# Development of disk-based baseband recorders and software correlators for radio astronomy

**Craig J West**

*A Dissertation*
*Presented in fulfillment of the requirments*
*for the degree of*
*Masters of Applied Science*
*at Swinburne University Of Technology*

*March 2004*

# Abstract

This thesis details work undertaken in the field of radio astronomy instrumentation. Specific components of the data collection and processing systems used by radio astronomers have been implemented using non-traditional approaches. Traditionally, the correlation of radio astronomy data has taken place on dedicated, specific hardware. This thesis deals with the implementation of equivalent correlators using software running on generic clusters of personal computers - the software approach to radio astronomy. Toward this end a system has been developed that records the raw telescope output onto computer hard drives, allowing easy access to the data on cluster supercomputers.

Part of this thesis describes the design, construction, testing and utilisation of these data recording systems. The correlator software used to process these data on supercomputers is also fully described, including extensive tests of the software and a detailed comparison between its output and the output of an existing hardware correlator. The software correlator is shown to produce output that agrees extremely well with the hardware correlator, verifying its accuracy and performance. Finally, results of on-going scientific investigations that use the software correlators described in this thesis are outlined, illustrating the flexibility and usefulness of the software approach to radio astronomy.

# Acknowledgments

I would like to thank Dr Steven Tingay, one of my supervisors, for the reading of thesis drafts and the motivation and direction given to this thesis. Also a thank you to Prof. Matthew Bailes, my other supervisor, for giving me the opportunity to study with him and complete the work I have done.

I would like to thank Dr Warwick Wilson of the Australia Telescope National Facility (ATNF) for helping me to understand the design and operation of the ATNF LBA correlator, and for assisting in data analysis. Dr John Reynolds, also of the ATNF, assisted with the preparation of the Parkes radio telescope, and the collection of data that form a large part of the correlator work comparison. I would also like to thank the staff at NASA's Deep Space Network station at Tidbinbilla, for also contributing to these observations. A general thank you to all the ATNF staff who have helped with the many observations that I have carried out as part of this thesis is warranted. The ATNF staff are always helpful and go out of their way to ensure that the users of their radio telescopes have the correct facilities with which to operate.

Dr Richard Dodson, formerly of the University of Tasmania, now at the Institute of Space and Astronautical Engineering in Japan, is thanked for providing an understanding of the Vela pulsar, as well as some of the observations and hardware correlation of the data. Dick Ferris, of the ATNF, provided a great deal of assistance in understanding the VLBI data acquisition systems used at the Australian telescopes.

Thank you to both Jouko Ritakari and Ari Mujunen from the Metsähovi Radio Observatory for the help in installation and use of the disk-based baseband recorders that are used as part of this thesis.

Thank you to the staff and students of the Centre for Astrophysics and Supercomputing at Swinburne University of Technology, who were able to help me

learn about radio telescopes and radio astronomy techniques, as well as spending numerous hours with at the Parkes radio telescope. I also need to thank the following individuals: Stuart Gill for being a good friend; Willem van Straten for the help in using *BasebandDSP*; Robin Humble for help with MPI; Paul Bourke for computer visualisations; and the Volleyball team for the useful frustration relief.

I would also like to thank my parents, Tony and Jenny West, for continuing to support me during my thesis and ensuring that I finally made it to the end. Those with whom I share a house, Sarah Jeffrey and Jon McNab have been very patient with me during the final stages of my thesis writeup.

This work was supported through a scholarship funded by the Major National Research Facility Program "Gemini and SKA: Australia's Astronomy Future".

# Declaration

This thesis contains no material that has been accepted for the award to the candidate of any other degree of diploma. To be the best of the candidate's knowledge this thesis contains no material that has been previously published or written, except where due reference has been made in the text. The work presented in this thesis is that of the candidate; joint research with collaborators and the contributions of colleagues have been acknowledged. Prof. Matthew Bailes and Dr Steven Tingay, the candidate's supervisors, have helped in the planning and execution of the observations and data analysis made as part of this thesis. Dr Tingay also gave guidance with regard to the content that appears in this thesis.

The *BasebandDSP* library and associated PSRCHIVE library that is used as a major part of this thesis was a joint effort involving many people: Willem van Straten, Russell Edwards, Haydon Knight, Aidan Hotan, Matthew Bailes and Dick Manchester. The original library from which these two libraries evolved was developed by Josh Kempner, Dan Stinebring, Maurizio Toscano, Matthew Bailes, and Matthew Britton. These new libraries are available to other radio astronomy groups and as such other people not mentioned may have been involved in small parts of the development.

Craig James West

31 March, 2004

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction and motivation

In this chapter a very brief introduction to radio astronomy and radio astronomy instrumentation is given, to place in context the work presented in later chapters.

Radio astronomy is the study of cosmic objects at wavelengths ranging from approximately one millimetre (300 GHz) up to approximately ten metres (30 MHz). Radio astronomy did not really become of major interest to the general astronomy community until after World War II. However, it was Karl Jansky that pioneered radio astronomy in 1932. His discovery that the galaxy was emitting signals at 20.5 MHz (Jansky 1933) sparked the interest of Grote Reber, who went on to build the first fully steerable radio telescope (see Figure 1.1) and many receivers to observe cosmic signals (Reber 1944). After the war many countries and groups started building and developing radio telescopes, growing out of the wartime effort put into radar technology (see contributions to the special issue of the Australian Journal of Physics, Goddard & Haynes 1994).

This thesis deals with very recent technological developments in computing, software, and data transport/storage and applies them to problems in radio astronomy that have traditionally been solved using dedicated hardware systems (see Chapters 2 and 3). This thesis also details observations of masers, pulsars and galaxies at radio wavelengths, objects that have been chosen because the required observations cover different areas of hardware and software techniques, and data processing requirements (see Chapter 4). Masers produce radio signals within narrow frequency ranges; pulsars are broad spectrum radio emitters that require high resolution temporal sampling; galaxies and AGN are bright, broad spectrum emitters that allow for the calibration of observations, as well as being scientifically

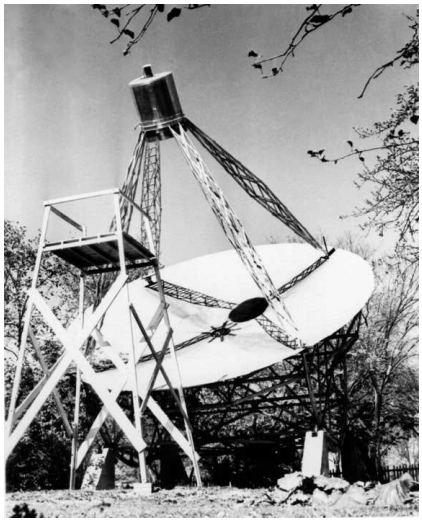

Figure 1.1: First steerable radio telescope, constructed by Grote Reber. Image courtesy of NRAO/AUI.

interesting in themselves.

The signals that are studied in radio astronomy are very weak (a strong source is approximately 1 Jy = $10^{-26}\mathrm{Wm}^{-2}\mathrm{Hz}^{-1}$)[1] and analog electronics used in telescopes amplify these signals so that they may be readily processed. Human-made signals can cause problems to astronomers, as they can be many orders of magnitude stronger than the cosmic signal of interest. Much effort is taken to ensure that telescopes are located in radio quiet areas and that the equipment at the telescopes generate a minimum of radio noise. The problems of human-made radio frequency interference (RFI) are also touched upon in this thesis (see Chapter 4).

In the rest of this chapter, very brief descriptions of both single dish and interferometric radio telescopes are given and the traditional hardware solutions and new software solutions for radio astronomy (the topic of this thesis) are outlined. Finally the aims of the thesis are stated.

[1] All the radio emission ever collected in the history of radio astronomy does not contain enough energy to raise the temperature of 1 teaspoon of water by 1 degree Celsius.

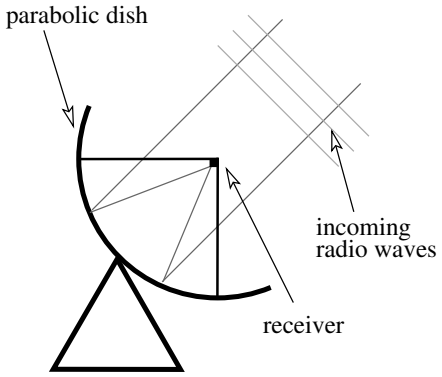


Figure 1.2: Schematic diagram of a parabolic dish receiving a radio signal and focusing it upon the receiver.

## 1.1 Single dish radio astronomy

To receive radio waves from cosmic sources most radio telescopes are built so that they collect radiation over a large area and ensure that it is concentrated coherently onto a radio receiver. The most common way of achieving this is to use a parabolic antenna (or dish) which focuses the incoming radiation onto a receiver (see Figure 1.2). The receiver, and associated electronics, which are often cryogenically cooled, amplify the signal and send it to further electronics for processing (Kraus 1986).

Fully steerable single dishes come in many different sizes. They range from small ~10 m diameter dishes up to the large 100 m diameter class dishes. Figure 1.3 shows the 64 m diameter Parkes radio telescope which was used for part of this thesis. Steerable telescopes (such as Parkes) allow observers to access a large part of the sky. Some dishes are non-steerable (such as Arecibo - the largest dish at 305 m diameter) and use the rotation of the Earth (plus limited steering of the receiver) to allow access to as much of the sky as possible. Observers usually spend many hours at a time viewing the sources they are interested in. Signal processing is used to detect the astronomical signals, usually performed in dedicated hardware processors designed specifically for the required tasks.

Observers use single dish telescopes to look at all types of cosmic sources. Single telescopes give better sensitivity the larger they get, but the area of the sky (beam size) that can be observed in a single pointing is reduced. The beam size

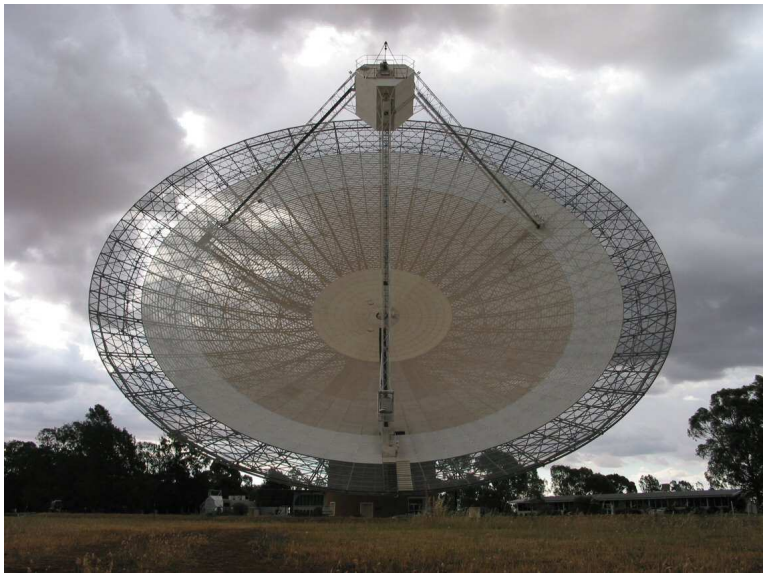

Figure 1.3: Parkes radio telescope, a 64 m diameter fully steerable dish. The enclosure containing the receivers can be clearly seen.

also decreases as the observing frequency increases. The beam size is described by the equation

$$\theta_S \simeq \frac{\lambda}{D_s},$$

where $\theta_S$ is the full width, half maximum (FWHM) beam size in radians, $\lambda$ is the wavelength of the radio waves, and $D_s$ is the diameter of the dish (Hecht 1987). Compared with the resolution of an optical telescope at 1 arcsecond, the Parkes radio telescope, when observing at 1 GHz, has a resolution of 0.25 degrees. To produce an angular resolution equivalent to a small optical telescope, a single radio dish observing at 1 GHz would have to be approximately 40 km in diameter.

## 1.2 Interferometry in radio astronomy

The technique known as interferometry was first developed at optical wavelengths by Michelson (1890). Interferometry was later developed for radio astronomy to allow multiple dishes to be synthesised into a single observing instrument. This technique is used to overcome the resolution problem that single dishes have, as mentioned above. In interferometry the equation describing the resolution of the

Figure 1.4: Four of the six 22m antennas that form the Australia Telescope Compact Array.

synthesised telescope is

$$\theta_I \simeq \frac{\lambda}{D_I},$$

where $D_I$ is now effectively the distance between the telescopes (baseline) rather than the diameter of an individual dish. When telescopes are placed close together, such as in the Australia Telescope Compact Array (ATCA), with its 6 km maximum baseline and six 22m diameter dishes (Figure 1.4 shows an image of four of the ATCA telescopes) a resolution of approximately 1 arcsecond can be obtained at a frequency of 8 GHz, a good match to the resolution of optical telescopes. The ATCA provides a larger effective collecting area than a single 22m dish, but only equivalent to a 54 m diameter dish (less than the 64 m diameter collecting area of Parkes for example). However, because these dishes are spaced apart the ATCA can produce much higher angular resolution images than Parkes.

Interferometry involves the combination of signals from pairs of telescopes within an array of N telescopes ($\frac{1}{2}$N[N $-$ 1] independent pairs). Interferometers measure the mutual coherence of the electric field at different spatial locations - the measurements of mutual coherence can be related to the structure of the cosmic radio source of which is being observed. Detailed descriptions of the theory of

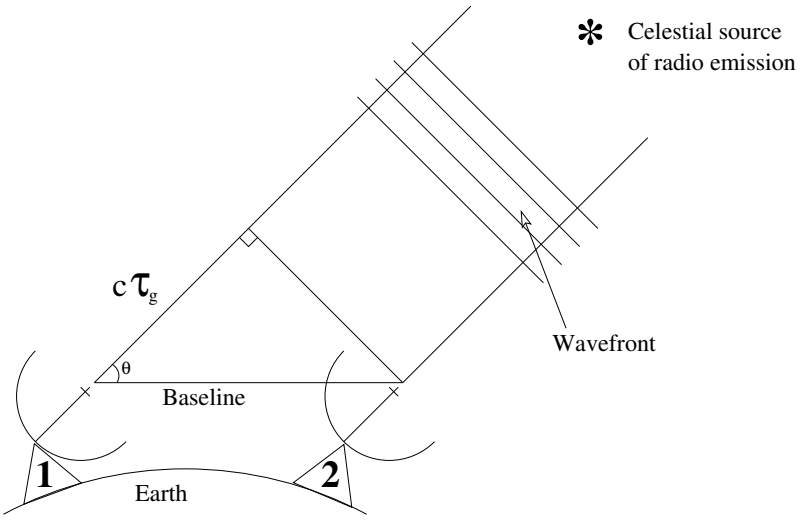


Figure 1.5: Schematic diagram of an interferometer (two telescopes).

interferometry are available in texts such as Thompson, Moran & Swenson (1994a) and Taylor, Carilli & Perley (1999) and will only be discussed briefly here.

Each telescope within the array of telescopes points at the same celestial source of radio emission (see Figure 1.5). A difference exists in the arrival time of any given wavefront at each telescope in a baseline pair (relative to some reference coordinate). In Figure 1.5 the reference coordinate corresponds to telescope 2 and the time difference (or equivalently the differential path length - known as the delay) is $\tau_g$. To measure the mutual coherence of the electric field, the delay between the signals must be first removed (other effects due to station electronics must also be accounted for - see the references). The data must then be combined, typically either by a process involving correlation of the two signals then Fourier transformation of the result (known as XF correlation), or alternatively by Fourier transformation of the individual signals and then multiplication of the Fourier transforms (known as FX correlation). A detailed comparison of XF and FX correlators can be found in Romney (1995). The resulting quantities, known as complex visibilities, are functions of frequency and time, and are related to the source structure via a Fourier transform (Lipson & Lipson 1981). Interferometry utilises the rotation of the Earth to measure the mutual coherence of the electric field at many points in the aperture plane (using fixed physical baselines). As Earth rotation causes $\tau_g$ to change, interferometers need to account for changes in $\tau_g$ as a function of time.

Very long baseline interferometry (VLBI) uses the same technique as the ATCA but with telescopes spread over much larger (continental and inter-continental) distances resulting in a resolution of milliarcseconds instead of arcseconds. Because the baseline lengths involved can be hundreds to thousands of kilometres, as well as spanning countries, it is not generally possible to collect the data at a central processing point until well after the observations have taken place; no data transport infrastructure is easily available that allows for all of the telescopes to be connected to a central processor in real-time. Chapter 3 deals with VLBI in more detail. There are many VLBI arrays around the world: the European VLBI network (EVN - `http://www.evlbi.org/`); the American very long baseline array (VLBA - `http://www.aoc.nrao.edu/vlba/`); the Australian long baseline array (LBA - `http://www.atnf.csiro.au/vlbi/`); the Japanese VLBI exploration of radio astrometry (VERA - `http://veraserver.mtk.nao.ac.jp/`); and others including one space based telescope dedicated to VLBI, named HALCA, part of the VLBI Space Observatory Programme (VSOP - `http://www.vsop.isas.ac.jp/`).

