 6. For Night 5, the best-fit model appears to trace the data well, but the noise seems to have been underestimated. For Night 9 the scatter in the real and simulated datasets is about the same, but the model does not trace the actual data well; there appears to be a jump in the sensitivity of the interferometer between hour angle 1.0 and 2.0. The specific cause is unknown, but possibilities are an incorrect focus setting or bad pointing that was subsequently corrected.

Table 6.2 summarizes the results of the fitting for the two sources. In the first instance, the ranges searched were quite large. They were then narrowed somewhat. The flux range was chosen to suit the data, the position angle was always taken in steps of 9°, and the angular radii in the range 0.08 to 1.25 arcseconds. The interferometer cannot distinguish angular sizes any less than about 0.3 arcseconds, so we usually only have an upper limit for the minor radius. The confidence intervals quoted in the table span 90% of the corresponding distribution in the histogram, i.e. they are nominally 90% confidence limits. This is a conservative estimate since I have made no attempt to account for correlations between the four parameters.

For nights where the sensitivity was known (see Table 6.1), I have included the relevant fluxes in Janskys. The flux for L1551-IRS 5 on Nights 5 and 9 appears to be significantly higher than that for Night 6. This is not unexpected, since the latter observations were

Source	Flux /Jy	γ /deg	r_{maj} /AU	r_{min} /AU
HL Tau	$2.52\,^{2.94}_{2.38}$	$126\,^{144}_{117}$	$60\,^{70}_{52}$	< 52
L1551-IRS 5	$2.24\,^{2.52}_{2.10}$	$162\,^{180}_{153}$	$80\,^{100}_{70}$	< 52

Table 6.3: A summary of the data from Night 6.

made at 345 GHz, compared to 354 GHz for the first two. The difference is consistent with the increase in flux with frequency normally associated with dust emission but, with the relatively large errors and the proximity of the two frequencies, the actual value of the spectral index is not well-defined by these observations.

Night 6 clearly gives the best data for both sources; Nights 3 and 4 have poor signal-to-noise and there are problems with both Nights 5 and 9. In the analysis that follows I will therefore use results derived from the fits to the Night 6 data. These are summarized in Table 6.3. Apart from the underestimate of the noise, Night 5 is also a well-constrained fit and is in good agreement with Night 6 for L1551-IRS 5.

One should be wary of the effects of systematic errors on the fit. For example, if there was a period of particulary unstable weather with large amplitude phase variation during a night, then the visibilities would be reduced for that period of time. On Night 6, observations of L1551-IRS 5, HL Tau and 0528+134 were interleaved, cycling through the three sources every 30 minutes or so. If there is no change in the sensitivity, then the flux of the quasar will remain constant throughout the night. Figure 6.17 shows the relevant data for 0528+134; a fit has been made for a point source to determine the flux, and simulated data have been added. The best-fit value of $\log \mathcal{P}$ for the real data lies well within the simulated distribution, implying that there is no significant excursion from constant flux behaviour, and that the sensitivity did not change significantly during the night

To test how robust the fitting procedure is in the presence of unknown fluctuations in the sensitivity, I simulated fluctuations at the level of 10%. The parameters subsequently obtained for the L1551-IRS 5 Night 6 data were still within the ranges given in Table 6.2. I conclude that, even though there was no evidence for variation in the flux sensitivity, small changes would not greatly affect the fitting process.

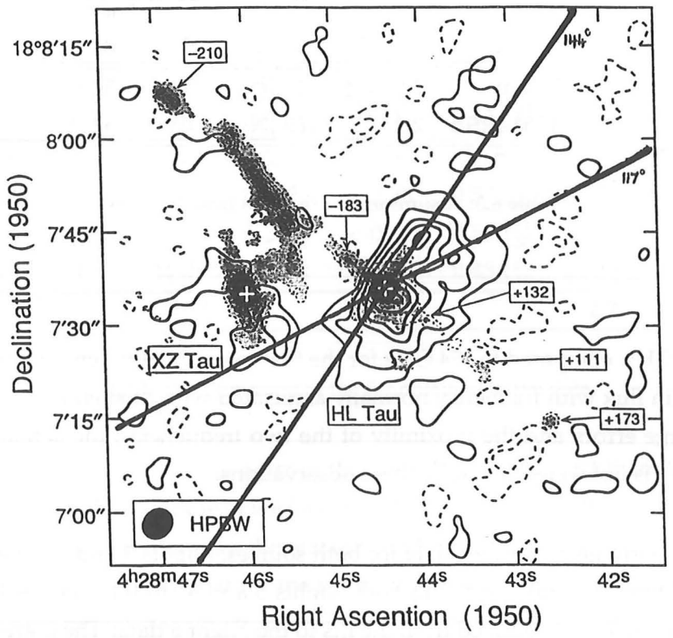


Figure 6.19: Map of HL Tau taken from Hayashi, Ohashi and Miyama (1993). The grey scale indicates the shocked [SII] emission from the bipolar outflow, and the contours show the intensity of the ^{13}CO (1–0) transition. The thick black lines have been added to show the 90% confidence range for the direction of elongation, and are not intended to represent the size of the disc. Note that the axes are labelled in seconds of time (1 second of time ≡ 15 arcseconds).

6.4 Further analysis and discussion

6.4.1 How do the disc orientations compare to the bipolar outflows?

If the emission that we detect is from accretion discs, then we would expect the elongation to be perpendicular to the outflow direction. This is easy to check.

I will first consider HL Tau; Figure 6.19 shows that the outflow direction appears to be normal to the plane of the disc inferred from our data. It now seems likely that the ^{13}CO traces the pseudo-disc predicted by Galli and Shu (1993), rather than the giant, rotating, molecular disc in the original interpretation by Sargent and Beckwith (1987).

In Figure 6.20 we can see that the elongation direction for L1551-IRS 5 is not perpendicular to the molecular outflow on large scales. On smaller scales, however, the direction of the outflow rotates to become almost east-west when seen in free-free emission by the