## 1.3 Thesis motivation: Hardware and software solutions for radio astronomy

### 1.3.1 The traditional hardware solution

Radio observatories (both single dish and interferometric) have traditionally used large purpose-built correlators to process the vast amounts of data that their telescopes collect. The data volumes are reduced due to this processing, formed into smaller, more manageable, data units that observers can transport on media such as tapes, CDs and (now) DVDs. These purpose built machines are very efficient at a small range of activities, for example, forming spectra from single dish observations or cross correlating the signals from a multi-element interferometric array. The purpose-built machines are flexible enough to allow the observer to customise (within reason) the output volume of data they want, and what they want the data to represent, depending on the scientific goals of their observations. For example, objects like masers can be observed with narrow frequency ranges and with high frequency resolution within those ranges. Objects like galaxies and AGN can

be observed with wide frequency ranges and relatively low frequency resolution. Pulsars, due to the need for high time resolution, are usually processed with specialised hardware which is not installed at all telescopes, and as such it is difficult to observe pulsars at those telescopes without the equipment.

The software available to observers for post-observation processing help to reduce the data volume further by averaging and allow the observers to make images and perform further analysis by extracting scientifically useful parameters from the data, such as the flux density, size, shape and spectra for the radio sources of interest. The software is typically limited to analysis of the modest data volumes that can be processed on single desktop computers.

While there are many advantages to dedicated hardware processors for radio astronomy observatories (such as efficiency and optimisation for the situations of primary interest), one of the problems with specialised processing hardware is that when new processing methods become available, or other generic hardware in the telescope system is updated, the specialised hardware often needs to be updated as well. This can be a time consuming and costly exercise. Every effort is usually made during the construction of the hardware to allow for future enhancements and additions, but this is not always possible.

This represents the traditional hardware solution for radio astronomy.

### 1.3.2 The new software solution for radio astronomy.

Generic computers can help solve the processing problems faced in radioastronomy in a different way. This is the software solution for radio astronomy. Large improvements have been made over the last 10 years in general purpose computing and the use of clusters of generic computers is becoming more common; high speed networking between cluster nodes allows rapid communication and sharing of the work load. Mass data storage now exists that can store astronomically useful amounts of raw telescope data cheaply and conveniently. Processor speeds are now such that they are starting to compete with dedicated hardware. Indeed, one of the biggest projects in radio astronomy at the moment is the Dutch Low Frequency Array (LOFAR) for which much of the signal processing will be done in software. These generic computers are available off-the-shelf, and with the economy of large-scale production have become a cost-effective alternative to some purpose built hardware. On the other hand, software to access and process radio

astronomy data, using these computers, needs to be written - the effort in designing hardware in the traditional approach is replaced by the effort in writing software. However, if a telescope is upgraded to produce double the data rate, for example, the addition of more or faster generic computers can ensure that the data is always processed on time, probably with very little change to the software. Further to this, if new processing methods are required, then anywhere from a few lines of the code, right up to the entire suite of software can be re-written without the need to rebuild the whole processor.

The raw data that are collected by a telescope (see Chapter 2) are usually sent straight to the hardware processors in the traditional approach, where they are reduced into a usable format, after which the raw measurements are discarded. This is acceptable in many cases, unless the data record is of a unique, never to be repeated event. Traditionally the storage of raw data, when required, has been to magnetic tape. Many different types of tape storage systems have been used over the years, from the spool-to-spool reel-based tapes, to the more modern video cassette tapes. These tapes are a cheap and cost effective way to store and transport data. However, when it comes to processing the data recorded on these tapes, specialised processors have generally been used.

In order to process these same raw data on a cluster of generic computers, it is necessary to transfer the data from these tapes onto computer hard drives. Systems for doing this exist, however they are time consuming and simply slow down the data processing. A better way to access the data on a generic computer is to record the data directly to hard disk, instead of first recording it to tape. High volume computer hard drives are now common. By using many of these disks in what is called a redundant array of inexpensive disks (RAID), a grouping of disks is made to look like one large drive. The data can be recorded to these disks, the disks can then be installed into a computer cluster, and data processing using software can be undertaken. A cluster of computers and random access to the data means that the computer can process the data at the speed with which it was recorded, or faster. If more speed is required, more or faster computers can be added.

The recording of raw data[2] to disk (or tape) is known as baseband recording. Once the data have been stored they are then available for processing using generic

[2]The raw data has already been partially processed as it is filtered, down-converted and digitised.

computers, although there is no reason why the data cannot be processed while being stored, and deleted when no longer needed. If the data are stored for some time, it is possible to test many variations of different data processing techniques, in order to maximise the science return from a given observation. Whilst clusters of computers may never be as fast and efficient as specialised hardware, they do allow tests of data processing methods, and they also become good general purpose computers for reducing the data after they have been processed.

The recent advances in high performance computing, and the possibilities for applying these technologies to radio astronomy, provide the motivation for this thesis, which will be an exploration of the new software solution for radio astronomy.

## 1.4 Thesis aims

With the description of the traditional hardware and new software approaches to radio astronomy outlined in the previous sections in mind, the aim of this thesis is to demonstrate software solutions for specific data processing problems in radio astronomy and to demonstrate that these solutions can be used to obtain useful scientific data.

In order to address these aims it has been necessary to develop a method for recording the raw data from radio telescopes that allows easy access to the data from generic computers. This was achieved by developing inexpensive disk-based baseband recorders. Chapter 2 summarises the use of baseband recorders in both older tape-based and newer disk-based systems and describes the development of a network of disk-based baseband recorders for Australian radio telescopes. Chapter 3 details the development and testing of data processing software for use in VLBI, a software correlator, in particular for an array of Australian radio telescopes. Chapter 4 discusses the use of both the disk-based recorders and the software correlator in scientific applications. The conclusions of the thesis are summarised in Chapter 5, along with ideas for future work.

# Chapter 2

# Development of disk-based baseband recording technologies for Australian radio telescopes

## 2.1 Introduction

Due to the high observing frequencies used by contemporary radio astronomers, greater than a few GHz, it is prohibitively expensive and unnecessary to directly record the waveform received by radio telescopes. In order to record the information contained in the waveform within a given frequency range, $\Delta\nu$, centred on an observing frequency, $\nu_0$, a series of frequency conversion and filtering stages are applied to the received analog signal, followed by quantisation. The frequency conversion and filtering stages preserve the variation of the analog waveform within the frequency range $\nu_0 - \frac{\Delta\nu}{2} < \nu_0 < \nu_0 + \frac{\Delta\nu}{2}$ and reproduce it at much lower frequencies, for example in the frequency range $0 < \nu_c < \Delta\nu$. The quantisation process produces a set of digital samples that represent the information contained in the frequency range $\Delta\nu$.

The digital samples for a band limited signal can be used to reconstruct the analog waveform if the sampling rate is equal to or greater than twice the highest frequency component present in the waveform (as stated by the well-known Nyquist theorem - Nyquist 1924). Thus, for example, if the frequency range is given as above, $0 < \nu_c < \Delta\nu$, a sampling rate of $2 \times \Delta\nu$ will allow the generation of a set of digital samples that contain all of the information present in the analog

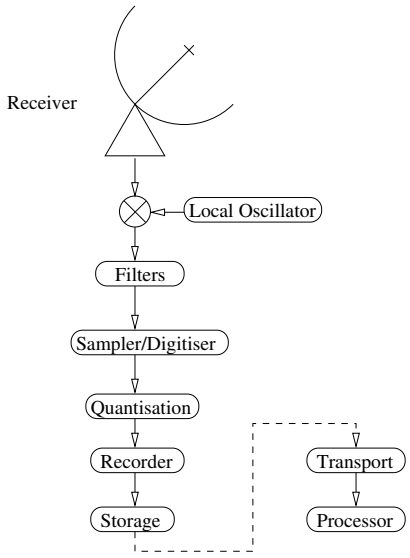


Figure 2.1: Schematic diagram showing acquisition and processing of baseband data from radio astronomy signals.

waveform [some loss of information occurs depending on how coarse the quantisation is - radio astronomy signals are typically digitised using a 4-level [2-bit] quantisation scheme, but corrections can be made to compensate for the loss of information (Jenet & Anderson 1998) that would otherwise distort spectra and pulse profiles].

Since the digital sampling rate reflects only twice the bandwidth of interest (typically 10s to 100s of MHz), rather than the observing frequency (typically greater than a few GHz), the digital data rate can be easily managed and recorded, or processed at a radio telescope. The digital data produced in this manner are known as baseband data (the term baseband reflects the fact that frequency conversion translates the analog waveform frequency components close to zero frequency). Any device that is capable of recording these baseband data can be called a baseband recorder. Figure 2.1 illustrates the typical frequency conversion, filtering, quantisation, and recording sequence that might be employed at a radio telescope (Thompson, Moran & Swenson 1994b).

## 2.2 Applications of baseband processing and example baseband recording systems

Baseband data that are stored for later processing are available to be processed in any way the user requires. For example, typical uses for baseband data are: the formation of spectra of astronomical (or man-made) radio sources; the high precision timing of pulsars - neutron stars that produce regular pulses of radio emission (Stairs *et al.* 2000; Bailes 2003); or the formation of interferometric fringes between two radio telescopes to achieve high angular resolution imaging of astronomical radio sources (Whitney 2002; Ritakari & Mujunen 2002). These applications will be explored in more detail in chapters 3 and 4 of this thesis. In many instances baseband data are processed as they are received in real-time at the telescope. In some applications (for example in the case of very long baseline interferometry [VLBI]) there is no practical alternative to recording the data for post-observation processing.

Data processed in real-time during observations are only processed once, and in one way. An advantage of recording the baseband data in all cases is that the data are then available to be processed using many different processing options at the user's leisure. Some of these options may be quite simple, such as increasing the spectral resolution of a radio source spectrum of interest. Others are more complex, such as removing from the recorded data the dispersion imposed on radio signals as they travel through the interstellar medium of our galaxy (Hankins 1971).

Traditionally in radio astronomy applications, when required, baseband data have been recorded to magnetic tape, the most well-known application being VLBI. Newer technologies allow for significant amounts of baseband data to be recorded to computer hard disk, enabling direct processing of the data using off-the-shelf computers running specialised software (Bailes 2003).

### 2.2.1 Tape-based baseband recording

Tape-based recorders for radio astronomy have existed for many years (Table 2.1). Some of the systems developed include the Mark I system (720Kb/s using video recorders and 1 bit sampling; Bare *et al.* 1967), the Mark II system (4Mb/s using

Table 2.1: Comparison of common tape-based baseband recorders

| Recording System | Sampling Rate (Mbps) | Tape duration (minutes) | Year | Reference |
|---|---|---|---|---|
| Mark I | 0.72 | 3.33 | 1967 | 1 |
| Mark II | 4 | 246 | 1971 | 2 |
| Mark III | 112 | 164 | 1977 | 3 |
| Mark IV | 1024 | 67.5 | 1993 | 4 |
| Canadian S1 | 8 | 60 | 1971 | 5 |
| Canadian S2 | 128 | 360 | 1993 | 6 |

1 - Bare *et al*. (1967), 2 - Clark (1973), 3 - Rogers *et al*. (1983),
4 - Whitney (1993), 5 - Moran (1976), 6 - Wietfeldt *et al*. (1996)

video recorders and 1 bit sampling; Clark 1973), the Mark III system (112Mb/s using instrumentation recorders, and 1 bit sampling; Rogers *et al*. 1983), and the Mark IV system (1Gb/s data-rate using 14 inch reel-to-reel tapes with 1 and 2 bit sampling; Whitney 1993). The Australian VLBI array adopted the Canadian-developed S2 system in 1994, based on super VHS video tapes and modified VHS recorders (Wietfeldt *et al*. 1996). The S2 system allows a maximum recorded data rate of 128 Mb/s. For example, one set of eight tapes in an S2 unit will allow continuous recording of 2 x 16 MHz bands sampled at the Nyquist rate with 2-bit quantisation (128 Mb/s) for 6 hours. Prior to 1994 the Australian VLBI array used the Mark II system.

### 2.2.2 Disk-based baseband recording

Over the last five years there has been a growing interest in disk-based recording of baseband data for radio astronomy, as the capabilities of cheap off-the-shelf computing equipment have improved exponentially. Several well-developed systems now exist. The Mark V recording system has been developed by the Haystack Observatory of MIT (supporting a data rate of 1 Gb/s using 16 hard drives making a 24 hour unassisted runtime possible; Whitney 2002). The Japanese K5 recorder (also known as PC-VSSP) has been developed at the Kashima Space Research Center - Communications Research Laboratory (capable of 1 Gb/s recordings using 4 120 GB hard drives; Kondo *et al*. 2003). An effort to use the cheapest possible off-

the-shelf hardware has been made in Europe, resulting in the PCEVN hardware (512 Mb/s recording speed using generic PC components and a specialised DMA card). As part of this thesis the PCEVN hardware system has been used as the basis for development of a disk-based baseband recorder for Australian radio telescopes (see Section 2.3).

An example of a disk-based baseband recorder, used for pulsar studies, is the second generation Caltech Parkes Swinburne Recorder (CPSR2; Bailes 2003). This system records baseband data but does not store it indefinitely, the data are automatically processed in near real-time (1024Mb/s recording rate with a total disk space of 4TB).

Disk-based recorders are also attractive because they can reduce costs. The initial outlay for a tape-based recorder is more than for a disk-based recorder capable of the same data-rate. A good example of this is the S2 recording system compared to the new Swinburne recording system. An S2 recorder costs approximately USD$60 K, whereas the Swinburne system costs around USD$20 K (including the disk space for a single 36 hour recording period at a rate of 128Mbps). However, the cost of the extra storage media needs to be considered. Data storage on S2 tapes costs ~$0.25/GB, whereas storage for the Swinburne recorders costs ~$1/GB. The cost of producing hard drives is reducing, with estimates that tape-based and disk-based recorder mediums will cost the same, per gigabyte of data, around 2005-6 (Whitney 2002).

## 2.3 The development of disk-based baseband recorders for Australian radio telescopes

### 2.3.1 Hardware

Disk-based baseband recorders have been developed for Australian radio telescopes, as part of this Masters thesis. The recorders consist of an interface to the VLBI data acquisition system (DAS) units available at the telescopes, data capture hardware, an off-the-shelf computer, and large volume disk drives. The interface (VLBI Standard Interface Converter [VSIC]) and data capture (VLBI Standard Interface I/O Board [VSIB]) hardware use the VLBI Standard Interface

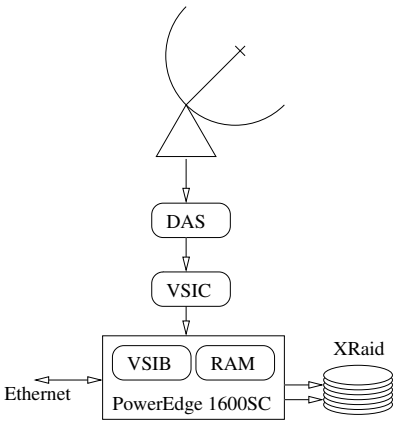


Figure 2.2: Schematic diagram of the Swinburne-MRO recording setup. The telescope includes all the necessary components to convert the incoming signal from the observed frequency to baseband for recording.

(VSI) format[1] and were purchased from the Metsähovi Radio Observatory (MRO) in Finland (this hardware also forms the basis of the PCEVN recorders mentioned in section 2.2.2) (Ritakari & Mujunen 2002). The off-the-shelf computer chosen for the recorders is a Dell PowerEdge 1600SC server class machine. The disk storage chosen was Apple XRaid disk drives (2.2 TB configured under Raid5). These drives were chosen as they were cost effective as well as cable of recording at 100MB/s with removable drives. Figure 2.2 shows a schematic diagram of the recorder setup, which has been named the Swinburne-MRO baseband recorder.

MRO provide a recommended configuration for the computer to be used in conjunction with the VSIC and VSIB hardware, specifying an NVidia NForce chipset. At the time the VSIC and VSIB boards were developed, this chipset provided very fast data transfer rates from the PCI slots to the RAM. MRO also suggest at least 512MB of RAM and that the disk space should be made from a redundant array of inexpensive disks (RAID). The remainder of the computer specifications are left up to the user's particular requirements.

After reviewing the recommended specifications, and looking at the user requirements, e.g. the use of the Apple XRaid external disk drives, it was decided that the recommended chipset would not be suitable; there were no motherboards available at that time that supported 64 bit PCI cards (needed for the XRaid units)

[1]Full details for the VLBI Standard Interface are available at: `http://web.haystack.edu/vsi/`

as well as the NVidia NForce chipset. After studying technical documents provided by Dell Computer Pty Ltd, the Dell PowerEdge 1600SC was chosen. This machine is a server class computer with 3 PCI buses utilising the ServerWorks GC-SL chipset. The VSIB board plugs into the 32 bit PCI bus (internal to the machine), the XRaid plugs into the 64 bit PCI bus, and an extra ethernet card plugs into the PCI-X (also 64 bit) bus. This configuration ensures that data can flow from the VSIB into RAM, and then back out to the XRaid with as little interruption as possible.

The machine is also equipped with 1 GB of DDR RAM and a 2.4 GHz Intel Xeon processor, with a spare slot for a second CPU. A further 3 GB of RAM can be optionally installed. The machine has a single internal 36 GB SCSI hard drive (with space for another 5 SCSI disks). A SCSI drive was chosen because the system architecture indicated that an IDE disk drive could cause a potential bottleneck in the data flow. The Apple XRaid is connected via copper fibre channel and supports a data transfer rate of 100 MB/s through each of two connectors. The XRaid is configured under RAID-5, which allows for the failure of a single disk without loss of data, but also ensures that the data transfer rates are kept high. The Apple XRaid units are made up of two RAID sets, each of seven disks. Both RAID sets are available at the same time but are separate units. Recording only takes place on one RAID set at a time.

The VSIC board was shipped as a pre-made circuit board. An external housing was built to fit the board, fitted with two inputs (from the digital acquisition system [DAS] units at many Australian radio telescopes - these units are used to digitally sample the data from the radio telescopes) and two outputs (to the VSIB board). The VSIC also requires a 4.5 V power supply with the ability to run at 1.5 A. There are pins on the VSIC boards that allow different data conversion modes to be selected. Switches were installed on the housing in order to set these pins so that the test mode or S2 recording modes could be easily selected (the DAS units output data formated to the S2 specification). There are other data formats that the VSIC can convert and these modes are also accessible via the switches.

The output connectors on the VSIC housing are located on the front of the case, along with the switches and power connector. The power can be disconnected if needed. The rear of the housing is used for the input connectors. Figure 2.3 shows the VSIC housing with the lid removed.


Figure 2.3: The top panel shows the front of the VSIC box as well as the power supply. The bottom panel shows the rear of the VSIC box as well as the switch wiring.

A users guide that explains in detail how to set up the recorder hardware at the telescope is included as Appendix A to this thesis.

### 2.3.2 Software

The MRO provided software with the VSIB card which allows for basic recording of data. This software was modified to add timestamp information into the first 16 characters of each file containing recorded data. This is critical for applications such as pulsar timing and VLBI. Future versions of the recording software will include larger headers with more detailed information to allow for easier processing of the recorded data, and extra recording modes. Another modification that was made to the shipped software was to set the recording speed to 256Mb/s. The DAS recorders can produce data rates as high as 256Mb/s via the S2 output port, however, the S2 recorders can only record at a maximum rate of 128Mb/s. As the

speed of 128Mb/s was not an available option for the recorders the next highest speed was chosen (256Mb/s), resulting in half the recorded data being invalid[2]. However the invalid data can be discarded during processing. The recording software could have been modified to remove the invalid data before it was stored to disk, but this was not attempted as part of this thesis. The data are recorded as a series of files, each file usually 10 seconds long, however the user can select longer or shorter file sizes as required. The 10 second file lengths were chosen to ensure files were not overly large and that if a file is corrupted only a small amount of data would be lost. Each observation consists of a large number of these 10 second files. During testing of the hardware and software no errors were encountered, and no data were lost. A 10 second file recorded at 128Mb/s uses 153MB of disk space.

The recording software uses a linux driver to allow access to the VSIB card and reserves a section of RAM that is made unavailable to the rest of the system. Reserving this RAM ensures that it is always available and that there is no delay in accessing RAM when recording starts.

The timestamp inserted at the start of each recorded file consists of 16 characters which make up a complete date. The format is kept simple and is set as 'YYYYMMDD:hhmmss\n', 4 digits for the year, two digits for the month, two digits for the day of the month followed by a colon, two digits for a twenty four hour readout, two digits for the minutes and two digits for the seconds. The '\n' is used in C programming to denote a newline character. There is no need for any more time precision beyond the seconds even though millions of samples are recorded each second. The reason for not requiring greater precision is that the recorder is forced to start precisely on a one second boundary. This is done by utilising the one pulse per second (1PPS) signal that is present at the telescopes. The 1PPS signal is linked to the very accurate clock (Hydrogen maser) at each observatory and is used to keep all station equipment running on the correct time.

The recording software was modified to ensure that it correctly starts on a 1PPS signal. The VSIB has a mode for triggering on a 1PPS signal. This was enabled. The timestamp information is derived from the PC clock, which is continuously checked and updated against the accurate station clock, via the network time protocol (NTP).

[2]The data is not always invalid, but may be duplicated depending on the mode of the DAS.

### 2.3.3 Post-observation processing software

Once the data are recorded onto disk any invalid data needs to be excised (see previous section). A program called `vsib_converter` was written to excise invalid data. This program also sorts the data into a format that is more suitable for post-processing. Future versions of the program will achieve this in real-time during recording. The main reason for not implementing this during testing was to ensure that the systems were not under too much strain. Adding a second processor (as mentioned in Section 2.3.1) would help to reduce any processing loads if the recording software was to become more complex.

Processing of the converted baseband data has been made easier by the development of a programming library called *BasebandDSP*[3]. This library has been developed at the Swinburne University of Technology, with some modifications and additions as part of this thesis work. These modifications covered methods for access to the new recorded data, seeking to specific locations in files, debugging and minor bug fixes of existing code, additional utilities and functions, extensions to the method of transferring data between multiple computers, and the addition of an interferometric library to correlation of data from multiple telescopes. *BasebandDSP* supports many file formats, including those produced by CPSR2 and S2 recorders. Most software written using *BasebandDSP* is able to read CPSR2 and S2 data formats. A new file format reader was written, which was then added to *BasebandDSP*, allowing any data recorded with the VSI format to be processed using the available suite of software in *BasebandDSP*, plus any *BasebandDSP* software written in the future.

A spectrum of the recorded frequency range (bandpass) is generated by taking the autocorrelation of the data and Fourier transforming it into the frequency domain. A small program called `bandpass` was written to autocorrelate the raw input data, Fourier transform the result, and allow for integration in time. This program produces spectra of the radio source being observed in the frequency range of interest.

A much more complex program that uses *BasebandDSP* is the prototype software correlator for VLBI that was developed as the major part of this thesis. Details of the software correlator are given in Chapter 3.

[3]Full details of the library are available at: `http://astronomy.swin.edu.au/pulsar/software/libraries/dsp`

### 2.3.4 Baseband recorder test observations at the Australia Telescope Compact Array

Initial tests of the Swinburne-MRO recording system in the laboratory ensured that the recorder could maintain the required data rates. It was found that these units would sustain a data rate of around 600Mb/s. A data rate of only 128Mb/s is required to record at the same rate as an S2 recorder. The DAS systems used at the telescopes are capable of producing 256Mb/s via the S2 connector port.

The first astronomical test observations were conducted at the Australia Telescope Compact Array (ATCA) between 2-6 October 2003. A recorder was installed at the telescope and testing began. An observation of the methanol maser G345.010+1.79 (see Figure 2.4) showed that the recorders were able to successfully store data. The software used to process the data to form the spectra shown in Figure 2.4 is detailed in section 2.3.3 above. The figure shows the left circular polarisation (LCP) in the lower panel, and a narrower frequency range at both LCP and RCP (right circular polarisation) in the upper panel.

The first VLBI observations took place during the VLBI observing session between 20-28 November 2003 using the ATCA and the Parkes radio telescopes. Pulsars and continuum radio sources were observed. Results from these observations are detailed in Chapter 4.

Further single dish tests were carried out on March 1, 2004. These tests were used to demonstrate recording at double the data rate of the S2 tape-based recorders. Figure 2.5 shows a comparison between a single 16 MHz recorded band, and a 32 MHz recorded band. The two recordings were taken 1 minute apart using the 50cm receiver at Parkes with a centre frequency of 685 MHz. It can be seen that the spectral features that appear in the 16 MHz band align correctly with the same feature in the 32 MHz band. Further details of this comparison and the other tests can be found in section 2.4.

## 2.4 Disk-based baseband recorder comparisons

After the testing of the Swinburne-MRO baseband recorders at double the data-rate mode, when compared with S2 systems, during March 2004 (see section 2.3.4), direct comparisons were made to ensure that the data were valid. Two one minute

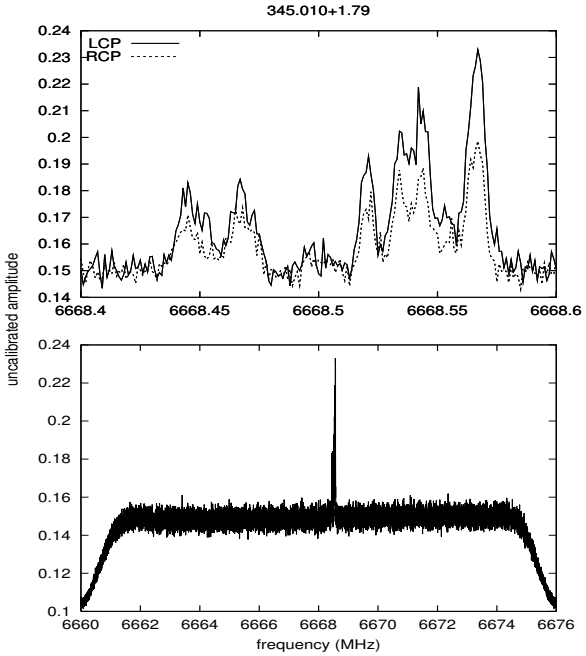


Figure 2.4: Spectral lines of maser G345.010+1.79. Both plots were generated using 16,000 frequency channels and an integration time of 60 seconds. The bottom plot shows the full frequency range, 16 MHz of bandwidth at LCP. The top plot shows the frequency range of 6668.4 to 6668.6 MHz for both LCP (solid) and RCP (dotted). Data were recorded at the ATCA using a tied array of 5 antennas and a centre frequency of 6668 MHz.

data recordings were made, one at 16 MHz and the other at 32 MHz, with a separation of one minute between the recordings. By forming twice the number of channels in the 32 MHz spectrum, compared with the 16 MHz spectrum, it is possible to obtain channels of the same width in both spectra. The result can be seen in Figure 2.5. Differences between the two plots and the spectral features are related to the data being observed at independent times. The spectral line features are from radio frequency interference (RFI) and are known to be time varying.

By making a comparison with another observing system it can be seen that the Swinburne-MRO recorders are producing results that are correct. The CPSR2 recorder (see section 2.2.2) produces a 64 MHz bandpass, and by ensuring that both CPSR2 and the Swinburne-MRO systems record data simultaneously using the same centre frequency, it is possible to align the bandpasses to produce plots that confirm the recorded signals are the same. Figure 2.6 shows just such a plot.

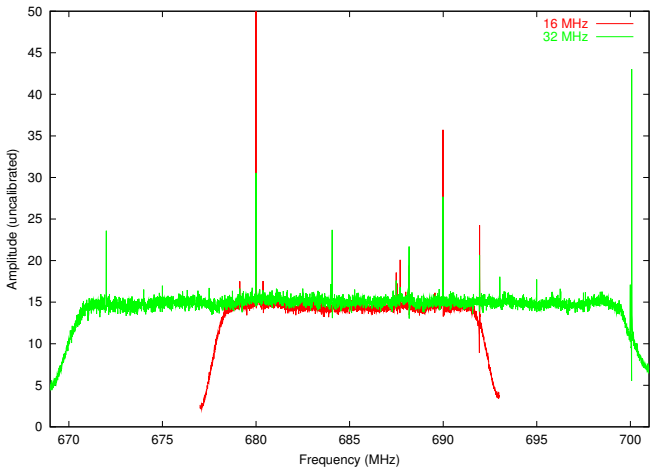


Figure 2.5: Swinburne-MRO recorders at 16 MHz (red) and 32 MHz (green), both spectrum are at LCP.

The centre frequency is 685 MHz and the source being observed was a continuum source. At 50cm (or 685 MHz) there is a large amount of known RFI, which produces the spikes that align in the two spectra. To ensure that a correct comparison can be made the channel widths of both the 64 MHz and 16 MHz bands have been set to be the same. This means that the 64 MHz band has four times more channels than the 16 MHz band. The differences seen in the overall bandpass shapes are attributed to the different electronics in the DAS compared to the filters in the CPSR2 front-end.

## 2.5 Conclusion

Baseband recorders have evolved from using tape based systems to more flexible computer recorders with large disks. The MRO hardware and off-the-shelf computing components have allowed the development of baseband recorders for radio astronomy using both single dish and interferometric telescopes. The development of these recorders for Australian telescopes has been successfully undertaken. The recorders successfully recorded data, with many of the results available in Chapter 4. The recorders have been tested under a number of situations and they appear to perform very well, matching and exceeding the capabilites of the currently used S2 system. In particular, data rates doubling the S2 maximum data rate have been

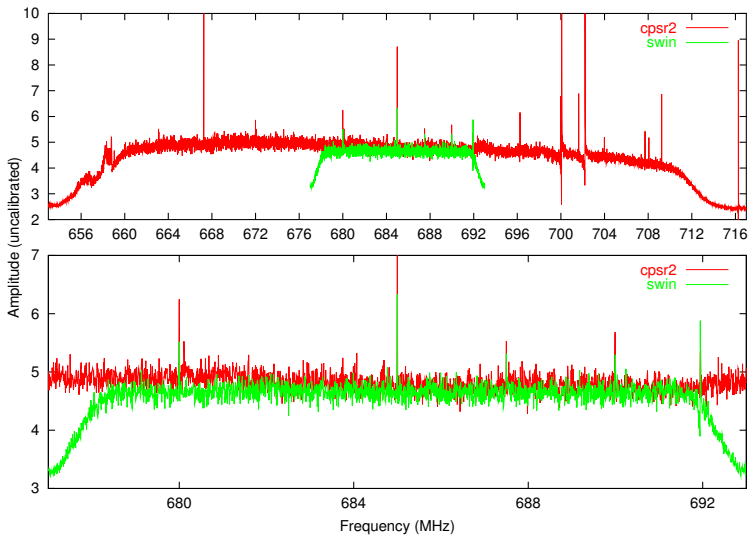


Figure 2.6: Comparison of a 64 MHz bandpass from CPSR2 and a 16 MHz bandpass from the Swinburne-MRO recorders. The upper panel shows the entire 64 MHz centred on 685 MHz. The lower panel shows the recorded 16 MHz from the Swinburne-MRO recorder, with the zoomed section of the CPSR2 recorder.

achieved with the Swinburne-MRO recorders.

By using *BasebandDSP* we are able to write software for these new recorders that will be able to use the output from other recorders. We can also use the already developed suite of software that exists in *BasebandDSP*.

# Chapter 3

# Software correlation of very long baseline interferometry data

## 3.1 Introduction

The first long baseline interferometers were built (using magnetic tape media for data recording and independent time standards at each antenna element) in the late 1960s and used general-purpose computers to correlate the recorded signals in software, in order to form the complex visibility functions and deduce information about the high angular resolution structure of the radio sources observed. The first software correlator for VLBI was implemented on an IBM 360/50 at the National Radio Astronomy Observatory (NRAO), Charlottesville, and was used to process data recorded using the NRAO-Cornell Mk I VLBI recording system (Bare *et al*. 1967). The first scientific result to be obtained using this system was a measurement of the angular size of 3C273B, $< 0.005''$ (Clark, Cohen & Jauncey 1967). At the same time a team at MIT were also developing software correlators for VLBI, initially on a CDC3300 at the Haystack Observatory and then on an IBM 360/92 at the Goddard Space Flight Center (GSFC), primarily for processing geodetic VLBI observations recorded on Mk I tapes and spectral line VLBI observations (Moran *et al*. 1967).