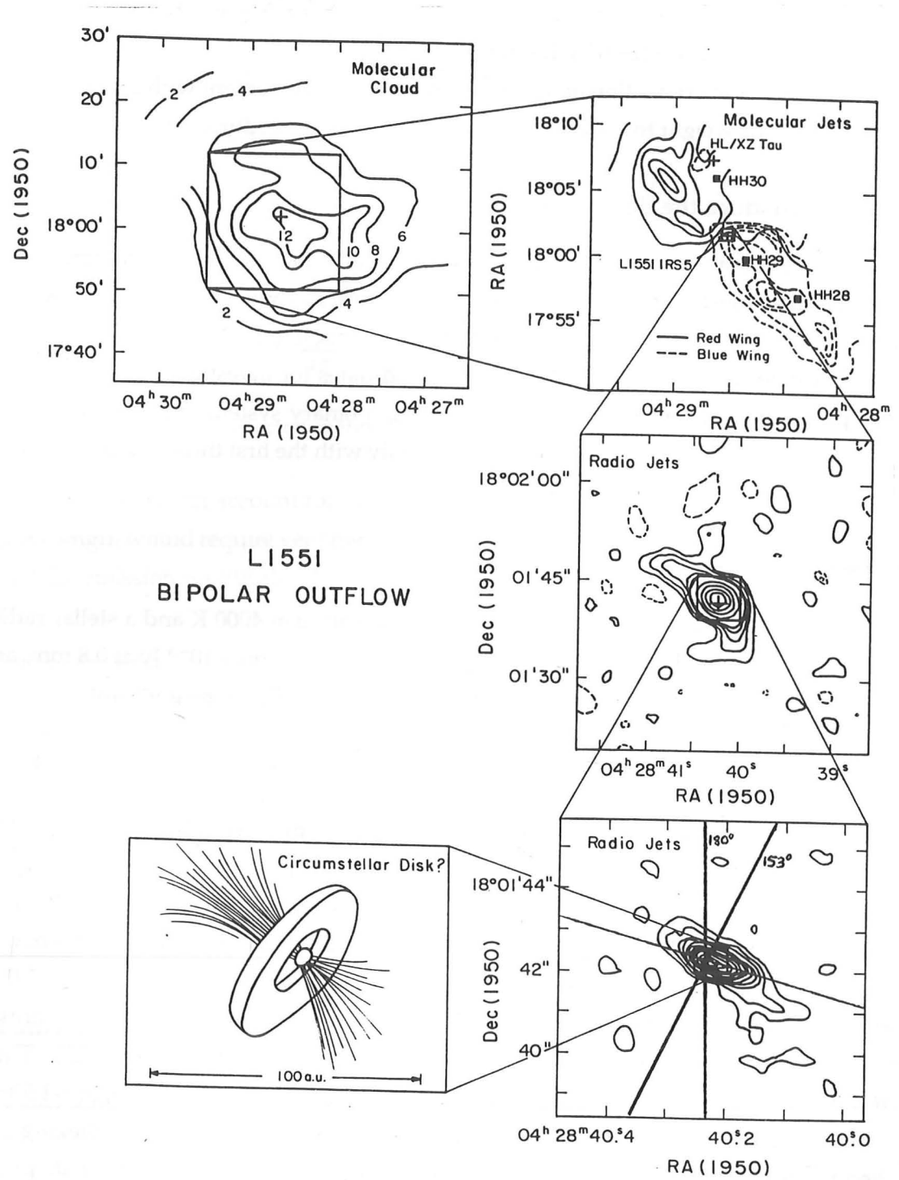


Figure 6.20: A series of maps of increasing resolution, from Snell *et al.* (1985). I have indicated the range of orientations for a disc that we can infer from our data, as well as the axis of the outflow very close to the central source.

VLA. This is consistent with the orientation of our disc. It is not clear whether the direction of the outflow has changed as a result of precession, or from interaction with density variations in the envelope. Our orientation is also similar to that reported for a much larger molecular disc seen in CS (1–0) by Kaifu *et al.* (1984).

In both cases the orientation of the emission is in good agreement with a disc interpretation. This lends weight to the other results of the fitting procedure.

6.4.2 Where does the flux come from?

If the interpretation of a circumstellar disc as the source of the emission is correct, then we expect that most, if not all, of the submillimetre flux comes from warm dust. We should estimate, however, the extent to which other emission mechanisms can contribute to the flux detected with the interferometer. Candidates for investigation are the stellar photosphere, free-free emission from ionised jets, a binary system and the envelope of material still falling into the centre. I will deal briefly with the first three before examining the envelope contribution in some detail.

The stellar photosphere

Keene and Masson (1990) suggest an effective temperature of 4000 K and a stellar radius of 0.028 AU for L1551-IRS 5. This is enough to muster a flux of only 10^{-5} Jy at 0.8 mm, and is clearly negligible. The contribution for HL Tau will be similarly insignificant.

Free-free emission

Figure 6.21 is a plot taken from Rodríguez (1994) showing the spectral energy distribution of HL Tau. The separate contributions from the free-free and dust components are quite clear. Extrapolation of the free-free emission to $\log \nu = 11.5$ (i.e. 350 GHz) suggests that it is weaker than the dust emission by almost two orders of magnitude at this frequency.

Free-free emission is seen from L1551-IRS 5; first Bieging and Cohen (1984) and then Rodríguez *et al.* (1986) used the VLA to show that there is emission from the ionised jet, and also from two unresolved point sources close to the centre (Figure 6.22). These two sources are separated by approximately 0.3 arcseconds and have fluxes of ~1.5 mJy at 2 cm. Two theories have been suggested to explain these point sources; Bieging and Cohen propose a binary protostellar system (see below) whereas Rodríguez *et al.* prefer it to be the inner, ionised boundary of a dusty disc seen almost edge-on. Neither can be ruled out by the observations presented here. It is unlikely, however, that these sources of free-free emission make a significant contribution to the submillimetre continuum flux; in the extreme case of both unresolved sources being optically thick all the way from 2 cm

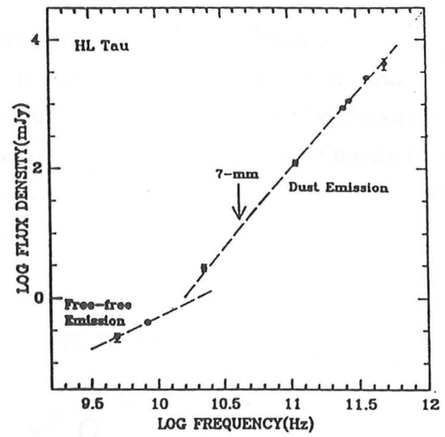


Figure 6.21: Spectral energy distribution of HL Tau from Rodríguez (1994).

to 0.8 mm, they can account for at best 1 Jy of flux. To be optically thick at such a short wavelength would require very high densities and temperatures and it is much more likely that the emission is optically thin and negligible at 0.8 mm.

Binary systems

It is also worth checking that the sources are not close binaries; such a system would have a non-spherical distribution of material that might resemble a disc, at least as far as our fitting procedure is concerned. Ghez *et al.* (1993) observed a number of T Tauri stars using speckle imaging at 2 μm and discovered that at least 60% are binaries. HL Tau was observed, but there was no sign of a companion within the sensitivity of their experiment. For a pair of equally bright sources they were able to detect binaries down to a separation of about 0.045 arcseconds.

L1551-IRS 5 is more embedded than HL Tau, in an envelope that is optically thick to the 2 μm radiation used in the speckle experiment. Bieging and Cohen (1984) suggested that the point sources seen in Figure 6.22 are the two components of a binary system. There is also some evidence for precession of the jet direction in this source (Figure 6.20), which could be explained if the central source is a binary. This could also give rise to the elongation of the continuum emission that we see, but our data do not enable us to distinguish between a disc and a close binary system. We can say, however, that a north-south binary system in L1551-IRS 5 must have an angular separation of less than 1 arcsecond (since we have a north-south baseline component of up to ~100 kλ), otherwise the visibility curve would show more than one minimum (for examples of this, see Figure 6.24).

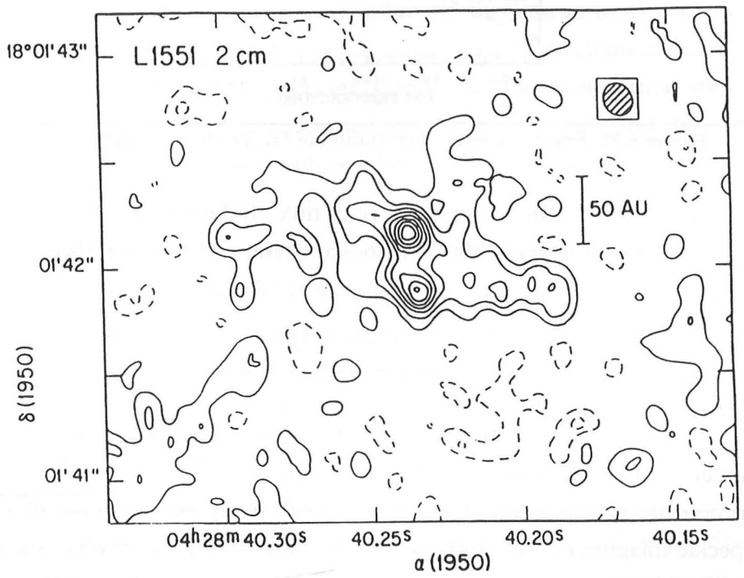


Figure 6.22: High-resolution VLA map of L1551-IRS 5, showing the free-free emission associated with the outflow and two point sources. Note how the jet appears to be aligned approximately east-west on this small scale. Taken from Rodríguez *et al.* (1986).

6.4.3 The envelope contribution

As pointed out by Terebey *et al.* (1993), it is important to account for the contribution of the in-falling envelope of the protostar. In this section I show that the envelope flux seen by the interferometer is relatively unimportant for both sources.

For HL Tau, the interferometer has detected all of the 2.6 Jy seen with the 16 arcsecond single-dish beam in the observations of Adams, Emerson and Fuller (1990), suggesting that there is very little continuum emission from outside the disc. For L1551-IRS 5, however, the 2.2 Jy emission associated with the disc at 0.87 mm accounts for only ~25% of the single-dish flux predicted by Keene and Masson (1990) for this wavelength.