Within a few years, the data rates produced at the antennas required faster correlators, which advances in digital electronics provided through the 1970s and 1980s. Some well-known examples of recent VLBI hardware correlators are: the correlator of NRAO's dedicated Very Long Baseline Array (VLBA – Rom-

ney 1995); the MkIV correlator of the Joint Institute for VLBI in Europe (JIVE – Casse 1999), used to correlate data from the European VLBI Network (EVN); the Canadian S2 correlator, used primarily for geodetic programs and supporting space VLBI observations which use the S2 recording format (Carlson *et al*. 1999); the Japanese VSOP space VLBI correlator at Mitaka (Horiuchi *et al*. 2000); and the S2 correlator of the Australia Telescope National Facility (ATNF), for processing data from its Long Baseline Array (LBA – Wilson, Roberts & Davis 1996), the only Southern Hemisphere VLBI network. The great progress in our understanding of compact radio sources over the last 20 years has been due to these and other sophisticated purpose-built machines and the organisation of several *ad hoc* and dedicated arrays of radio telescopes.

In the 1990s great advances in hard disk storage and computer processor speed were made, following the celebrated Moore's law, in which processor speeds double approximately every 18 months (Moore 1965). With these increases in computing power there has been a renewed interest in software correlation in radio interferometry, generally for niche applications. For example, fringes have been detected between the 34 m antenna at Kashima (Japan) and the Westford 20m radio telescope at the Haystack Observatory (USA) using a software correlator developed by the Communications Research Laboratory (CRL) in Japan (Kondo *et al*. 2003). This software correlator is planned to be used as a real-time fringe checking system for the European VLBI Network (EVN). The first EVN e-VLBI image was produced on the 15th of January 2004[1].

A software correlator is being developed at the Astro Space Center (Russia) based on recording using the S2 system. First fringes using this system were obtained for the source 0235+164 using a 2 MHz bandwidth and correlating 2048 frequency channels, on the baseline from Bear Lakes (Russia) to Tidbinbilla (Australia) (Chuprikov *et al*. 1996).

A software correlator has been developed to cross-correlate data from the Rapid Prototyping Array (RPA) at the University of California, Berkeley (Harp 2002). The RPA is a test-bed system to prove technology and techniques for the much larger Allen Telescope Array (ATA), which will be used to search for extra-terrestial intelligence and is considered a possible model for the future Square Kilometre

[1]Details about the process and results can be found at `http://www.evlbi.org/evlbi/tevlb8/tevlb8.html`

Array (SKA).

A mixture of software correlation and hardware correlation is planned for monitoring the descent of the Huygens probe into the atmosphere of Titan (Saturn's largest moon). The JIVE correlator (hardware) will be used to obtain phase closure of an Earth wide VLBI network monitoring the descent. After this the data, which is to be recorded onto Mk V (and other disk based recorders), will be sent to a linux cluster for software correlation. The data is to be filtered into a narrow bandwidth where the probe is transmitting. It is expected that there will be enough accuracy to report the descent of the probe to sub-kilometre resolution as it enters into the moon's atmosphere (Pogrebenko *et al*. 2003).

While it is true that software correlators will probably never be as fast or as efficient for large arrays of antennas as dedicated hardware correlators, it is equally true that in certain circumstances software correlators have some unique advantages over hardware correlators. In this chapter these advantages are outlined and the case is made for software correlators. The technical details of a prototype software correlator that has been written to correlate VLBI data from the Australian Long Baseline Array (LBA), are presented.

This work was motivated by several factors. Firstly, global VLBI arrays are beginning to move to hard disk recording of the sampled data produced at each array element, rather than traditional magnetic tape based recording systems (see the proceedings of the IVS symposium Korea 2002 [Minh 2002] and Chapter 2 of this thesis). This move is prompted by the cost per byte of storage and the ease with which data can be transfered once it is recorded on disk (for example via optical fibre links to a remote correlator via standard transfer protocols). Since data in the future will be written to hard disk, the data are easily available for processing on a software correlator.

Secondly, computer processing and storage capabilities are now at the point where scientifically useful amounts of recorded data can be correlated in a reasonable amount of time, on machines that are continually becoming faster and more affordable. Software processing of baseband recorded data has been shown to be highly successful in the case of pulsar timing observations at single dish radio telescopes (for example the CPSR2 system - see section 2.2.2).

Finally, software correlation of interferometer data allows vastly different correlator architectures to be implemented, modified, extended, and compared simply

with modifications of software, with no particular knowledge of digital electronics. Also, once efficient software is developed to perform the required tasks, the capability of the correlator can be increased very quickly simply by adding more processing power to parallel computing networks. Correlation of the signals across an array of antennas is an "embarrassingly parallel" computing task which is perfectly suited to parallel supercomputers. Therefore, software correlators are intrinsically extensible. The flexibility, extensibility, and upgradability of software correlators means that many different correlation schemes can be tested on small scale systems before being fully implemented on large-scale systems, making the initial cost outlay for the correlator small and correlator development attractive for small to medium sized arrays of antennas.

Another point in favour of software correlators in certain situations is that in the future the data rates produced from VLBI observations will be far greater than they are today; the extreme will be the data rate produced by the proposed Square Kilometre Array (`http://www.skatelescope.org/`). Regardless of whether software, hardware, or hybrid correlators are used to produce the correlated data, large parallel supercomputers will be required to form wide-field, high-sensitivity images from the correlated data. Pioneering work in this field is being done by Garrett and collaborators (Garrett *et al.* 1999). Given that powerful computing resources will be required to process the correlated data it makes sense to explore the concept of software correlation. Correlation of the data, image formation, and image processing on the same computer may be an elegant and efficient solution for radio interferometry.

In a wider sense, large-scale parallel supercomputing is one of the growth areas of modern science and offers a chance to bring radio astronomy in close touch with information technology. For example, if a parallel supercomputer does not need to be dedicated completely to software correlation, it could also be used in a diverse range of activities: simulating human brain function; computing atmospheric or oceanic models; predicting the structure of complex proteins; or calculating evolutionary models of galaxies, the universe, or relativistic jets; simply to name a few possible supercomputing applications.

In the following sections of this chapter the technical details of a prototype software correlator for VLBI that runs on a parallel supercomputer at the Swinburne University of Technology are described and example correlations are shown

to illustrate the performance and scientific usefulness of the correlator. Detailed scientific results on the Vela pulsar obtained using this correlator are presented in Chapter 4.

## 3.2 Software correlator: description of correlation scheme

### 3.2.1 Overall strategy

We adopt an XF style algorithm, in order to mimic the operation of the LBA correlator of the ATNF (Wilson, Roberts & Davis 1996). A geometric delay model for each baseline in the array is generated[2]. To the geometric delay model is added an additional delay as a function of time which is due to relative clock offsets and drift rates at the different array stations. The total delay is applied to one data stream on each baseline to align the streams to within $\pm 0.5$ samples (the fractional sample error). The fractional sample error results from the data being digitally sampled, while the delay present between the telescopes is an analog delay, thus the samples can only be aligned to the nearest digital sample. This fractional sample error is known to high precision and is retained in order to correct the resultant phase slope across the frequency band in post-correlation processing. Data from each aligned pair of data streams is passed to an Intel Integrated Performance Primitive (IPP) routine that returns the cross correlation function. The cross correlation function is Fourier transformed to provide the real and imaginary parts of the complex visibility, as a function of frequency, for that time period. Note that no fringe-rotation is performed on either of the sampled data streams.

The complex visibility, as described, is used to form amplitude and phase as a function of frequency. Since no fringe-rotations or corrections for fractional sample errors were performed on the data prior to correlation, substantial phase slopes exist as a function of both frequency and time. Both of these slopes are also variable with time. The phase slope in time in the correlated output will be equal to the fringe-rotation rate that was not applied prior to correlation,

[2]Geometric delay models are generated using the CALC (`http://gemini.gsfc.nasa.gov/solve/solve.shtml`) software running on a Digital VAX VMS machine at the ATNF.

$$\frac{d\phi}{dt} = 2\pi\nu_0 \left[ \frac{d\delta}{dt} + \frac{dc}{dt} \right], \tag{3.1}$$

where $\frac{d\phi}{dt}$ is the derivative of the phase with respect to time (phase slope in time), $\nu_0$ is the effective LO frequency, $\frac{d\delta}{dt}$ is the derivative of the geometric delay with respect to time, and $\frac{dc}{dt}$ is the derivative of the clock offset between the two antennas with respect to time. Provided that the data are correlated on suitably short timescales, it is feasible to almost completely remove this phase slope from the post-correlation data in post-correlation fringe-rotation.

The fractional sample error is an error in the delay and manifests itself as a phase slope in frequency across the observing bandwidth. The phase is zero at the lower edge of the band and has a linear slope across the band, having a phase at the upper edge of the band equal to $\pi f$, where $f$ is the fractional sample error and varies between $\pm 0.5$ in this implementation. As the fractional sample error is known to high precision, the corresponding phase slope in frequency can be calculated and removed from the data in post-correlation processing, again assuming that correlations are performed on sufficiently short timescales.

Since the fractional sample error is corrected in post-correlation processing, a small decorrelation of the amplitude across the band occurs, due to changes in the fractional sample error during the time range over which the correlation is formed. The degree of decorrelation is calculated to be, assuming that the integration time is short compared to the time taken for the delay to change by one sample, $(2\Delta\nu\frac{d(\delta+c)}{dt})^{-1}$,

$$\mathrm{sinc}(\Delta\theta_n) = \frac{\sin(\frac{\pi n}{n_{chan}}\frac{d(\delta+c)}{dt}\Delta\nu\Delta t)}{\frac{\pi n}{n_{chan}}\frac{d(\delta+c)}{dt}\Delta\nu\Delta t}, \tag{3.2}$$

where $n$ is the channel number, $n_{chan}$ is the total number of frequency channels across the band, $\Delta t$ is the period over which the correlation function is formed, and $\Delta\theta_n$ is the phase change in channel $n$ in the period $\Delta t$.

Similarly, and more seriously, there is a decorrelation of the amplitudes over the time $\Delta t$ due to the fact that fringe rotation is performed as a post-correlation step. The expected decorrelation of the amplitude due to post-correlation fringe rotation is,

Table 3.1: Maximum decorrelation due to post-correlation fringe rotation for given frequencies, bandwidths and number of channels.

| Frequency (GHz) | Bandwidth (MHz) | Channels | Decorrelation (%) | Example observing mode |
|---|---|---|---|---|
| 1.4* | 4 | ≤512 | <0.25 | Neutral hydrogen |
| 1.4* | 8 | ≤1024 | <0.25 | Neutral hydrogen |
| 1.4* | 16 | ≤512 | <0.02 | Low frequency wide field imaging |
| 2.3 | 16 | ≤512 | <0.12 | Low frequency wide field imaging |
| 4.8 | 16 | ≤512 | <0.5 | High frequency continuum |
| 8.4 | 16 | ≤512 | <2.0 | High frequency continuum |
| 22.2 | 16 | ≤256 | <3.0 | High frequency continuum |
| 6.7 | 4 | ≤512 | <16 | Methanol masers |
| 6.7 | 8 | ≤1024 | <16 | Methanol masers |
| 22.2 | 4 | ≤128 | <11 | Water masers |
| 22.2 | 8 | ≤256 | <11 | Water masers |

* Ceduna not included as there is no 1.4 GHz receiver.

$$\mathrm{sinc}(\Delta\phi) = \frac{\sin(\pi\nu_0 \frac{d(\delta+c)}{dt}\Delta t)}{\pi\nu_0 \frac{d(\delta+c)}{dt}\Delta t}, \tag{3.3}$$

where all terms have been previously defined and $\Delta\phi$ is the change in phase over the period $\Delta t$.

Using the expressions above, it is possible to determine the useful practical limits of this correlation scheme, since when the correlator integration time and/or the time derivative of the delay and/or the frequency of observation become too large, an unacceptably high level of decorrelation will be encountered. When the decorrelation is small the correlated amplitudes can be corrected in a straightforward way, since they are known, with only a very small loss in signal to noise. A discussion of the limits of applicability of the correlator, from this point of view follows.

### 3.2.2 Applicability of strategy

Table 3.1 gives examples of the degree of decorrelation in the amplitude of the complex visibility, as a consequence of post-correlation fringe rotation, as de-

scribed by equation 3.3, calculated for various real observing modes of the Australian LBA[3]. The integration time for each entry in Table 3.1 is taken as the minimum number of samples required to form the required lag spectrum for the specified number of frequency channels ($\Delta t = n_{chan}/\Delta\nu$). After each individual lag spectrum is produced, Fourier transformed, and corrected for fringe rotation and the fractional sample error, to produce the complex visibilities, the visibilities can be integrated coherently over longer periods of time. Alternatively, if the decorrelation losses are very small for the observation mode of interest, a number of lag spectra can be accumulated and then Fourier transformed and corrected. The latter case will provide greater efficiency in correlation.

Similarly, the decorrelation in the amplitude due to the fractional sample error can be calculated via equation 3.2. However, it can be shown that even for the most extreme case encountered by the LBA, that the percentage decorrelation is less than 0.1%. Thus, these numbers are not tabulated here.

It should be noted that the tabulated decorrelation of the amplitude is for the worst case scenario, for the largest delay-rate produced on the worst affected baseline available in the array at that observing frequency. The average decorrelation on the worst affected baseline is in general approximately half of the value tabulated. Also, other baselines in the array will be affected substantially less than indicated in Table 3.1 for the examples given. The examples given are calculated for sources near the equator; the decorrelation becomes substantially less on all baselines when a source in the far southern sky is considered, which is typical for the LBA. Therefore, the decorrelation losses for the majority of the time on the majority of the baselines in the array are far less than indicated in Table 3.1.

Table 3.1 clearly shows the consequences of different observing setups. In general, our strategy is broadly applicable to all typical observing modes at low frequency (1.4 and 2.3 GHz, all bandwidths) and maximum bandwidth (16 MHz, all frequencies). As expected the performance of the system is sub-optimal for narrow band observations, particularly for large numbers of frequency channels and at high observing frequencies. Our strategy is therefore most particularly suited to continuum observations that use the maximum bandwidth available for the purposes of sensitivity. In this case a large number of frequency channels, particularly at low frequencies, for the purposes of wide-field imaging, can be produced.

[3]Detailed parameters for the LBA are given at `http://www.atnf.csiro.au/vlbi/`

### 3.2.3 Amplitude calibration

The lag spectra calculated by the correlator are in the form of correlation coefficients. At each lag produced by the correlator, the correlation coefficient is calculated as (between antennas $x$ and $y$):

$$\rho = \frac{N\Sigma xy - \Sigma x\Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}, \tag{3.4}$$

where $x$ and $y$ denote the samples on antennas $x$ and $y$ starting at a given point in time in each antenna sample stream and $N$ is the number of sample pairs that contribute to an individual spectrum. The individual lag spectra are then Fourier transformed to produce complex visibilities. The visibility amplitudes are then calibrated into flux density units (Jy) by multiplying the amplitude in each frequency channel by the correlator $b$-factor [$b(\nu,t)$], multiplying by the square roots of the measured system temperatures (in units of degrees Kelvin [K]) for antennas $x$ and $y$, and dividing by the square roots of the measured gains for antennas $x$ and $y$ (in units of K/Jy), following Cohen *et al.* (1975),

$$S(\nu,t) = b(\nu,t)\rho(\nu,t)\sqrt{\frac{T_x}{G_x} \times \frac{T_y}{G_y}} \tag{3.5}$$

The correlator $b$-factor, as discussed by Cohen *et al.* (1975), contains all the known sources of signal loss within the data collection and correlation system. In particular, in the case of the software correlator, the following sources of loss are known:

$b_1 = 1/0.88$, due to 2-bit sampling of the analog signal;

$b_2(t)$, due to post-correlation fringe-rotation, as discussed in detail in section 3.2.1. $b_2$ is variable as a function of time, since the time derivative of the delay varies with time;

$b_3(\nu,t)$, due to post-correlation correction of the fractional sample error, also as discussed in detail in section 3.2.1, and is a function of both time and frequency,

$$b(\nu,t) = b1 \times b2(t) \times b3(\nu,t)$$

Other losses, such as those due to imperfect frequency standards at each antenna, are negligible. Further calibration of the flux density scale is possible by

comparison of the visibility amplitudes with the amplitudes of an appropriately observed calibration source within the same experiment.

## 3.3 Software correlator: description of hardware and software

### 3.3.1 The parallel supercomputer

The Centre for Astrophysics and Supercomputing at the Swinburne University of Technology is home to a linux cluster of 212 Pentium class processors. There are 60 $\times$ 2.2 GHz and 30 $\times$ 2.0 GHz dual processor Intel Xeon machines as well as 16 $\times$ 933 MHz dual processor Intel Pentium III Xeon machines, all with a minimum of 1 GB of RAM.

The software correlator uses the 30 dual processor Xeon 2 GHz machines as they are connected via a high speed, low latency interconnect (Giganet) and are home to 60 GB Raid0 stripped disks on each machine. These drives allow for fast loading of data into the correlator. The interconnects allow for high data rates between interconnected nodes - allowing processes on different machines to communicate with each other at high speeds.

### 3.3.2 The software correlator code

The software correlator uses the *BasebandDSP* library to gain access to data recorded with the disk-based systems described in Chapter 2 or using the S2 tape-based recording system[4]. The correlator is also capable of using data recorded with the newly developed baseband recorders CPSR2 and Swinburne-MRO (see section 2.2.2). All these systems use real sampled data. At present the software correlator is only capable of correlating real sampled data.

The correlator loads all the required information from a correlator configuration file (see Appendix B) that is very similar in structure to the LBA correlator configuration file, with more advanced fields and extra options. The extra options allow for things such as separate clock rate offsets in each polarisation, definition of the number of lags to be generated, and the size of the data block to be passed

[4]Access to the raw S2 data is obtained using the S2-TCI interface (Wietfeldt *et al.* 1998).

to the correlator. In having a variable sized data block, significant increases in correlator speed can be made when small delay rates are present. This also allows a small degree of decorrelation to be maintained and reduces the output data volume.

The correlator uses the packed version of the raw data when transmitting data to the processing nodes.

Application of the delay between telescopes in the correlator is handled by reference to externally generated files that contain relative delay as a function of time, delays listed at one second intervals. Quadratic fits are made to these delay data to best determine delay offsets on sub-second intervals.

### 3.3.3 Multiple processor implementation

In order to process the data at as fast a rate as possible it is necessary to distribute the correlation work load over many processors. The software correlator code has been implemented to use many machines by using the Message Passing Interface (MPI) standard (`http://www-unix.mcs.anl.gov/mpi/`). The code employs a hierarchical distribution configuration. This configuration is such that the processing is performed by processor nodes, while baseline nodes distribute the data, with a head node maintaining the processor nodes status and collecting the processed data. This allows the data to flow, via the network, to all the computers with a reduced impact than if the same computer was sending and receiving all the data. See Figure 3.1 for a diagrammatic view of the data flow.

Upon execution of the correlator software, all the machines are initiated. One machine is designated as the head node. There can be more than one baseline node depending on how many baselines there are to be processed. The number of processor nodes is dependant upon the time required to process each data segment that is distributed. Any number of processor nodes (must be at least one) can be employed to process the data. The head node maintains a list of all the processor nodes. This list contains the state of the node. If the node is busy the job number the processor node is processing is also recorded so that the data can be reassembled in correct time order upon completion of the correlations. The head node allocates the processor nodes to the baseline node as they are requested. The processor nodes are allocated in a round robin fashion, with nodes being skipped if they remain busy. The head node also receives the data from the processor nodes, and stores these data to disk.

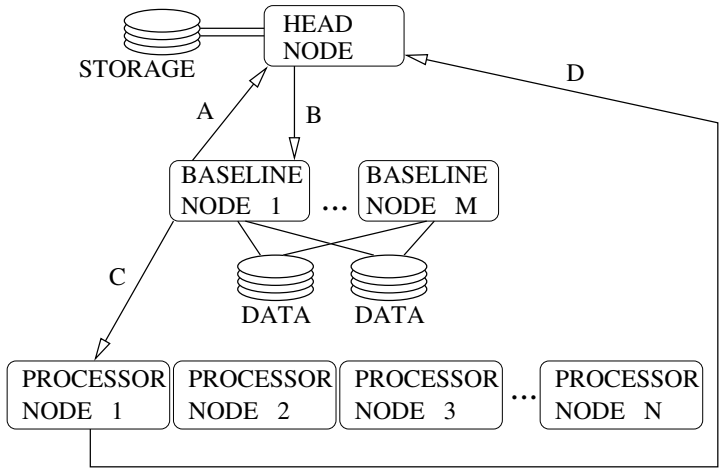


Figure 3.1: MPI node layout. Data is routed via A then B then C and finally D.

The baseline nodes load data and send it to the processor nodes. The data is loaded from the disk drives, which should be physically attached to the node, and not accessed via network file sharing (NFS). The data streams are aligned, and then sent to the processor node. Choice of processor node to distribute to is determined by asking the head node for a free processor, allowing for flexible allocation depending upon the cluster layout. Once a node is allocated, the distribution node sends the head node information about the job (timestamps the job number) and then forwards the data to the processor node. The distribution node continues to request new processor nodes until it either runs out of data, or there are no free processors. In the case of no free processors it simply waits a predetermined amount of time and then asks again.

The processor node accepts the data from a distribution node, unpacks the 2-bit binary data to 32-bit floating point numbers and then correlates the data. Post-processing methods are then applied to the data – FFTs are performed, fringe rotations applied, fractional sample errors are removed, and nominal amplitude calibration is applied. This increases the time taken to complete each correlation, but yields amplitude calibrated results.

For large numbers of baselines, it is possible to define the number of distribution nodes to be the number of baselines you need to process, or use a number smaller than the total number of baselines. In the later case the system will process the smaller number of baselines until each baseline completes. Once a distribution

node finishes processing a baseline it will continue onto the next available baseline. This process is handled automatically until all baselines are processed. The head node attempts to maintain a fair use policy of the processor nodes for each baseline by allocating the processors in a first come first served method.

### 3.3.4 Integration of data

Hardware correlators generally have long correlator accumulation times, of the order of 1-30 seconds. Because the software correlator performs all the correlations prior to fringe rotation it is necessary for this timescale to be short. These times are generally around 1ms. This vastly increases the amount of storage space required to store a correlated observation. As such, a further post-correlation process integrates the short timescale correlator output over much longer periods of time. This integration time is flexible and is mainly dictated by the amount of storage space required and the science requirements of the observation. The integration is a vector average of the complex correlator output.

### 3.3.5 Correlator performance

The software correlator is designed to run primarily on Intel based architectures, with Xeon processors being preferred due to the SSE2 extensions. Intel processors are used because Intel have released a performance library called Intel Integrated Performance Primitives (IPP). This library allows access to very fast correlation and Fourier transforms.

In a bid to discover further areas where the correlator could be improved, an extensive benchmarking session was performed. The results of this benchmarking are shown in Figure 3.2. This figure shows how the correlator performs as a function of both the number of frequency channels being processed as well as the number of processing nodes being used. All the tests use a single baseline of data. Many tests were conducted using 120 seconds of data from the source PKS B0826−373 (see section 3.4). If the correlator runtime for a test using these data is 120 seconds that means that processing occurs in real-time. The correlator appears to hit a limit where it doesn't perform any faster as more nodes are added. This is mostly due to the read and write speeds of the computer hard drives as well as the network interconnections. To test this limitation the correlator was run in a special

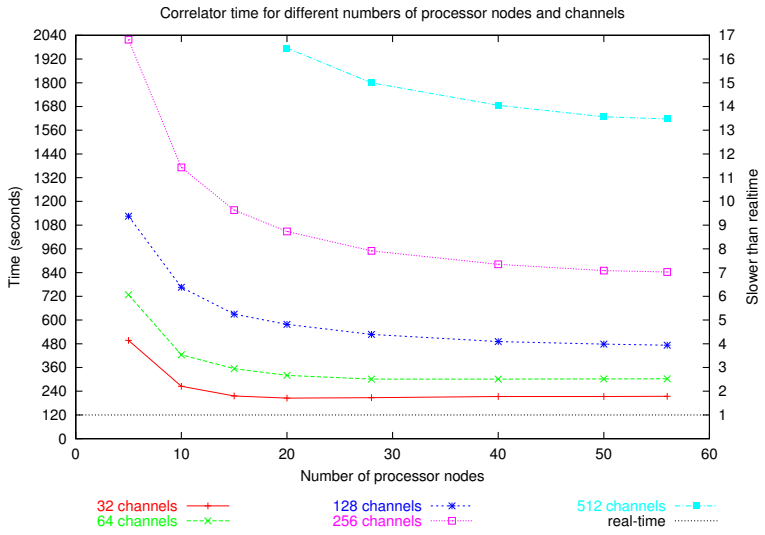


Figure 3.2: Correlator performance as a function of the number of processing nodes and the number of frequency channels being formed.

mode where the output data is not written to disk, while all of the other steps are performed as normal. This allowed a correlation time of 110 seconds for the 28 processor nodes test, at 32 channels, which is marginally faster than real-time.

Further tests were also carried out to see the effect of transferring large sections of data to the processor nodes. The nominal data length passed to the correlator corresponds to 1 ms of the observation ($120\text{s}/10^{-3} = 120,000$ data sections individually passed to the correlator), which was used for all tests in Figure 3.2. By increasing the size of the data length being processed (which reduces the total number of sections to be processed), there should be a reduction in the amount of time taken to correlate the data. This is mostly due to making larger reads of data from the hard disk, as well as making smaller writes to disks. Even though the input data size increases, the output data size remains the same – the overheads are being reduced. There is a trade-off to increasing the amount of input data which is to increase the decorrelation due to post-correlation fringe rotation (see Section 3.2.1). The result of increasing the amount of input data was to reduce the amount of time it took to process the full data set. This test was done without saving the output, using 64 frequency channels and 28 processor nodes for 1, 2, 5 and 10 ms of data (in these cases corresponding to a maximum of 2.5, 10, 50 and 100% decorrelation of the visibility amplitudes), which yielded run times of 290, 180, 100 and

90 seconds, respectively. This indicates that there is a non-linear increase in speed as the data size increases, but due to the extra loss of signal due to post-correlation fringe rotation there would be no need to use this mode if real-time performance is not needed.

The correlator has also shown a very stable increase in the amount of time taken to process the data for a given number of processing nodes, and increases in the number of channels being processed. Although not shown directly in Figure 3.2, it can be seen that there is a linear increase in time taken to process data as the number of channels being processed increases.

## 3.4 Example correlation: PKS B0826−373

PKS B0826−373 was observed as a fringe-check source (as described below), accompanying a longer (2.5 hour) observation of the Vela pulsar (see chapter 4). Since PKS B0826−373 is a strong and compact source, it is ideal for examining test correlations using the software correlator and for verifying the performance of the software correlator by comparing the amplitude and phase outputs of the correlator (as functions of time and frequency) to those from an established hardware correlator.

PKS B0826−373 is a relatively obscure radio source within approximately 10 degrees of the plane of our galaxy and with no optical counterpart yet identified (Costa & Loyola 1999). The radio source is part of the International Celestial Reference Frame and has been observed numerous times for the purposes of VLBI astrometry (Russell *et al.* 1994; Reynolds *et al.* 1994; Johnston *et al.* 1995; Ma *et al.* 1998). VLBI observations at 22 GHz, as a pre-cursor to the VSOP space VLBI mission, showed PKS B0826−373 to have approximately 0.26 Jy on intercontinental baselines, giving a lower limit to its brightness temperature of $1.8 \times 10^{10}$ K (Moellenbrock *et al.* 1996) at this frequency.

### 3.4.1 The observations and data recording

PKS B0826−373 was observed for approximately 2 minutes on 2001 March 25 with two antennas of the Australian LBA, the 70 m antenna of the Deep Space Network at Tidbinbilla, near Canberra, and the 64 m antenna of the Australia Tele-

scope National Facility (ATNF) near Parkes. The baseline length for Parkes to Tidbinbilla is approximately 275km. Data were recorded using the S2 VLBI recording system (Cannon *et al*. 1997) and consisted of 2-bit samples at 32 million samples per second for each of two 16 MHz bands, giving an aggregate data rate of 128 Mbps. One 16 MHz band was configured to record left circular polarisation and the other 16 MHz band was configured to record right circular polarisation. The centre frequency for each band was 2290 MHz. The effective local oscillator frequency corresponds to the lower edge of the frequency band, 2282 MHz.

### 3.4.2 Correlation of data with the ATNF correlator

The ATNF LBA correlator is an XF hardware correlator which uses the correlator chips designed for the ATCA correlator (Wilson, Roberts & Davis 1996; Tzioumis 1997) and interfaces to units that playback data that were recorded using S2 recorders at each antenna. The LBA correlator has been in regular use since 1994 for correlating VLBI data from Australian and other antennas and has participated in many large projects over the last ten years.

The PKS B0826−373 data were correlated on the LBA correlator using a single baseline, dual-frequency correlator configuration which utilised 64 frequency channels across each of the two 16 MHz observing bands. At the correlator the fringes were located manually and a model for the clocks at the two stations was added to the correlator configuration. An approximate 0.1 $\mu$s delay offset was found between the right and left polarisations. Since the LBA correlator uses a single clock model for both polarisations, the clock offset determined for the model was 2.38 $\mu$s, approximately 50 ns in error for both left and right polarisations, leading to substantial phase slopes as a function of frequency in the correlated data. At the time that the data were correlated, no linear clock drift rate was determined for inclusion in the clock model.

The correlated data were written to disk in the RPFITS data format, for further analysis and comparison to the correlated data from the software correlator.

### 3.4.3 Correlation of data with the software correlator

The software correlator has been described above. To correlate the PKS B0826−373 data a correlator accumulation time of 1 ms was used and a correlator configura-

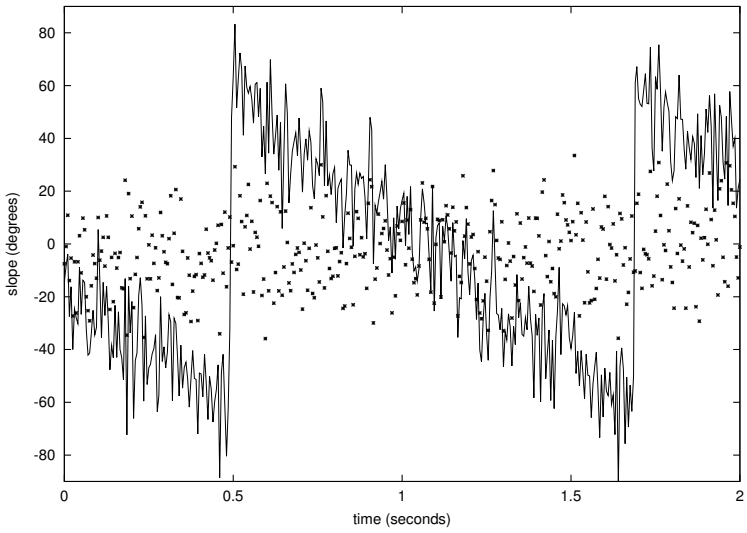


Figure 3.3: Example of fractional sample removal. The line shows the fit to the slope across the band after Fourier transformation. The dotted points show the fit to the band after removal of the fractional sample error. The data is noisy due to the short (1 ms) integration period used. Integration after correlation reduces the noise.

tion that utilised 64 frequency channels across each 16 MHz observing band. The fringes were located manually and a model for the clocks was determined, in a similar fashion to the LBA correlator as described above. An important difference to the LBA correlator is that different clock models can be applied to the different polarisations, allowing us to track the fringes close to zero delay, rather than incurring a 50 ns delay error. The clock offset between the two telescopes was found to be 2.338 $\mu$s for LCP and 2.431 $\mu$s for RCP. A linear clock drift rate was also included in the clock model, $3.639 \times 10^{-7} \mu\mathrm{s/s}$. A 1 ms correlator accumulation time, for the delay rates and frequencies experienced during this observation, produces approximately 2.5% amplitude loss due to decorrelation of the signal because of post-correlation fringe rotation. This loss in amplitude was corrected in post-correlation processing.

Figure 3.3 shows the post-correlation removal of the fractional sample error over a 2 second sub-section of the PKS B0826−373 data. Figure 3.4 shows the post-correlation removal of the fringe rate from the same 2 second sub-section of PKS B0826−373 data.