Here, my aim is to calculate the visibility curve for the envelope of L1551-IRS 5, in order to estimate the relative importance of the envelope as a source of flux in our measurements.

Modelling the envelope

I will adopt the model of Adams, Lada and Shu (1987) for the density and temperature of a spherically-symmetric, isothermal cloud undergoing collapse:

$$\rho(r) = \rho_o \left(\frac{r}{r_o}\right)^{-1.5}, \tag{6.11}$$

$$T(r) = T_o \left(\frac{r}{r_o}\right)^{-(0.5+\delta)}, \tag{6.12}$$

where ρ_o and T_o are the respective (undetermined) values at a radius of r_o. The nominal value of δ is zero, but I will also investigate the case $\delta = 0.5$. Assuming that the envelope is optically thin at submillimetre wavelengths, the flux seen by an observer at distance D is given by an integral over the volume of the source:

$$S_{\rm env} = \frac{1}{D^2} \int \rho(r) \kappa_\nu B_\nu dV \tag{6.13}$$

$$= \frac{1}{D^2} \int \rho(r) \kappa_o \left(\frac{\lambda_o}{\lambda}\right)^{1.5} \left(\frac{2kT}{\lambda^2}\right) dV \tag{6.14}$$

$$= \frac{A}{\lambda^{3.5}} \int \frac{1}{r^{(2+\delta)}} dV, \tag{6.15}$$

where

$$A = 2kT_o \rho_o \kappa_o \lambda_o^{1.5} r_o^2 D^{-2}. \tag{6.16}$$

I have taken the specific opacity κ_ν to have a power law dependence of $\lambda^{-1.5}$ (Terebey *et al.*, 1993), and approximated the black-body brightness function B_ν with the Rayleigh–Jeans limit. If this is not valid, we will overestimate the flux.

The emission from the envelope is strongly peaked towards the centre and it is the material close to the central source that gives rise to any flux detected with the interferometer.

The rest will be extended with respect to our response fringe and will be resolved out. It is therefore sufficient, when calculating the visibility curve for the envelope, to consider only the emission from a central sphere. A sphere that subtends an angular size comparable to the beam of the JCMT will easily contain all of the flux detected by the interferometer. I will therefore take a sphere of radius 1200 AU, which has an angular size of 17 arcseconds when viewed at a distance of 140 pc, and determine the visibility curve for this sphere. A good starting point is to find its total emitted flux.

The total flux from a sphere of envelope radius R

When the volume of integration in equation 6.15 is a sphere of radius R, the corresponding flux is given by

$$S_{\mathrm{env},<\mathrm{R}} = \frac{A}{\lambda^{3.5}} \int_0^R \frac{1}{r^{(2+\delta)}} 2\pi r dr \tag{6.17}$$

$$= 4\pi A \lambda^{-3.5} \frac{R^{(1-\delta)}}{(1-\delta)}. \tag{6.18}$$

For $\delta = 0$, the flux simply scales with the radius. The unknown quantity is the constant A, and we can find this from fluxes measured using a single dish.

The flux measured by a single dish

The beam of a single dish forms a cylinder with a Gaussian profile that passes through the envelope. The envelope flux seen with a beam of full-width-half-maximum angular size θ_b, and at wavelength $\lambda_{\mathrm{single}}$, is

$$S_{\mathrm{single}} = \frac{A}{\lambda_{\mathrm{single}}^{3.5}} \int \frac{1}{r^{2+\delta}} \exp\left[-4\ln 2 \left(\frac{\theta}{\theta_\mathrm{b}}\right)^2\right] dV, \tag{6.19}$$

where θ is the angle subtended from the volume element dV to the line of sight. For the case of $\delta = 0$ this integral can be evaluated analytically to give

$$S_{\mathrm{single}} = \frac{\pi^{1.5}}{2\sqrt{\ln 2}} A \lambda_{\mathrm{single}}^{-3.5} \theta_\mathrm{b} D. \tag{6.20}$$

Now we can express the flux from the sphere of radius R in terms of the observed single dish flux, effectively eliminating the unknown constant A:

$$S_{\mathrm{env},<\mathrm{R}} = \left(\frac{8\sqrt{\ln 2}}{\pi^{1.5}}\right) \left(\frac{R}{\theta_\mathrm{b} D}\right) \left(\frac{\lambda_{\mathrm{single}}}{\lambda}\right)^{3.5} S_{\mathrm{single}}. \tag{6.21}$$

For non-zero values of δ, the integral of equation 6.19 must be evaluated numerically (I did this using the MAPLE package). For $\delta = 0.5$ we obtain

$$S_{\mathrm{env},<\mathrm{R}} = 1.81 \left(\frac{R}{\theta_\mathrm{b} D}\right)^{0.5} \left(\frac{\lambda_{\mathrm{single}}}{\lambda}\right)^{3.5} S_{\mathrm{single}}. \tag{6.22}$$

Real observations can now be used to quantify the flux from spheres of different radii. The values of Keene and Masson (1990) give a flux of 1.4 Jy for the envelope, once the estimated disc flux (1 Jy) has been subtracted. This was measured using the 27 arcsecond beam of the CSO at $\lambda_{\mathrm{single}} = 1.25$ mm. Inserting these numbers into equation 6.21 gives a flux of 1.90 Jy for $R = 1200$ AU and 0.08 Jy for $R = 50$ AU, both appropriate for $\delta = 0$. Corresponding values for $\delta = 0.5$, using equation 6.22, are 3.4 Jy and 0.7 Jy, respectively. The reason for calculating the flux within 50 AU of the centre becomes apparent in the next section.

The visibility curve for a sphere radius R

To calculate the visibility curve for a 3-dimensional sphere of the envelope, we must first project all of the emission onto one dimension, in a direction that lies in the plane of the sky. This gives the 1-dimensional brightness distribution, $B_1(\psi)$, where ψ is the angular offset along the chosen direction. Straightforward integration with $\delta = 0$ yields

$$B_1(\psi) \propto \ln\left(\frac{1200AU}{\psi D}\right), \tag{6.23}$$

and for $\delta = 0.5$

$$B_1(\psi) \propto \left(\frac{1200AU}{\psi D}\right)^{0.5} - 1. \tag{6.24}$$

These curves are shown in Figure 6.23a. The visibility curve is the Fourier transform of $B_1(\psi)$; I evaluated these numerically with a Fast Fourier Transform algorithm and the resulting curves are depicted in Figure 6.23b, normalized to a peak flux of unity. This peak flux is simply the total flux from the 1200 AU sphere, a value already calculated in the previous section. At the shortest projected baseline ($\sim$50 kλ) that was used in the observations of L1551-IRS 5, the interferometer will see only 0.2 Jy (i.e. 0.1×1.9 Jy) for $\delta = 0$, or 1.0 Jy (i.e. 0.3×3.4 Jy) if $\delta = 0.5$.

The actual value is likely to be less than this, since the envelope cannot extend all the way down to the star. The angular momentum of the in-falling material means that at small radii it will fall onto the disc, rather than the star. A conservative estimate for the inner cut-off of the envelope is 50 AU and we should neglect any flux that comes from within this region. Using the figures that we have already calculated, and the fact that a sphere of 50 AU radius is not significantly resolved for a baseline of 50 kλ, the maximum contribution of the envelope to our data is reduced to 0.12 Jy (i.e. 0.2 Jy – 0.08 Jy) for $\delta = 0$, or 0.3 Jy (i.e. 1.0 Jy – 0.7 Jy) for $\delta = 0.5$.