The correlation of the PKS B0826−373 data used a supercomputer configura-

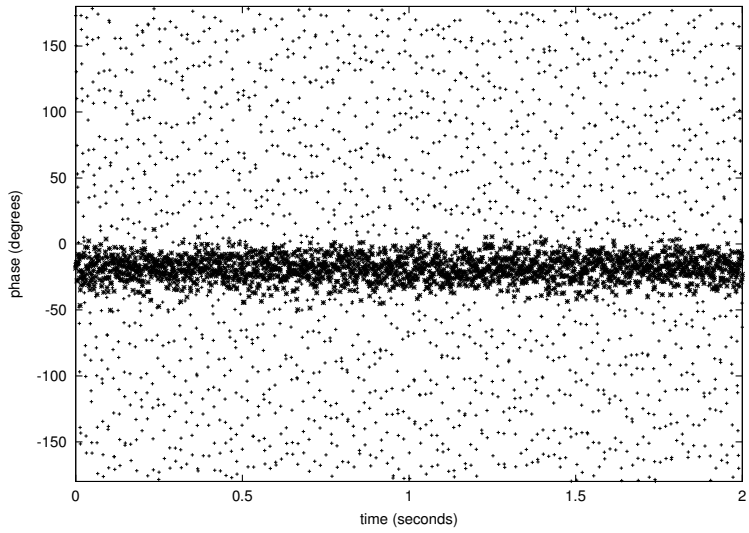


Figure 3.4: Example of post-correlation fringe rotation. The band of points at the centre of the plot show the data after it has been fringe rotated. The small points that are everywhere else are the data values after correlation, but before fringe rotation.

tion consisting of one head node, one baseline node and 56 processor nodes (a total of 58 processors). The correlation of 2 minutes of data, with 64 frequency channels, required approximately 300 seconds of time on the supercomputer, a factor of 2.5 times slower than real-time correlation.

### 3.4.4 Correlator comparison

Figure 3.5 shows the amplitude and phase measured across the full 16 MHz observed band in the left circular polarisation for PKS B0826−373. The asterisks denote amplitude and phase output from the LBA correlator for a single two second correlator integration time (at time = 08:13:35 UT on the 25 of March 2001), using 64 frequency channels across the 16 MHz band. The amplitudes and phases produced by the LBA correlator were extracted by loading the correlated data into the MIRIAD software package (`http://www.atnf.csiro.au/computing/software/miriad/`) and using the task UVSPEC to extract the data from a single 2 second correlator integration period. Nominal system equivalent flux densities (in Jy) have been applied at the LBA correlator, along with other corrections, to calibrate the correlator output into flux density units (Jy). The system equivalent

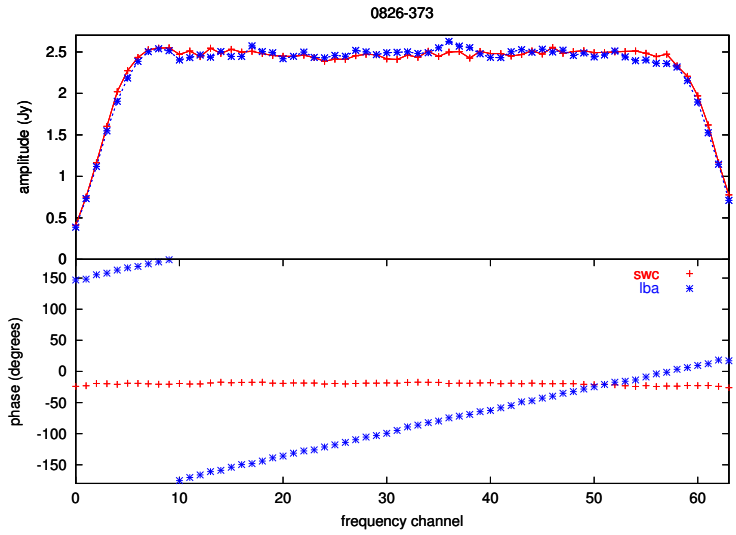


Figure 3.5: Bandpass comparison of software correlator and LBA correlator using LCP. The top panel shows the amplitude of the signal as a function of frequency. The red points are from the software correlator, and the blue points from the LBA correlator. The two correlators show amplitudes that agree to within the uncertainties. The bottom panel shows the phase of the signal as a function of frequency. The LBA correlator has a rotating phase across the band due to a delay offset between the LCP and RCP signals. This offset meant that the data were correlated using a delay half-way between the delays for the two polarisations (there is approximately 100ns of delay between the polarisations).

flux densities applied in this case are: Tidbinbilla, 20 Jy; Parkes, 25 Jy. A large phase slope across the band is apparent because the data were correlated with a delay error of approximately 50 ns, as mentioned above.

The crosses in Figure 3.5 denote the amplitudes and phases produced from the software correlator, for the same two second period of data, and again for 64 frequency channels across the band. In this case the data were correlated using a 1 ms correlator integration time, corrected for fractional sample error and fringe-rotation (as described in section 3.2.1), and integrated over the two second interval (as described in section 3.3.4). The correlator output was calibrated into flux density units as described in section 3.2.3, using the same system equivalent flux densities as used above to calibrate the data from the LBA correlator.

It is immediately apparent that the overall agreement between the amplitude measurements from the LBA correlator and the software correlator is very good. In

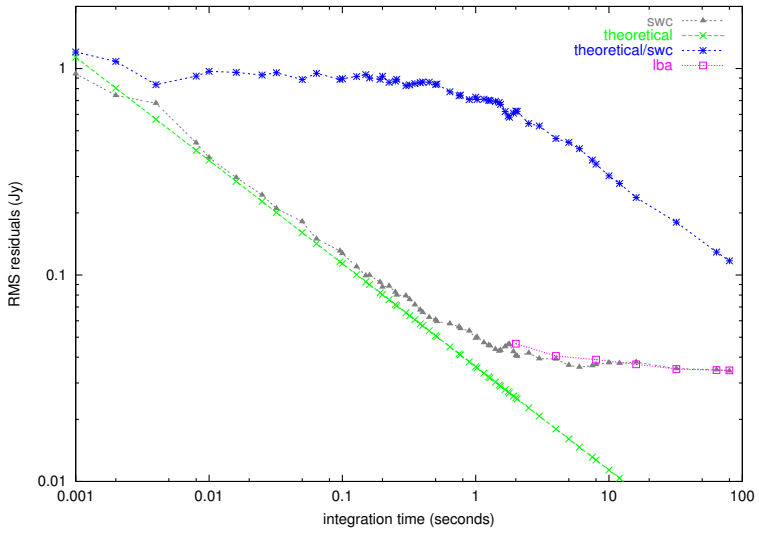


Figure 3.6: RMS residuals from LBA and software correlator for given integration times. The green points show the theoretical noise estimate for both correlators. The pink points show the measured noise for the LBA correlator (starting at the smallest integration length of 2 seconds). The grey points are for the software correlator, which was used with a minimum integration length of 1 ms. The blue points show the theoretical noise divided by the measured noise from the software correlator.

particular, a linear, zero slope fit to the amplitudes across the central 68.75% of the band gives a mean flux density of 2.47 Jy for the data from the software correlator, with an RMS deviation around this mean of 0.04 Jy. In comparison, a similar fit for the LBA data gives a mean flux density of 2.48 Jy and an RMS deviation around the mean of 0.05 Jy. The mean (scalar averaged) flux densities from the two correlators agree to within these uncertainties. The RMS deviations also agree very well. The theoretical RMS deviation around the mean amplitude due to thermal noise is calculated to be approximately 0.03 Jy, based on a 2 second integration in a 250 kHz frequency channel of the 16 MHz band, and using the nominal system equivalent flux densities given above, for 2-bit sampled data (Crane & Napier 1989). Thus, the measured RMS noise is approximately a factor of 1.5 higher than the theoretical value, using nominal calibration parameters for these antennas.

However, a deeper investigation of the measured RMS noise relative to the theoretically expected RMS noise shows better agreement at shorter integration times. Figure 3.6 shows the theoretical RMS noise for integration times between 1 ms and

100 s (green diagonal crosses; using calibration parameters and frequency ranges as used above), compared to the measured RMS noise from the software correlator (grey filled triangles) and the measured RMS noise from the LBA correlator (pink open squares). The measured values are all formed using vector averages over the given integration times. Also shown is the ratio of theoretical to measured RMS noise for the software correlator measurements (blue asterisks). Agreement between the software correlator RMS noise and the theoretically expected values is very good over the integration period range of 1 ms to 0.5 s. At an integration time of approximately one second the measured RMS noise flattens to a constant value, due to the existence of un-modeled structure in the bandpasses. The RMS noise measured using a zero slope fit at the LBA correlator over the range in integration times from 2s to 100s agrees closely with the data from the software correlator and both measurements flatten to the same value. A standard bandpass calibration using the measured data may allow improvements in the RMS noise with integration times longer than one second.

Finally, Figure 3.7 shows a comparison of the data from the LBA correlator and the software correlator as a function of time. Measurements of the amplitude using a two second integration time for the LBA correlator, and a one second integration time for the software correlator, as discussed above, for each of the correlators is shown in the top panel. A linear, zero-slope fit to these data gives a mean flux density of 2.499 Jy and RMS of 0.006 Jy for the software correlator and similarly 2.481 Jy and 0.010 Jy for the LBA correlator. The bottom panel shows the phases as a function of time (with and without a clock drift rate present) from the software correlator and also the phases determined from the LBA correlated data by fringe-fitting the data using the AIPS task FRING. The phases generated by the software correlator and the LBA correlator also agree to within the uncertainties.

## 3.5 Conclusions

A software correlator well suited to observations of continuum radio sources with the LBA has been developed on a parallel supercomputer. The correlator implements a relatively straight-forward scheme whereby corrections for fringe-rotation and fractional sample errors are made as part of post-correlation processing. The trade-off in this approach is that the timescale on which correlations are formed

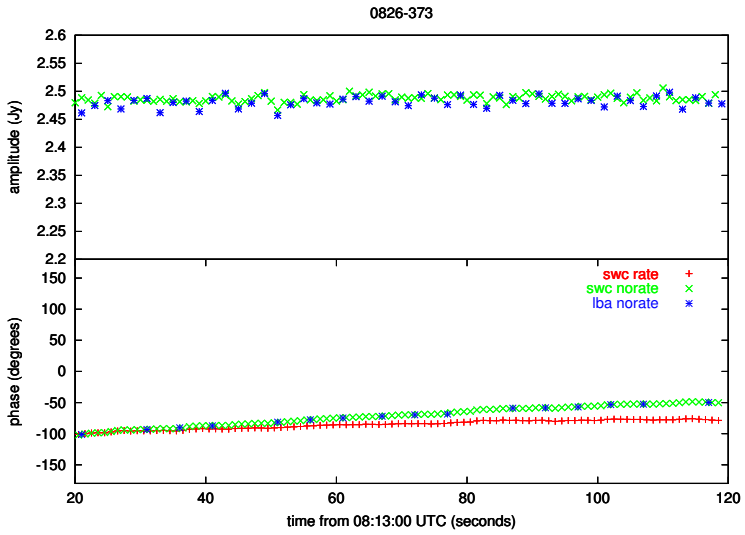


Figure 3.7: Software correlator and LBA correlator comparison for 100 seconds of data at LCP. The blue points indicate from the LBA correlator, and the green points indicate from the software correlator. Both correlators agree to within the uncertainties in amplitudes. The red points in phase are for the software correlator with a clock drift rate present, which reduces the rate of change of phase with time.The green points are the software correlator with no clock drift rate. The blue phase points are the fringe-fitted phases from the LBA correlator. The data is offset from 08:13:00 UTC.

must be kept short. While the approach is sub-optimal for narrow bandwidth, high frequency resolution observations at high frequency, the approach works very well in the situations for which it was designed, wide bandwidth continuum observations.

Performance of the correlator has been verified via comparison with an established hardware correlator. The software correlator illustrates the fact that different correlation schemes can be implemented in software for the purposes of testing. The correlator described here can be considered a prototype and future extensions will include pre-correlation fringe rotation, development of an FX style correlator, and an improved data distribution scheme.

# Chapter 4

# Applications of the baseband recorders and software correlator

## 4.1 Introduction

The two preceding chapters dealt with descriptions of the hardware and software developed for recording and correlating baseband data from radio telescopes for astronomical applications. In this chapter, some of these applications are explored, illustrated by observations made at three radio telescopes, the Parkes 64 m "Dish", the Australia Telescope Compact Array (ATCA), near Narrabri, NSW, and the NASA Deep Space Network (DSN) 70 m telescope at Tidbinbilla (DSS43).

Single dish applications, such as the measurement of source spectra, the monitoring of radio frequency interference (RFI), and the timing of pulsars are investigated. Also explored are applications in VLBI.

In the sections below, the concepts behind the observations are briefly explained and the observational data presented. In some cases, these observations are being used in scientific investigations, in collaboration with other investigators. In particular, the analysis of the OH maser G345.003−0.224 which is a collaborative project with Drs Frank Stootman and Miroslav Filipovic of the University of Western Sydney and Drs Shinji Horiuchi and Steven Tingay of the Swinburne University of Technology. Also the analysis of the Vela pulsar emission region is being done in collaboration with Drs Steven Tingay and Stephen Ord, and Prof. Matthew Bailes of the Swinburne University of Technology and Prof. Jim Cordes of Cornell University.

## 4.2 Single dish applications

As mentioned in Chapter 1, single dish telescopes have been used in this thesis to study masers, pulsars and RFI. The following sections briefly describe studies of masers and RFI made with the ATCA and pulsar observations from both the ATCA and the Parkes radio telescopes. The observations from the ATCA are in the tied-array mode, therefore considered a "single dish" in this case.

### 4.2.1 Masers

The term maser is the acronym for microwave amplified stimulated emission of radiation. Masers are the naturally occurring microwave equivalents of lasers (light amplified stimulated emission of radiation) and are found in molecular clouds, the emission coming from compact regions of the clouds. Some of the well-known and well-studied masers are water ($H_20$), hydroxyl (OH), silicon monoxide (SiO), and methanol ($CH_3OH$). Masers produce emission at specific frequencies resulting in strong spectral lines.

#### 4.2.1.1 G345.003−0.224

One maser that has been observed and considered in some detail during this thesis, and is being studied in an on-going program, is an OH maser at 1720 MHz in the star-forming region G345.003−0.224 (Caswell 1999). This maser was observed by Dr Frank Stootman using the University of Western Sydney's high resolution spectrometer at the Parkes radio telescope in January 2003 and found to have a simple near-Gaussian spectral profile in the left circular polarisation (LCP) and no detectable emission in right circular polarisation (RCP). OH masers usually exhibit circular polarisation and the fact that no RCP was found indicates that this source may be a single maser feature that is 100% circularly polarised.

In the classical model of masers, at high saturation, the spectral profile of a single maser feature reflects the Gaussian velocity distribution of the masing molecules (Emmering & Watson 1994). However, recently it has been recognised that the exact shape of maser spectral profiles, in particular deviations from a Gaussian profile, may be used to measure the maser amplification (Watson, Sarma & Singleton 2002; Moscadelli *et al.* 2003; Watson & Wyld 2003).

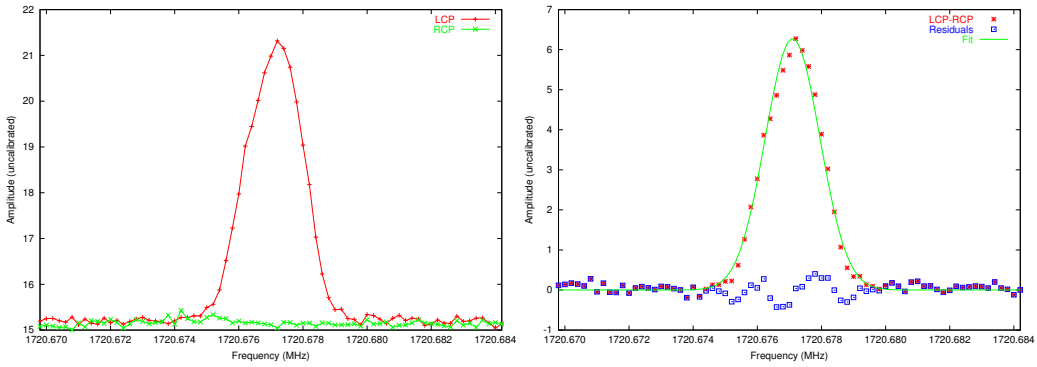


Figure 4.1: OH maser at ~1720 MHz in star forming region 345.003−0.224. The left panel shows the left and right circular polarisation emission for the maser. The right panel shows the fit to the difference of the polarisations. The green line is the fit (to the red data), with the blue squares showing the residuals to the fits.

The possibility that the maser observed by Frank Stootman is a single feature makes it potentially interesting for tests of the models of Watson, Sarma & Singleton (2002). A potential problem with the observations of Stootman is that the Parkes radio telescope, at a frequency of 1720 MHz, has a FWHM beam size of approximately 13 arc minutes, representing a massive area of the sky compared to the area of the single maser spot (probably far less than one arcsecond in extent). It is possible that chance maser emission at the same frequency, but well separated from the source of interest on the sky, may complicate any analysis of the spectral profile. That is, this may be a confused region of the sky at this frequency and with this telescope.