Therefore, the nominal envelope contribution to our data for L1551-IRS 5 is $\sim$0.1 Jy, less than 5% of the estimated disc flux. To ensure that the fitting procedure for the disc model was not disrupted by this extra flux, I subtracted the predicted envelope visibilities

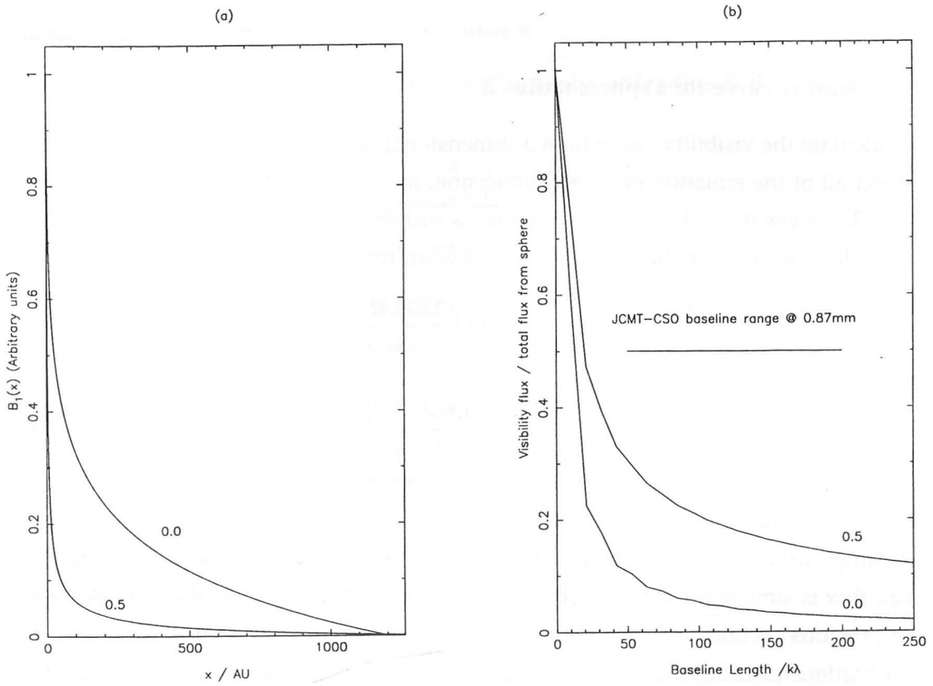


Figure 6.23: a) 1-dimensional brightness distribution B_1 plotted as a function of radius ($= \psi D$), for the two cases $\delta = 0$ and $\delta = 0.5$. b) The Fourier transform of B_1 as a function of baseline length. For a baseline of 50 kλ, the interferometer will see about 10% of the total flux if $\delta = 0$ or 30% if $\delta = 0.5$.

		Flux @ λ_1	Flux @ λ_2	Spectral index
HL Tau	Disc	2.52 Jy[a] @ 0.87 mm	0.112 Jy[b] @ 2.7 mm	2.7
	Envelope	∼0 Jy[c] @ 0.87 mm	—	—
L1551-IRS 5	Disc	2.24 Jy[a] @ 0.87 mm	0.13 Jy[d] @ 2.73 mm	2.5
	Envelope	7.8 Jy[e] @ 0.87 mm	0.16 Jy[d] @ 2.73 mm	3.4

[a] JCMT–CSO Interferometer.
[b] Sargent and Beckwith (1991).
[c] Using the 2.6 Jy flux of Adams, Emerson and Fuller (1990).
[d] Keene and Masson (1990).
[e] Using estimate of 10 Jy for disc and envelope from Keene and Masson (1990).

Table 6.4: The spectral indices for the disc and envelope of HL Tau and L1551-IRS 5.

from our measured data before the fitting process; as expected there was a small decrease in the flux attributed to the disc, but no change in the position angle and sizes that were determined. I conclude, therefore, that the envelope is not an important source of flux in our data, for either HL Tau or L1551-IRS 5.

6.4.4 Spectral indices for the disc and envelope emission

The spectral index of dust emission reflects both the nature of the dust grains and the optical depth. For optically-thin emission, the flux $S_\nu \propto \nu^{2+\beta}$, where β specifies the power-law dependence of the specific opacity, $\kappa_\nu \propto \nu^\beta$. Early results (Draine and Lee, 1984) suggested that $\beta = 2$, but later observations have found $1 \leq \beta \leq 2$ (Terebey *et al.*, 1993). Recent evidence suggests that β can be as low as zero for sufficiently large (∼10 cm) or 'fluffy' grains (Pollack *et al.*, 1994). If, on the other hand, the dust is optically thick, then it emits as a black body and $S_\nu \propto \nu^2$.

Table 6.4 lists the fluxes that we can infer for the disc and envelope of each source at two frequencies, along with the corresponding spectral indices. The relatively low indices for the discs suggest either that they are becoming optically thick, or that the grains are large and $\beta < 1$. The envelope of L1551-IRS 5 has a higher spectral index than the disc, in line with the findings of Keene and Masson and the assumption that the envelope is optically thin.

6.4.5 Disc brightness temperatures

It is possible to distinguish optically-thick dust emission from that of large grains by finding the brightness temperature of the disc, T_b. If the dust is optically thick everywhere, then

T_b should be the same as the physical temperature of the dust, T_{dust}; if, however, the emission is coming from a relatively dilute distribution of large grains then we would expect $T_b < T_{dust}$. The average physical temperature of the disc, and therefore T_{dust}, will be ~100 K (Butner *et al.*, 1994).

With a flux from a source of known size it is straightforward to determine the effective brightness temperature, T_b. If the source has flux S_ν and subtends solid angle Ω, then in the Rayleigh–Jeans limit, $T_b = \lambda^2 S_\nu / 2k\Omega$. It can easily be shown that an ellipse whose brightness falls off as a Gaussian subtends an effective solid angle of $[\pi(\ln 2) r_{min} r_{maj}]$, where r_{min} and r_{maj} are the radii to the contour of half-maximium brightness. Unfortunately, we only have an upper limit for the r_{min} of each source and can only set a lower limit on T_b. Using the nominal values for S_T and r_{maj}, and the upper limit for r_{min} (Table 6.3), we obtain lower limits on T_b of 36 K for HL Tau and 28 K for L1551-IRS 5. These are an appreciable fraction of the predicted physical temperature, and we can conclude that the average filling factor cannot be less than about 20%. It therefore seems likely that at least part of each disc is optically thick.

6.4.6 An accretion disc model

In an attempt to gain further insight into the structure and mass of the discs, I compared our results against the predictions of the 'standard' model for an accretion disc.

The standard model

The accretion disc of a young star is assumed to have circular symmetry with a thickness that is small compared to the radial extent. The temperature can be considered to be uniform with depth and, together with the surface density, is taken to have a power-law variation with radius in the disc. Adopting the notation used by Beckwith *et al.* (1990):

$$\Sigma_{disc} = \Sigma_1 \left(\frac{r_{disc}}{AU} \right)^{-p}, \tag{6.25}$$

$$T_{disc} = T_1 \left(\frac{r_{disc}}{AU} \right)^{-q}. \tag{6.26}$$

Σ_1 is the surface density at a distance of 1 AU from the centre and I shall assume that $1.0 \leq p \leq 1.5$. The index q is predicted from theory to be 0.75 for both actively accreting discs (Lynden-Bell and Pringle, 1974) and passive discs that reprocess the radiation from the central source (Adams and Shu, 1986). Beckwith *et al.* (1990), however, found that many T Tauri stars have $q \simeq 0.5$. In addition they determined that T_1, the disc temperature at 1 AU, is ~300 K. I shall take $T_1 = 500$ K; this more generous value results in a slight underestimate of the disc mass, but helps to ensure that we obtain a good lower limit to the mass.

The total flux from a disc, seen at a distance D, is given by

$$S_{disc} = \frac{1}{D^2} \int_{disc} \kappa_\nu B_\nu(T_{disc}) f(\tau) dM, \tag{6.27}$$

where $B_\nu(T_{disc})$ is the Planck black-body function and κ_ν is the specific opacity law, for which I adopt Hildebrand's (1983) formulation:

$$\kappa_\nu = 0.1 \left(\frac{250\,\mu m}{\lambda} \right)^\beta \text{ cm}^2\,\text{g}^{-1}. \tag{6.28}$$

Observed spectral indices and theoretical considerations suggest $0.5 \leq \beta \leq 2.0$. The function $f(\tau)$ is included in equation 6.27 to account for the effects of optical depth; for high surface densities the disc is no longer optically thin and some of the emission is re-adsorbed. For a simple parallel slab geometry,

$$f(\tau) = \frac{(1 - e^{-\tau})}{\tau}, \tag{6.29}$$

$$\tau = \frac{\kappa_\nu \Sigma_{disc}}{\cos\theta}, \tag{6.30}$$

where θ is the inclination of the disc axis to the line of sight. Equation 6.27 can now be rewritten as

$$S_{disc} = \frac{1}{D^2} \int_{r_{in}}^{r_{out}} \kappa_\nu B_\nu(T_{disc}) \left(\frac{1 - e^{-\tau}}{\tau} \right) \Sigma_{disc}\, 2\pi r dr \tag{6.31}$$

$$= \frac{1}{D^2} \int_{r_{in}}^{r_{out}} B_\nu(T_{disc})(1 - e^{-\tau}) \cos\theta\, 2\pi r dr. \tag{6.32}$$

Choosing solutions

For a given combination of the parameters p, q and β, within the ranges specified above, we are free to choose the values of Σ_1, θ, r_{in} and r_{out} such that the model gives the correct total flux and spectral index, and reproduces the general shape of the visibility curves. I used the previously determined position angles of 162° for L1551-IRS 5 and 126° for HL Tau.

I employed the MAPLE mathematics package to evaluate the flux for each set of parameters using numerical integration. To find the visibility curve for a given model, I wrote a FORTRAN program that calculated the visibility fluxes that would be detected with the JCMT–CSO Interferometer at ten different values of the hour angle, for a source at the same declination as L1551-IRS 5 and HL Tau (i.e. ~20°). I could then compare the shape of this curve with the actual measured data. It was not feasible, however, to include this in the same kind of automated fitting procedure that was used with the simpler Gaussian model.

For every permutation of p, q and β that I tried, I was able to find suitable values of Σ_1, θ, r_{in} and r_{out} that adequately reproduced our data. In all cases the choice of the inner disc radius r_{in} was not important and I set it to the estimated stellar radius of ~0.02 AU.

Source	p	q	β	Σ_1 /g cm^{-2}	θ	$r_{\rm out}$ /AU	$M_{\rm disc}$ /M$_\odot$
L1551	1.0	0.5	0.5	170	45°	200	0.02
HL Tau	1.0	0.5	0.5	260	45°	140	0.03
L1551	1.0	0.5	1.0	310	45°	190	0.04
L1551	1.0	0.75	0.5	1300	45°	190	0.18
L1551	1.5	0.5	0.5	1900	70°	200	0.04
L1551	—	0.5	—	—	53°	150	—
HL Tau	—	0.5	—	—	38°	130	—

Table 6.5: Examples of different solutions for an accretion disc model that give good agreement with the data

The disc mass is determined by Σ_1, p and $r_{\rm out}$:

$$M_{\rm disc} = \int_{r_{\rm in}}^{r_{\rm out}} \Sigma_{\rm disc}\, 2\pi r dr \tag{6.33}$$

$$\simeq \frac{2\pi\Sigma_1\,(1\,{\rm AU})^p}{2-p}\, r_{\rm out}^{2-p}. \tag{6.34}$$

A lower limit to the disc mass

The lowest disc masses were obtained using the lowest values for each of the three indices: $p = 1.0$, $q = 0.5$ and $\beta = 0.5$. Example solutions that gave good agreement with the data are shown in the first two lines of Table 6.5. Most of the emission for these solutions is optically thin: the disc is only optically thick within ~10 AU of the centre. As a result, the flux is relatively insensitive to the inclination angle, until seen nearly edge-on. The disc masses of ~0.02 M$_\odot$, which can be considered a good lower limit, are in line with the estimated minimum mass required to form the solar system. But note that this low value of β implies large grains; the Hildebrand (1983) opacity law is then likely to be in error, and we will have underestimated the disc mass.

Optically-thick solutions

The bottom two lines in Table 6.5 are examples of totally optically-thick solutions. The emission no longer depends on the density and opacity, and this treatment gives no upper limit to the mass. The inclination and the outer radius of the disc are the only two free parameters. The spectral index is 2.0. The middle three lines show solutions to the L1551-IRS 5 emission for some other combinations of p, q and β.

Summary

This analysis shows that the current data do not constrain the structure of the discs very well, although we can set a lower limit to the disc mass of ~0.02 M$_\odot$ for both sources. This compares with the value of 0.1 M$_\odot$ estimated by Beckwith *et al.* (1990) for the disc mass of HL Tau.

It is not possible to choose between optically-thick emission and a more spatially dilute emission from large grains to explain the low spectral index, although the the discs cannot be entirely optically thick (our spectral indices are too high), or any less than about 20% 'filled-in' (as the brightness temperatures are quite high). If the disc inclination can be found, either from the orientation of the outflow or, preferably, from future observations that resolve the minor radius, then tighter constraints on the disc will be possible.

6.5 Detecting binaries with the interferometer

Before concluding this chapter, it is appropriate to describe briefly observations that we made of NGC1333-IRAS 4 using the interferometer. This protostellar source was recently discovered to be binary, with a separation of 31 arcseconds (Sandell *et al.*, 1991).

6.5.1 Observations of NGC1333-IRAS 4

We observed both components (designated IRAS 4a and IRAS 4b) over one night in November 1993. Figure 6.24 shows the visibility curves for each source. It appears that there is a sharp dip in the visibility for both components, in addition to the gradual increase of flux arising from resolved circumstellar emission. The interleaved quasar observations showed no such deviation, and the simplest explanation is that each component is itself a binary.

6.5.2 A simple model

The solid curves in Figure 6.24 show models that have been fitted to the data using the method of §6.3.2 to account for the non-Gaussian measurement errors. Figure 6.25 depicts the simple binary model that was used. The components are assumed to be of equal flux, each with a circularly-symmetric dust distribution that falls off as a Gaussian. Parameters are the total flux of the binary system, the angular separation, the position angle of the separation, and the radius to half-maximum of the circumstellar emission. Where the flux falls to zero we expect a large change in the phase of the data. There was some evidence for this, but it is hard to be certain in the face of phase calibration problems.

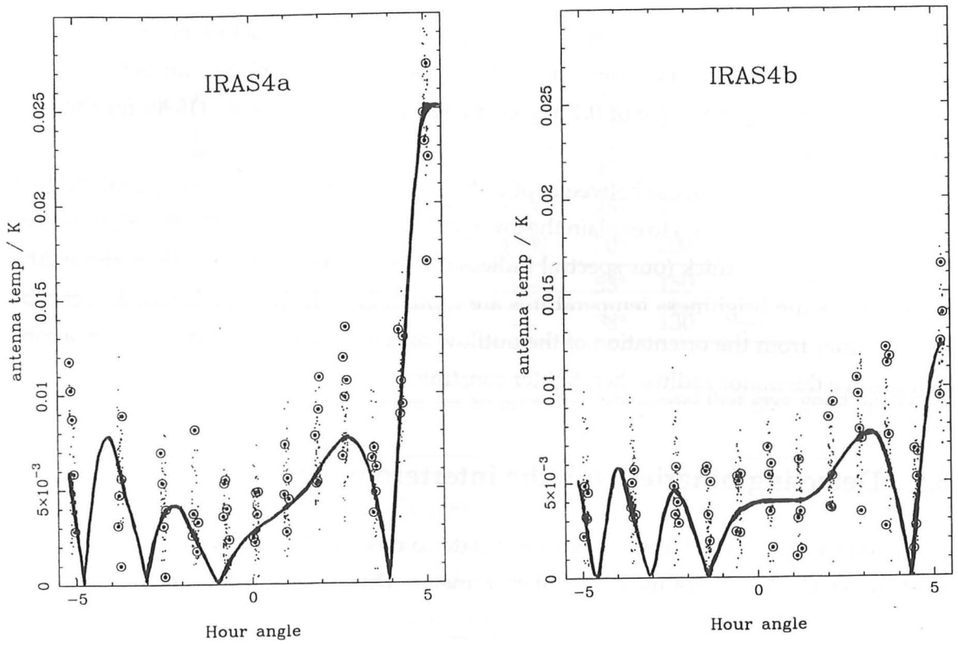


Figure 6.24: Visibility data and best-fit model curves for the binary components NGC1333-IRAS 4a and IRAS 4b. Fitted parameters for these models are: $S_T = 4.8$ Jy, $\gamma = 54°$, $r_{sep} = 3.0$ arcsecs and $r_{env} = 0.36$ arcsecs for IRAS 4a; $S_T = 2.0$ Jy, $\gamma = 54°$, $r_{sep} = 3.3$ arcsecs and $r_{env} = 0.27$ arcsecs for IRAS 4b.