The Swinburne-MRO recorders described in Chapter 2 were used at the ATCA to record high angular resolution data at baseband. The telescope was operated in the tied array mode, which simply coherently adds the signals from all 6 antennae. This is equivalent to using a single telescope (albeit with a strange primary beam shape) with a 6 km diameter but the sensitivity of a 54 m diameter dish. The angular resolution of this setup is approximately 7 arc seconds, a vast improvement in angular resolution over the 13 arc minute Parkes beam. Using the ATCA in the tied array mode, it was possible to isolate the maser emission from the one star-forming region of interest.

The maser was thus observed for 30 minutes using a 16 MHz band at LCP and a 16 MHz band at RCP over the same frequency range. Spectra of the maser

were formed in software, as described in Chapter 2. The left panel in Figure 4.1 shows the LCP and RCP spectra derived from these observations, using 200 Hz frequency channels (equivalent to a velocity resolution of 0.034 km/s). As can be seen there is no significant detection of the maser in the RCP, showing the maser to be only left circularly polarised. The data have been fitted to a pure Gaussian profile model, the model and the residuals from the fit are show in the right panel of Figure 4.1. An interesting frequency structure can be seen in the residuals, which will bear more investigation and further observation at higher spectral resolution. A similar residual frequency structure was found in the Stootman data from the Parkes radio telescope.

In order to further study this source, proposals have been submitted to the ATCA and the Australian LBA for observations in the observing semester for May 2004. These observations have multiple aims:

1. The ATCA observations will be used to image the field around the maser at 1720 MHz, to confirm that no confusing emission exists at the exact frequency of the maser;

2. During the ATCA observations, the tied array mode will be running in parallel with the imaging observation, allowing the baseband data to be recorded using a Swinburne-MRO recorder. The final spectra should show 3 times the signal to noise and 4 times the spectral resolution of the spectra shown in Figure 4.1 for a 12 hour observation;

3. The LBA observations will be used to measure the high angular resolution structure of the maser and identify any positional offsets as a function of frequency, which would indicate that the maser is not a pure single maser.

All of these measurements are important for testing and evaluating the models of Watson, Sarma & Singleton (2002) and will be reported in future publications.

#### 4.2.1.2 Other maser observations

Further maser observations using the Swinburne-MRO recorders were undertaken near the VLBI observations of November 2003 (see Chapter 2). A list of the observations undertaken can be found in Table 4.1 along with the recorded profiles of some water masers in Figure 4.2 and a Hydroxyl and Methanol maser in Figure

Table 4.1: List of maser observations undertaken for testing of Swinburne-MRO recorders

| Source Name | Frequency | Duration | Channels | Channel size |
|---|---|---|---|---|
| Water Masers | | | | |
| ORI-KL | 22.2 GHz | 5 mins | 8,000 | 2 kHz |
| ORI-S | 22.2 GHz | 5 mins | 3,200 | 5 kHz |
| MRC0513-694b | 22.2 GHz | 10 mins | - | - |
| MC24 | 22.2 GHz | 10 mins | - | - |
| MON-R2 | 22.2 GHz | 5 mins | - | - |
| Hydroxyl Masers | | | | |
| G345.003−0.224 | 1.7 GHz | 30 mins | 80,000 | 200 Hz |
| SGRA | 1.7 GHz | 5 mins | - | - |
| G349.7−0.2 | 1.7 GHz | 10 mins | - | - |
| Methanol Masers | | | | |
| G213.71−16.60 | 6.6 GHz | 2 min | 2000 | 2 kHz* |
| G232.62+0.99 | 6.6 GHz | 2 min | 2000 | 2 kHz* |

Dashed entries ( - ) indicate that the data have not been fully processed

* These observations were taken using a 4 MHz bandwidth,
as opposed to the 16 MHz bandwidth for the rest of the observations.

4.3. The observations are short as they are purely for testing the Swinburne-MRO recorders. As such, those with good signal-to-noise ratios (SNR) have been plotted. These data have not been amplitude or bandpass calibrated, and therefore will not be compared in this thesis to the existing profiles from previous observations. However, simple comparisons were made to ensure the spectral lines appeared in the same frequency ranges as seen in other published results for the same sources.

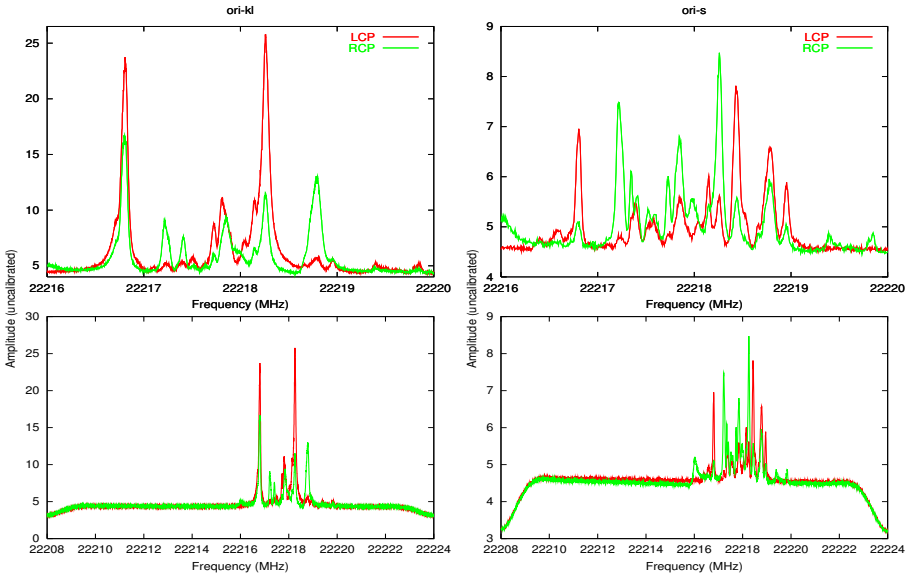


Figure 4.2: 22.2 GHz water masers ORI-KL and ORI-S. The upper panels show an expanded section of each band, with the lower panels showing the complete 16 MHz band. ORI-KL is on the left, ORI-S is on the right.

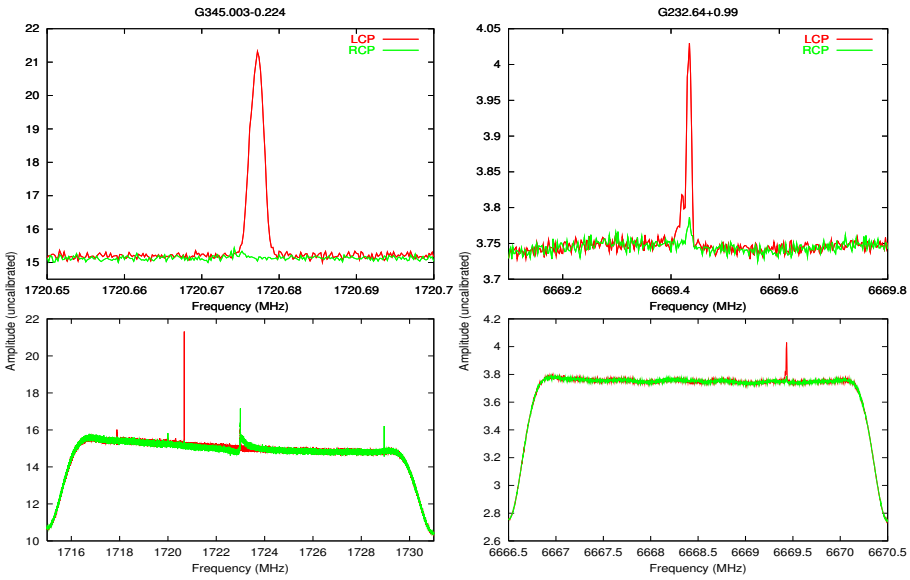


Figure 4.3: 1720 MHz hydroxyl maser G345.003−0.224 and 6668 MHz methanol maser G232.62+0.99. The upper panels show an expanded section of each band, with the lower panels showing the complete bandwidth. The left panels show G345.003−0.224, observed over a 16 MHz bandwidth. The right panels show G232.62+0.99, observed over a 4 MHz bandwidth.

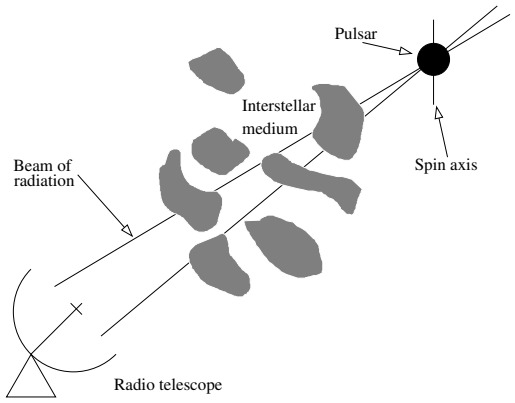


Figure 4.4: Schematic diagram of a pulsar beam intersecting the line of sight of a radio telescope.

### 4.2.2 Pulsars

Pulsars were first discovered at radio wavelengths in 1967 (Hewish *et al.* 1968) and have been widely accepted as rapidly spinning, magnetic neutron stars (Gold 1968). The rotation periods of these stars are very stable and are found to be between 1.5ms and 8 seconds in duration[1] (Taylor, Manchester & Lyne 1993). As the star rotates, and if the star/observer geometry is suitable, beams of emission intersect the observer's line of sight and the observer sees regular pulses of emission from the star (see Figure 4.4). Pulsars can be observed at X-ray, optical and gamma-ray wavelengths (Romani 2003) and are believed to be the result of supernovae, in which stellar cores collapse into extremely dense neutron stars. A pulsar can also get a kick from this explosion, which propels it through space away from the site of the supernova (Kohri & Nagataki 2000; Burrows & Hayes 1996). The transverse speeds for these pulsars can range from 0 to >1000 km/s (Harrison, Lyne & Anderson 1993; Brisken *et al.* 2003). With the high angular resolution available with VLBI arrays it is possible, over many years, to see the movement of the pulsar with respect to the other objects around it (for example Dodson *et al.* 2003; Bailes *et al.* 1990).

Pulsars are regularly timed by many radio astronomy groups around the world. The discovery of relativistic binary pulsars (Hulse and Taylor were awarded the

[1] A catalog of pulsars and the associated parameters can be found at `http://www.atnf.csiro.au/research/pulsar/psrcat/`

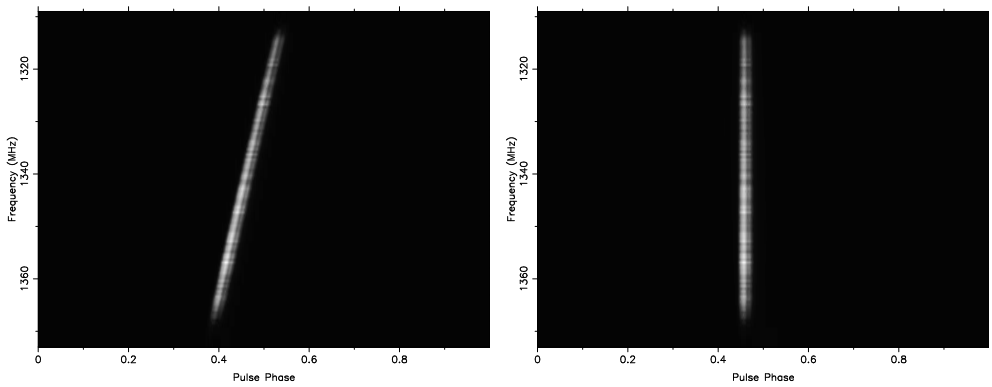


Figure 4.5: Dispersion and dedispersion of the Vela pulsar. The figures show a grey scale of the pulse intensity as a function of both observing frequency and pulse phase. The left panel shows the Vela pulsar signal as it received by the telescope, dispersed. The right panel shows the result of modeling and removing the dispersion, a pulse profile unaffected by the dispersive smearing effect.

Noble prize for the discovery PSR B1913+16; Hulse & Taylor 1974) and accurate timing of these pulsars have lead to tests of general relativity (Taylor & Weisberg 1982; van Straten *et al*. 2001). Pulsar timing is performed using high time resolution recording instruments such as the second Caltech Parkes Swinburne Recorder (CPSR2 - see section 2.2.2).

The signals that arrive at the radio telescope from the pulsar (and all other radio sources) have not only traveled vast distances, but have been slightly modified in their travels. The interstellar medium (ISM) that the signal travels through is ionised, which causes the signal to be dispersed by the time it reaches the radio telescope. The overall effect is that the lower frequencies arrive some time after the higher frequencies. Figure 4.5 shows the Vela pulsar signal as it is received at the telescope, the curve showing the dispersion that is present. Also shown in Figure 4.5 is the result of dedispersion, which is the removal of the dispersion. Each pulsar has a different dispersion measure (DM) that is related to the column density of free electrons along the line of sight to the pulsar. The ISM is not uniform in nature, but rather lumpy as depicted in Figure 4.4. Different areas of the galaxy have larger and smaller electron densities, for example the arms of a spiral galaxy are denser than the areas in between the arms. Cordes & Lazio (2002) have published an updated model of the electron densities for the Milky way galaxy.

Table 4.2: Observations of pulsars for comparisons of baseband recorders.

| Date | Pulsar | Recorder | Duration | Frequency | Telescope |
|---|---|---|---|---|---|
| Mar-2001 | J0835−4510 | S2 | 2.5 hrs | 2.3 GHz | Parkes, Tidbinbilla |
| Jan-2003 | J0835−4510 | CPSR2 | 10 mins | 1.4 GHz | Parkes |
| Nov-2003 | J0835−4510 | Swinburne-MRO | 30 mins | 2.3 GHz | Parkes, ATCA |
| Nov-2003 | J0835−4510 | Swinburne-MRO | 2.2 hrs | 4.8 GHz | Parkes, ATCA |
| Nov-2003 | J0437−4715 | Swinburne-MRO | 10 mins | 1.4 GHz | ATCA |
| Nov-2003 | J1141−6545 | Swinburne-MRO | 10 mins | 1.4 GHz | ATCA |

Much work has been done to develop methods for dedispersion (Hankins 1971). The *BasebandDSP* library has implemented a coherent dedispersion function for use in the processing of pulsar data. The Swinburne software correlator has a pulsar processing mode (see Appendix B) which allows for the study of on and off pulse phase of the pulsar period, as well as the ability to dedisperse the data before correlation. This is a method that is not generally implemented at correlators, and allows for a much better understanding of the pulsar from VLBI observations. An example of how this can be used is shown in section 4.3.3.

Three pulsars have been observed with the Swinburne-MRO recorders. They are the Vela pulsar (PSR J0835−4510), the well-known pulsar PSR J0437−4715, and PSR J1141−6545. The reason for observing the Vela pulsar is that it is the brightest pulsar from Earth. This is due to the Vela pulsar being relatively close (7th closest of the known pulsars), at a distance of approximately 287 parsecs (Dodson *et al*. 2003). The brightness of the Vela pulsar makes it useful for testing the baseband recorders, as it can be easily located in the raw data with high signal to noise. J0437−4715 was chosen as it is observed regularly by the Swinburne pulsar group, and because it is a fast pulsar (5.8 ms period), which allows for testing the time stability of the Swinburne-MRO recorders. The third pulsar, J1141−6545, was chosen because it is slower than both J0835−4510 and J0437−4715 (393.9 ms period). Table 4.2 shows a summary of the observations that have been made of pulsars as part of this thesis. Comparisons of the pulsar profiles will be made where possible. The observations from the CPSR2 recorders were made during pulsar timing observations in collaboration with Prof. Matthew Bailes, Dr Stephen Ord, Mr Aidan Hotan and Mr Haydon Knight. The CPSR2 observations of the Vela pulsar cannot be directly compared with the S2 or Swinburne-MRO observations

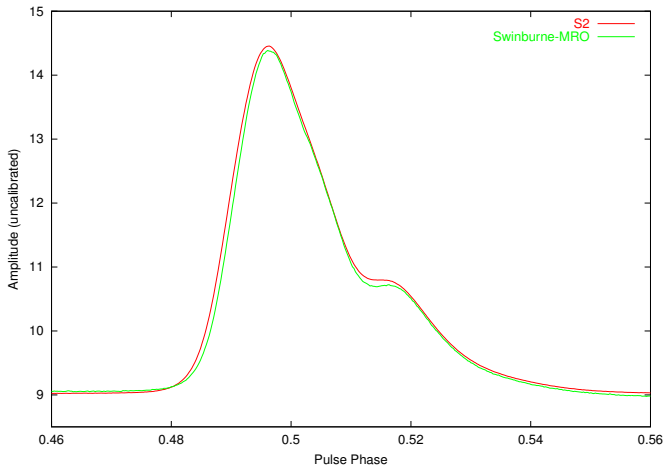


Figure 4.6: Vela pulsar profile comparison of S2 and Swinburne-MRO baseband recorders as a function of pulse phase (over a limited range). The S2 profile was generated using 2 x16 MHz bands, an integration of 2.5 hours and a centre frequency of 2290 MHz. The Swinburne-MRO profile was also generated with 2 x 16 MHz bands, but a lesser integration of 0.5 hours and a centre frequency of 2268 MHz. Both profiles are dedispersed.

due to the different frequencies of the observations. Pulsars are known to have different pulse profiles at different frequencies (Karastergiou *et al.* 2003), so direct comparisons require data that are recorded at the same frequency.