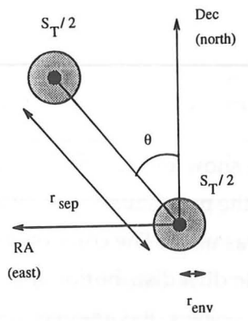


Figure 6.25: A simple model for a binary system.

Of course this a very simple model, and I describe it more in the hope of showing that the shape of the visibility curves can be explained by a binary system, than of actually obtaining any significant quantitative data.

6.6 Future work

Protostellar systems are probably the most fertile area of research for submillimetre interferometry and there is great potential for future observations. We have time scheduled for October 1994, during which we hope to extend the sample of protostars and to get good observations of molecular line emission.

6.6.1 A survey of protostars

By observing many embedded protostars we can learn how common discs are, when they develop, how long they last and how big they are. Our current methods of observation require sources with a continuum flux in excess of 0.5 Jy or so, but this will improve when we can integrate coherently and reliably for periods longer than 100 s. I estimate that we will be able to observe two or three sources per night.

From studies of T Tauri stars by, for example, Ghez *et al.* (1993), it is becoming clear that single stars may be the exception rather than the rule in the star-formation process. It is therefore very important to search for binaries and higher order multiple systems at very early stages in their formation, when they are still embedded in the parent molecular cloud. Traditional methods of finding binaries, for example optical spectroscopy, infrared speckle and lunar occultation, do not penetrate the obscuring envelope material, and submillimetre interferometry is one of the only tools available with which to do this. The model-fitting process for binaries will also be more effective once the phases are more meaningful.

If successful, such a survey will provide invaluable information and help to settle some of the unresolved issues in star formation.

6.6.2 Observing the molecular line emission

We detected CO (3–2) line emission from L1551-IRS 5 with the interferometer. The fact that this emission is seen against the disc continuum implies either that it is more optically thick than the continuum or that it has a higher filling factor in the interferometer beam. With spectral line data we have the exciting possibility of finding a rotation curve for the disc, which would establish beyond any doubt that we are indeed seeing an accretion disc. In addition to the relatively high velocities expected from the inner regions of an

accretion disc, we would expect to see slightly red-shifted absorption from the cool, infalling molecular material in the envelope.

Observations of CO (3–2) may also enable us to distinguish between optically-thick disc continuum emission and emission from a more spatially dilute collection of large grains. The CO (3–2) emission from the dense protostellar core is likely to be optically thick, assuming that it is not substantially depleted onto grain surfaces. If we can attribute it to emission from the disc, rather than, for instance, from an outflow, then the ratio of the line antenna temperature to that of the continuum sets an upper limit to the average optical depth of the dust emission. It may also be possible to measure the temperature of the disc as a function of radius to differentiate between models with $q = 0.5$ and those with $q = 0.75$.

To achieve these aims we need good quality molecular line observations, and these will take much longer to obtain than the continuum measurements. Long periods of coherent integration are required, although it should be possible to use the continuum as a phase reference. Other molecular transitions, for example HCO^+ (4–3), should also be observed.

6.7 Summary

The JCMT–CSO Interferometer has detected elongation of the continuum emission from the protostellar sources L1551-IRS 5 and HL Tau for the first time. It is likely that this emission comes from the warm, dusty accretion discs expected to exist around young stars. Table 6.6 summarizes the parameters that I was able to derive for the two sources.

We also discovered that NGC1333-IRAS 4, a very young and embedded protostellar binary system, is actually a hierarchical, quadruple source.

		L1551-IRS 5	HL Tau	See section
Disc	870 μm Flux	2.2 Jy	2.5 Jy	6.3.4
	r_{maj} to 50%	80 AU	60 AU	6.3.4
	r_{min} to 50%	< 52 AU	< 52 AU	6.3.4
	Pos. angle	162°	126°	6.3.4
	Spectral Index	2.7	2.5	6.4.4
	Perpendicular to outflow?	yes	yes	6.4.1
	Brightness temp.	> 28 K	> 36 K	6.4.5
	M_{disc}	$> 0.02\ M_\odot$	$> 0.03\ M_\odot$	6.4.6
Envelope	870 μm flux, 17″ beam	$\sim$3 Jy	$\sim$0	6.4.3
	Spectral index	3.4	—	6.4.4

Table 6.6: Summary of the parameters determined for the discs and envelopes of L1551-IRS 5 and HL Tau.

Chapter 7

SUMMARY AND CONCLUSIONS

Each chapter has ended with a summary of the work that it has covered. Here I reiterate what I consider to be the important results and conclusions of my work, and describe the current status and future prospects of the JCMT–CSO Interferometer project, emphasizing the aspects for which I have been responsible.

7.1 Observing with the JCMT–CSO Interferometer

There have been four observing runs with the interferometer to date. Most of this time has been spent debugging and calibrating the system. This involved observing bright sources, particularly quasars, that are of limited intrinsic interest. With the improved understanding of the system, the emphasis is now shifting to astronomy, and the next run, scheduled for October 1994, has been made open to the astronomical community.

Observations so far have been made in the 'A' and 'B' atmospheric windows (see Table 1.1); in October we hope to make the first interferometric observations in the 'C' band. In the longer term, we look forward to collaboration with the Harvard–Smithsonian Submillimeter Array (SMA); when the JCMT and the CSO are connected to this array, maps with sub-arcsecond resolution will be possible, with the JCMT and the CSO providing high sensitivity on the longer baselines. There is also the more tentative possibility of including the Keck telescopes in an array with the JCMT and the CSO.

7.2 Technical status

7.2.1 Hardware

Richard Hills and John Carlstrom were responsible for the basic design of the system. To make interferometry possible, this design required two major additions to the existing

hardware at the JCMT and the CSO. The phase-lock system ensures that the local oscillators at the two receivers are coherent. This was designed and built by John Carlstrom at Caltech. It appears to work well, except for the generation of the 10-MHz signal at the JCMT. I found that this was causing phase jumps in the system, and the problem will be fixed properly for the next run.

I was responsible for implementing most of the IF system. This takes the outputs of the receivers and delivers them to the correlator at the JCMT. Optical fibres, in conjunction with commercially available 10-GHz transducers, are used to transmit signals between the two telescopes. Equalization of the two different path lengths that the signals travel from the source to the correlator is a two-stage process: (1) Coarse compensation is effected using switchable lengths of optical fibre. (2) Post-processing of the correlator output is then used to achieve the fine compensation, utilizing the digital correlator's range of built-in delays. I calculated the number of delay lines that were required, designed a configuration with which to achieve this, and built a compact unit to house the fibre, optical switches and couplers that were used.

Two Direct Digital Synthesizers are used for sideband rejection, phase-switching and tracking the source. Their frequencies and phases, which must be very accurate, are set by a microcomputer that I interfaced and programmed.

The IF system is now working as intended.

7.2.2 Calibration

I have spent a lot of time improving the calibration of the instrument, particularly the modelling of the interferometer phase response as a function of time. There are many aspects to this, including a determination of: the baseline components (the two non-polar components are now known to a precision of ± 2 mm), the non-intersection of the azimuth and elevation axes (the combined offset for the two antennas is ~ 6 mm), the lengths of cabling in the system including the delay lines, and the dish deformations due to gravity. Once these have been corrected, there is a typical residual drift of $\sim 90°$ per hour, much of which is due to thermal effects. This variation can be followed by interpolating from regular observations of a phase calibrator source, making possible long, coherent integrations on faint sources. Until now we have been observing relatively bright sources that can be detected in a time within which there is little drift and decorrelation. The errors should be reduced further, but future work in this area should be aimed specifically at measuring the polar component of the baseline and reducing the phase drift for sources close to the horizon.

The system temperature of the JCMT–CSO Interferometer is the geometric mean of the individual system temperatures of the JCMT and the CSO. These are now measured

automatically by inserting hot loads into the two beams. The conversion factor from effective antenna temperature to units of flux should in theory be about 50 Jy K^{-1}, but to date this figure has been much higher in practice. Following the correction of a problem with the phase-switching, this figure has been reduced from $\simeq$140 Jy K^{-1} to $\simeq$100 Jy K^{-1}. Reducing this further should be the top priority for the run in November 1994; a value of 70 Jy K^{-1}, for example, would halve the integration time that is needed on a given source.

I have also described a new technique for measuring the passband response of an interferometer, on the assumption that the fine structure is due to the correlator. The broad shape is determined from observations of a quasar, and the fine structure from injecting correlated noise into the correlator. This in itself is not new; my idea concerns the way in which the correct passband can be obtained from these measurements. I have yet to try this on real data and the assumption about the fine structure should first be tested, but the method should prove especially useful for observations of bright spectral lines.

7.2.3 Data reduction

I have written almost all of the software that is used to process the data. The INTEREDUCE program, written in FORTRAN, is very flexible and new routines can be added easily. The raw data is currently stored for each 10 s integration, requiring large amounts of disk space. I intend to add the option of saving the data in a more advanced state of processing, probably as the vector average of ten 10 s integrations. The original raw data can still be kept on digital tape, but the smaller files will be much easier to work with. I also hope to improve the 'help' facilities, to assist new observers who are unfamiliar with the system.

7.3 Astronomy

We have observed a variety of objects using the JCMT–CSO Interferometer. Our results indicate that the continuum flux at millimetre and submillimetre wavelengths from the Galactic Centre object Sagittarius A* is unlikely to be the result of dust emission; a compact and highly variable synchrotron source is more probable. We also detected narrow, red-shifted absorption lines towards the nucleus of the nearby radio-galaxy Centaurus A, for both the CO (2–1) and CO (3–2) transitions. These must be due to molecular clouds moving in highly non-circular orbits, possibly within an inner bar structure. More analysis of these results is needed to estimate the temperatures and densities of these clouds.

For me, the most interesting results so far concern the evolved star VY CMa and the protostellar discs of L1551-IRS 5 and HL Tau.

7.3.1 VY CMa: a binary system

Observations with the JCMT–CSO Interferometer have confirmed that the 321-GHz H_2O emission is the result of maser amplification. A compact continuum source, spatially offset from the maser emission, was also detected, as was CO (3–2) emission.

I showed that these results point to a binary model for VY CMa, a theory supported by the wide range of existing data. In this model, the secondary accretes some of the material being lost by the primary and has many features commonly associated with a protostar—specifically an accretion disc and outflowing jets. I suggest that it is this additional source of luminosity that has led to the past confusion over the nature of this source.

Although the evidence for a binary system is compelling, I wish to make further observations for confirmation. These include mapping the source with the Owens Valley Millimeter Interferometer at the CO (2–1) transition to look for evidence of a disc and jets from the secondary, and possibly optical observations to search for shocked Hα emission. I also hope to observe similar sources, such as OH 231+4.2, with our submillimetre interferometer.

7.3.2 Protostars: resolved discs

The JCMT–CSO Interferometer has easily resolved the continuum emission associated with the protostars L1551-IRS 5 and HL Tau. My analysis showed that the emission is in each case elongated in the direction perpendicular to the outflow, consistent with an accretion disc interpretation. The disc radii to half-maximum brightness are 80 AU for L1551-IRS 5 and 60 AU for HL Tau, in line with expectations from theory. By modelling the structure of the accretion disc, I set a lower limit of $\sim$0.02 $M_\odot$ on the disc masses of the two sources.

I also showed that the visibility curve for each component of the young binary source NGC1333-IRAS4 is itself indicative of a binary system, so that this is a hierarchical, quadruple source, embedded in a molecular cloud.

Protostellar systems are the prime target of the JCMT–CSO Interferometer, and most of the October 1994 run will be dedicated to observing them. In addition to the continuum observations, there is the prospect of observing spectral line emission, where the challenge will be to separate out the contributions from the disc and the outflow. We also hope to make the first unambiguous detection of infall towards a protostellar source, by observing red-shifted absorption against the warm dust in the core.

The JCMT–CSO Interferometer is the pioneering instrument in the field of submillimetre interferometry. After only 24 nights of observing time, it has already generated a number of exciting results. Now that the system is more fully understood and debugged, the interferometer should be capable of realising its full potential.

REFERENCES

Adams, F. C., Emerson, J. P., and Fuller, G. A.: 1990, *Astrophysical Journal* **357**, 606

Adams, F. C., Lada, C. J., and Shu, F.: 1987, *Astrophysical Journal* **312**, 788

Adams, F. C. and Shu, F. H.: 1986, *Astrophysical Journal* **308**, 836

Baade, W. and Minkowski, R.: 1954, *Astrophysical Journal* **119**, 215

Beckwith, S. V. W., Sargent, A. I., Chini, R. S., and Güsten, R.: 1990, *Astronomical Journal* **99**, 924

Bieging, J. H. and Cohen, M.: 1984, *Astrophysical Journal* **289**, L5

Bos, A.: 1986a, *Functional design of a wideband digital spectrometer*, Technical Report 179, Netherlands Foundation for Radio Astronomy

Bos, A.: 1986b, *A General Purpose Correlator Board: Functional Description*, Technical Report 173, Netherlands Foundation for Radio Astronomy

Bos, A.: 1986c, *The N.F.R.A. Correlator Chip*, Technical Report 176, Netherlands Foundation for Radio Astronomy

Bos, A.: 1988, *The D.A.S.: System description*, Technical Report 187, Netherlands Foundation for Radio Astronomy

Bowen, G. H.: 1988, *Astrophysical Journal* **329**, 299

Bowers, P. F.: 1991, *Astrophysical Journal Supplement Series* **76**, 1099

Bowers, P. F., Claussen, M. J., and Johnston, K. J.: 1993, *Astronomical Journal* **105**, 284

Bowers, P. F., Johnston, K. J., and Spencer, J. H.: 1983, *Astrophysical Journal* **274**, 733

Burns, J. O., Feigelson, E. D., and Schreier, E. J.: 1983, *Astrophysical Journal* **273**, 128

Butner, H. M., Natta, A., and Evans, N. J.: 1994, *Astrophysical Journal* **420**, 326

Calvert, N. and Peimbert, M.: 1983, *Revista Mexicana de Astronomía y Astrofísica* **5**, 319

Carlstrom, J. E., Higgs, L., Hills, R. E., Raimond, E., and Webster, A.: 1990, *Short-baseline interferometry at the James Clerk Maxwell Telescope*, Technical report, The JCMT Board

Cernicharo, J., Bujarrabal, V., and Santarén, J. L.: 1993, *Astrophysical Journal* **407**, L33

Chapman, J. M. and Cohen, R. J.: 1986, *Monthly Notices of the R.A.S.* **220**, 513

Christiansen, W. N. and Högbom, J. A.: 1985, *Radiotelescopes*, Cambridge University Press, second edition

Cox, J. P.: 1980, *Theory of Stellar Pulsation*, Princeton University Press

Daniel, J.-Y.: 1982, *Astronomy and Astrophysics* **111**, 58

Deguchi, S., Good, J., Fan, Y., Mao, X., Wang, D., and Ukita, N.: 1983, *Astrophysical Journal* **264**, L65

Dent, W. R. F., Matthews, H. E., Wade, R., and Duncan, W. D.: 1993, *Astrophysical Journal* **410**, 650

Draine, B. T. and Lee, H. M.: 1984, *Astrophysical Journal* **285**, 89

Dutrey, A., Guilloteau, S., and Simon, M.: 1994, *Astronomy and Astrophysics* **286**, 149

Eckart, A., Cameron, M., Genzel, R., Jackson, J. M., Rothermal, H., Stutzki, J., Rydbeck, G., and Wiklind, T.: 1990, *Astrophysical Journal* **365**, 522

Elias, J.: 1978, *Astrophysical Journal* **224**, 857

Eliasson, B. and Bartlett, J. F.: 1969, *Astrophysical Journal* **155**, L79

Elitzur, M.: 1992, *Astronomical Masers*, Kluwer Academic Publishers

Feibelmann, W. A. and Bruhweiler, F. C.: 1990, *Astrophysical Journal* **354**, 262

Galli, D. and Shu, F. H.: 1993, *Astrophysical Journal* **417**, 243

Gardner, M. G.: 1979, *Phaselock techniques*, Wiley–Interscience

Ghez, A. M., Neugebauer, G., and Matthews, K.: 1993, *Astronomical Journal* **106**, 2005

Grasdalen, G. L., Strom, S. E., Strom, K. M., Capps, R. W., Thompson, D., and Castelaz, M.: 1984, *Astrophysical Journal* **321**, 473

Hamann, F. and Simon, M.: 1988, *Astrophysical Journal* **327**, 876

Harvey, P. M., Thronson, H., and Gatley, I.: 1979, *Astrophysical Journal* **231**, 115.

Hayashi, M., Ohashi, N., and Miyama, S.: 1993, *Astrophysical Journal* **418**, L71

Herbig, G. H.: 1970, *Astrophysical Journal* **162**, 557

Herbig, G. H.: 1972, *Astrophysical Journal* **172**, 375

Hildebrand, R. H.: 1983, *Quarterly Journal of the R.A.S.* **24**, 267

Icke, V. and Heske, F. A.: 1992, *Astronomy and Astrophysics* **258**, 341

Israel, F. P., van Dishoeck, E. F., Baas, F., de Graauw, T., and Phillips, T. G.: 1991, *Astronomy and Astrophysics* **245**, L13

Kaifu, N., Hasegawa, T., Morimoto, M., Inatani, J., Nagane, K., Miyazawa, K., Chikada, Y., Kanzawa, T., Akabane, K., and Suzuki, S.: 1984, *Astronomy and Astrophysics* **134**, 7

Kawabe, R., Ishiguro, M., and Omodaka, T.: 1993, *Astrophysical Journal* **404**, L63

Kaye, T. H. and Laby, G. W. C.: 1986, *Tables of Physical and Chemical Constants*, Longman, fifteenth edition

Keene, J. and Masson, C. R.: 1990, *Astrophysical Journal* **355**, 635

Kenyon, S. J. and Hartmann, L.: 1987, *Astrophysical Journal* **323**, 714

Knapp, G. R., Sandell, G., and Robson, E. I.: 1993, *Astrophysical Journal Supplement Series* **88**, 173

Knowles, S. H., Mayer, C. H., Cheung, A. C., Rank, D. M., and Townes, C. H.: 1969, *Science* **163**, 1057

Koerner, D. W., Sargent, A. I., and Beckwith, S. V. W.: 1993, *Astrophysical Journal* **408**, L93

Lada, C. J. and Reid, M. J.: 1978, *Astrophysical Journal* **219**, 95

Lewis, B. M.: 1992, in *IAU Colloquium No. 135*

Lucas, R., Bujarrabal, V., Guilloteau, S., Bachiller, R., Baudry, A., Cernicharo, J., Delannoy, J., Forveille, T., Guélin, M., and Radford, S. J. E.: 1992, *Astronomy and Astrophysics* **262**, 491

Lynden-Bell, D. and Pringle, J. E.: 1974, *Monthly Notices of the R.A.S.* **168**, 603

Malin, D. F., Quinn, P. J., and Graham, J. A.: 1983, *Astrophysical Journal* **272**, L5

Masson, C. R.: 1994, in M. Ishiguro and W. J. Welch (eds.), *Astronomy with Millimeter and Submillimeter Wave Interferometry*, Astronomical Society of the Pacific Conference Series

Matese, J. J., Whitmire, D. P., and Reynolds, R. T.: 1989, *Icarus* **81**, 24

Matthews, H. E.: 1991, *The James Clerk Maxwell Telescope. A guide for prospective users*

McCarthy, D. W.: 1979, in J. Davis and W. J. Tango (eds.), *High angular resolution stellar interferometry*, IAU Colloquium 50, p. 18.1, University of Sydney

Mendoza, E. E.: 1966, *Astrophysical Journal* **143**, 1010

Menten, K. M., Melnick, G. J., Phillips, T. G., and Neufeld, D. A.: 1990, *Astrophysical Journal* **363**, L27

Morris, M.: 1987, *PASP* **99**, 1115

Morris, M. and Bowers, P. F.: 1980, *Astronomical Journal* **85**, 724

Ohashi, N., Kawabe, R., Hayashi, M., and Ishiguro, M.: 1991, *Astronomical Journal* **102**, 2054

Pascoli, G.: 1992, *P.A.S.P.* **104**, 350

Planesas, P. and Martín-Pintado, J.: 1992, *Astrophysical Journal* **386**, L23

Podsiadlowski, P. and Clegg, R. E. S.: 1992, in P. Gondelakar (ed.), *RAL Workshop on Astronomy and Astrophysics: Dusty Discs*, p. 69

Pollack, J. B., Hollenbach, D., Beckwith, S. V. W., Simonelli, D. P., Roush, T., and Fong, W.: 1994, *Astrophysical Journal* **421**, 615

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: 1992, *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press

Quillen, A. C., de Zeeuw, P. T., Phinney, E. S., and Phillips, T. G.: 1992, *Astrophysical Journal* **391**, 121

Reid, M. J. and Menten, K. M.: 1990, *Astrophysical Journal* **360**, L51

Reipurth, B.: 1987, *Nature* **325**, 787

Rodríguez, L. F.: 1994, *Revista Mexicana de Astronomía y Astrofísica*, In press

Rodríguez, L. F., Cantó, J., and Torrelles, J. M.: 1986, *Astrophysical Journal* **301**, L25

Rohlfs, K.: 1986, *Tools of radio astronomy*, Springer–Verlag

Ryle, M. and Vonberg, D. D.: 1946, *Nature* **158**, 339

Sandell, G., Aspin, C., Duncan, W. D., Russell, A. P. G., and Robson, E. I.: 1991, *Astrophysical Journal* **376**, L17

Sargent, A. I. and Beckwith, S. V. W.: 1987, *Astrophysical Journal* **323**, 294

Sargent, A. I. and Beckwith, S. V. W.: 1991, *Astrophysical Journal* **382**, L31

Schwarz, H. E.: 1992, *Astronomy and Astrophysics* **264**, L1

Seaquist, E. R. and Bell, M. B.: 1990, *Astrophysical Journal* **364**, 94

Seaquist, E. R. and Plume, R.: 1991, *Astrophysical Journal* **367**, 200

Shlosman, I., Frank, J., and Begelman, M. C.: 1989, *Nature* **338**, 45

Soares, S. F. and Padin, S.: 1994, in M. Ishiguro and W. J. Welch (eds.), *Astronomy with Millimeter and Submillimeter Wave Interferometry*, Astronomical Society of the Pacific Conference Series

Tayler, R. J.: 1972, *The Stars: their structure and evolution*, The Wykeham Science Series

Terebey, S., Chandler, C. J., and André, P.: 1993, *Astrophysical Journal* **414**, 759

Thompson, A. R., Moran, J. M., and Swenson, G.: 1986, *Interferometry and synthesis in radio astronomy*, Wiley–Interscience

Ulich, B. L. and Haas, R. W.: 1976, *Astrophysical Journal Supplement Series* **30**, 247

van der Hulst, J. M., Golisch, W. F., and Haschick, A. D.: 1983, *Astrophysical Journal* **264**, L37

Weaver, H., Williams, D. R. W., Dieter, N. H., and Lum, W. T.: 1965, *Nature* **208**, 29

Wood, P. R.: 1979, *Astrophysical Journal* **227**, 220

Woody, D. P., Scott, S. L., Scoville, N. Z., Mundy, L. G., Sargent, A. I., Padin, S., Tinney, C. G., and Wilson, C. D.: 1990, *Astrophysical Journal* **337**, L41

Wright, M. C. H., Carlstrom, J. E., Plambeck, R. L., and Welch, W. J.: 1990, *Astronomical Journal* **99**, 1299

Yamashita, T., Hodapp, K., and Moore, T.: 1994, *Astrophysical Journal* **422**, L21

Yates, J. A.: 1993, *Ph.D. thesis*, University of Manchester

Zhou, S., Evans, N. J., Kömpe, C., and Walmsley, C. M.: 1993, *Astrophysical Journal* **404**, 232

Zylka, R., Mezger, P. G., and Lesch, H.: 1992, *Astronomy and Astrophysics* **261**, 119