The observations that were made of Vela with the S2 recorders are compared to those made by the Swinburne-MRO recorders in Figures 4.6 and 4.7. The S2 observations were taken during March 2001. The profiles of the pulsars are shown as a function of pulse phase. The pulse phase is a value between 0 and 1 that represents the position in the pulsar period (fraction of pulsar period), independently of what the period actually is. The Vela pulsar produces a favourable comparison when shown over the on pulse region (Figure 4.6). There are some slight differences in the profile which can be attributed to the difference in observing frequencies between the two observations and the offset baseline (as mentioned below). The S2 data were taken with a centre frequency of 2290 MHz, and the Swinburne-MRO used 2268 MHz. The frequencies are different as the Swinburne-MRO observations were taken during a pulsar timing experiment which was configured for a different frequency due to different processing requirments.

The SNR for these observations could not be used as a direct comparison due

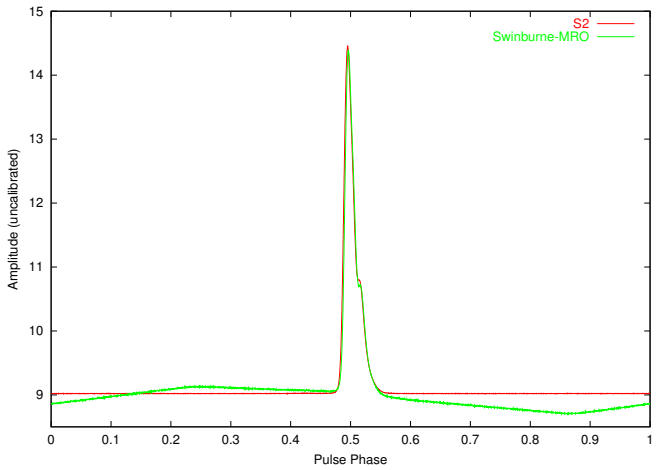


Figure 4.7: Vela pulsar profile comparison of S2 and Swinburne-MRO baseband recorders as a function of pulse phase (over the full pulse phase range). The S2 profile was generated using 2 x 16 MHz bands, an integration of 2.5 hours and a centre frequency of 2290 MHz. The Swinburne-MRO profile was also generated with 2 x16 MHz bands, but a lesser integration of 0.5 hours and a centre frequency of 2268 MHz. Both profiles are dedispersed.

to a calibration tone being present during the recording of the Swinburne-MRO data. These calibration tones are used during normal VLBI observations to allow for extraction of the system temperatures of the radio telescope. Figure 4.7 shows the overall effect of the calibration tone (which is a square wave). Due to the calibration tone being set to an on/off cycle frequency close to a harmonic of the Vela pulsar (approximately 11 Hz), the square wave is not added perfectly, and produces a corrupted baseline signal as a function of pulse phase. This corrupt baseline is formed by taking the square wave and shifting it slightly in one direction, then adding it to itself, and repeating this step. This baseline makes it difficult for the analysis software to determine the SNR. The corrupt baseline explains the majority of the difference seen between the Swinburne-MRO and S2 results shown in Figure 4.6. Indeed, the removal of the baseline resulted in the difference between the profiles being very small, and attributed to the difference in the observed frequencies. Simultaneous tests at the same frequency with the CPSR2 and Swinburne-MRO recorders would provide a better comparison (without the calibration tone present!).

Due to the short integration times for both J0437−4715 and J1141−6545 the

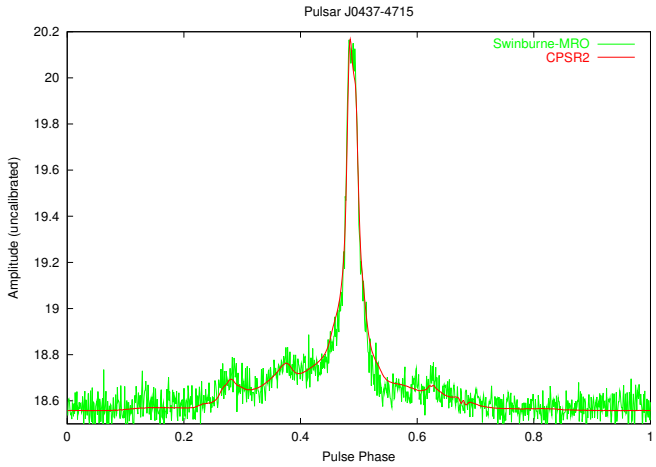


Figure 4.8: Comparision of pulsar J0437−4715 from Swinburne-MRO and CPSR2 recorders. The CPSR2 data were taken from the Parkes radio telescope with a centre frequency 1405 MHz, a bandwidth of 64 MHz and an integration time of 3.5 hours. The Swinburne-MRO data were taken at the ATCA with a centre frequency of 1384 MHz, a bandwidth of 16 MHz and an integration time of 10 minutes. Both profiles are dedispersed.

SNRs are relatively low. Figure 4.8 shows the comparison of J0437−4715 between the CPSR2 recorder at Parkes, with a bandwidth of 64 MHz and 3.5 hours of integrated data, and the Swinburne-MRO recorders at the ATCA, with a 16 MHz bandwidth and 10 minutes of integrated data. The comparison shows a very good result, as the Swinburne-MRO recorder data is visibly similar to the CPSR2 data, albeit nosier as expected due to the reduced integration length and bandwidths as well as the smaller collecting area of the ATCA telescope. J1141−6545 has not been shown due to the signal not being strong enough for a meaningful comparison.

### 4.2.3 Radio frequency interference

Human-made radio frequency interference (RFI) is often a problem when observing with radio telescopes. Not only do radio telescopes receive signals from a cosmic origin, but the telescopes themselves generate signals (from electronic components that make up the telescopes, power supplies, etc). Much of this self-interference is understood, and can be minimised through careful design. One of the largest generators of noise are power supplies. In Australia the normal mains

power supply is 240V at 50Hz. The 50Hz signal generates significant radio noise. Also, modern computers now run at clock frequencies that are in the range of popular radio astronomy wavebands, however it is generally the bus speeds of the computers that cause problems in the harmonics (Brett Presig - private communication). Careful shielding of computer equipment can minimise the effects of these interfering sources.

While the RFI generated at the telescope can usually be understood, it is sources away from the telescope and outside of the control of the observatory that cause the most problems. Television transmitters and mobile phones are some of the worst offenders. Most observatories ask visitors to turn off mobile phones when on site. Not only are there terrestial-based RFI generators, but there are transmitting satellites that orbit the Earth that also have to be dealt with. They pass in and out of the telescope beam and can cause a large amount of RFI - usually making the affected section of data useless. The Australian Communications Authority (ACA) acknowledges that radio telescopes make observations in the frequency ranges[2] of 1250-1780 MHz, 2200-2550 MHz, 4350-6700 MHz, 8000-9200 MHz and 16000-26000 MHz. The ACA also makes a note of observations at 440 MHz, 660 MHz, 840-845 MHz and 12-15 GHz, but the radio astronomy community does not have exclusive use of these frequency ranges.

As we build more and more data transmitters and RFI generators, we need to keep a constant eye on what RFI is present in the local area. Digital television transmitters are an example of a new source of RFI at Australian observatories.

Disk-based baseband recorders are extremely useful for monitoring and analysing RFI because they allow high time and frequency resolution observations of the interference. From an analysis of monitoring measurements, methods for actively canceling RFI can be investigated (Leshem, van der Veen & Boonstra 2000; Bell *et al*. 2001). Traditional methods of RFI mitigation used at telescopes usually involve placing narrow-band filters around the frequency range of the RFI to passively cancel the interference. This leads to data not being recorded in these narrow bands, which is a problem if this is where specific astronomy signals are expected. By actively canceling RFI it is possible to still record data in the entire frequency range of interest.

[2]Full details of the ACA frequency assignment can be found at: `http://www.aca.gov.au/radiocommunications.htm`

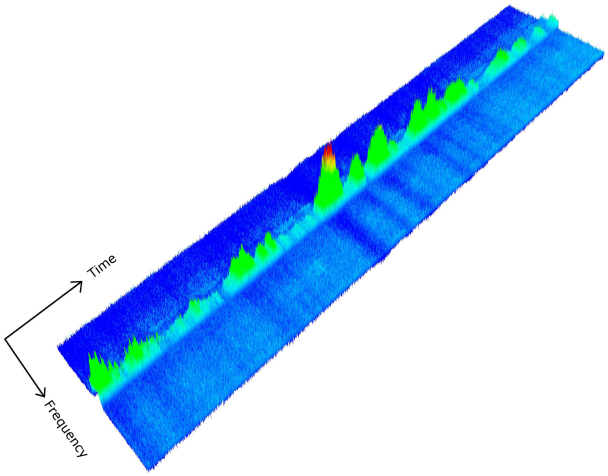


Figure 4.9: Dynamic spectrum of RFI from Narrabri. A total of 30 minutes of data with 512 frequency channels and 1 second integrations. The centre frequency is 2492 MHz and the central 8 MHz of a 16 MHz band are shown.

The Swinburne-MRO recorders were used to capture 30 seconds of a known RFI signal at the ATCA (Figure 4.9), from a Globalstar satellite near 2496 MHz. As can be seen the RFI is mostly random in nature, and large RFI spikes (the red peaks) can corrupt the data across the whole band. The corruption can be seen as a drop in the overall amplitude across the band (the darker blue sections). This drop in amplitude of the signal across the entire frequency range at particular times is due to the RFI dominating the signal so much that it causes the averaging processes used at the telescope to reduce the amplitude of the rest of the bandpass to maintain the correct statistics in the recorded data. Figure 4.10 shows a 10 second integrated bandpass from the same dataset as shown in Figure 4.9. The centre of the band shows the ~1.3 MHz wide section that is polluted with RFI. The strength of the RFI is quite strong. When compared to the size of the cosmic signal, the RFI is the dominating source, making it difficult to observe the actual signal of interest.

In the future, the Swinburne-MRO recorders will be used to undertake detailed RFI surveys at the ATCA and Parkes observatories, in preparation for studies of active RFI mitigation techniques. The initial results shown in Figure 4.9 show that this is feasible.

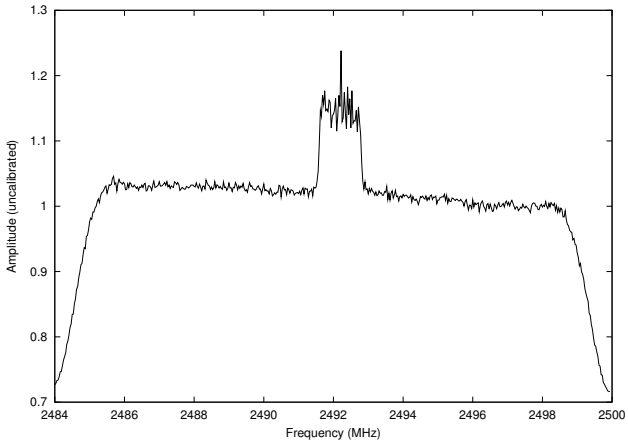


Figure 4.10: A Globalstar satellite RFI signal. A centre frequency of 2492 MHz was used with an integration length of 10 seconds. This shows how the RFI signal can dominate the total signal.

## 4.3 Interferometric applications

In order to obtain higher resolution observations than possible with current single dish telescopes, interferometric observations are made (Chapters 1 and 3). This is done by combining the signals from two or more radio telescopes. The Swinburne software correlator (as discussed in Chapter 3) is designed to combine these signals correctly using a supercomputer, as opposed to dedicated hardware processing as detailed in Chapter 1.

VLBI observations were made of the Vela pulsar, and a reference source PKS B0826−373 with S2 baseband recorders during March 2001. The data collected were transported to the ATNF LBA correlator where it was correlated as well as translated onto disks so that it would be accessible to the software correlator on the Swinburne supercomputer. Comparisons between the LBA and Swinburne software correlator have already been made for the reference source in Chapter 3.

Further observations have been made using the Swinburne-MRO recorders to ensure they functioned correctly when used for VLBI observations. These observations were made of the Vela pulsar as well as PKS J0538−4405 and Centaurus-A, and totaled over 200 GB of recordings for each of the ATCA and Parkes radio telescopes. The Vela data have been correlated and compared to the data from the S2 recorders.

### 4.3.1 VLBI tests using disk-based recorders

During the November 2003 VLBI observing session (mentioned in Chapter 2) test time was requested to carry out Australia's first disk-based VLBI observations. These observations were performed between the Parkes radio telescope and the ATCA in Narrabri, a baseline distance of 321 km. Three test sessions during the regular week of VLBI observing were allocated. The first was to ensure that recordings could be made between the two telescopes correctly and involved the source PKS J0538−4405. The second observation was of the Vela pulsar at 2.3 GHz, and the third was a four hour long recording session aimed at ensuring the recorders would be stable over larger recording times, with observations made of both Vela and Centaurus-A at 4.8 GHz. There was a total of 200 GB of data recorded at each telescope during the VLBI recording phases. This accounts for ~10% of the total capacity of the recorders as they stood at the time of recording.

#### 4.3.1.1 Near-real-time fringe checking

For VLBI purposes a "fringe checker" program was written to allow for confirmation that fringes were found between the telescopes in near-real-time. The fringe checker takes data from two telescopes and forms a lag spectrum of the data in order to find the fringe and determine a rough delay. The fringe checker program is used by sending small sections of data from each telescope in the observation to a central processor. The data are sent via the internet (~512kbps connections during Nov-2003 tests). Once all the required data arrives at the central processor, each baseline can be checked. This program was created specifically for testing a single baseline experiment, and future versions will be able to semi-automatically search for fringes over many baselines.

The program was successfully tested during the November 2003 VLBI observations, where fringes from the observation of the Vela pulsar were detected approximately 15 minutes after the start of recording. This was accomplished by sending one second of data from the ATCA to the Parkes radio telescope. Once the data arrived, they were fringe checked by stepping through the data until a potential fringe was found, at which point the software paused to let the observer confirm or reject the potential fringe. This software detected a large (~60 microsecond) offset between the clock at the ATCA and the clock at Parkes. This

offset was later confirmed when the data were returned to Swinburne University for correlation with the Swinburne software correlator. This clock offset was also confirmed in other data recorded during the same VLBI session by the LBA correlator at the ATNF. This proved to be a very useful exercise as the 60 microsecond offset information was forwarded onto those who processed the data at the LBA correlator, saving the LBA operator considerable time. It is not uncommon to have invalid data taken at one of the telescopes during the VLBI. This can lead to the loss of many baselines of data (6 telescopes involves 15 baselines, 5 telescopes only have 10 baselines). Both the VLBA and EVN VLBI arrays are implementing near-real-time fringe checking programs. With the deployment of more of the Swinburne-MRO recorders, and faster network speeds to the telescopes, the LBA will also be able to achieve this near-real-time fringe checking capability.

#### 4.3.1.2 Swinburne-MRO observations and software correlation of the Vela pulsar.

Observations of the Vela pulsar were made with the Swinburne-MRO recorders as part of the November 2003 VLBI session. The amplitude and phase data as a function of time shown in Figure 4.11 has an integration length of 1 second (~11 pulses). The structure in the amplitudes is due to scattering of the ISM and pulse-to-pulse variabilities of the Vela pulsar (Krishnamohan & Downs 1983). These data were correlated less than one week after they were recorded. This is a vast improvement on the usual turn-around time for VLBI observations which can take a few months. This improvement is due to the disk-based baseband recorders allowing for easier access to the data and the software correlator being able to process the pulsed data faster than it was recorded. This is achievable because only the on-pulse regions need to be correlated (approximately 5% of the total data).

#### 4.3.1.3 Near real-time VLBI using disk-based recording systems

Following the first VLBI tests of the Swinburne-developed baseband recording system, plans are now in place to undertake a global disk-based recording experiment in April 2004 involving telescopes in Australia, South Africa, Japan and the USA. In Australia, telescopes at Parkes and the ATCA will have the Swinburne-

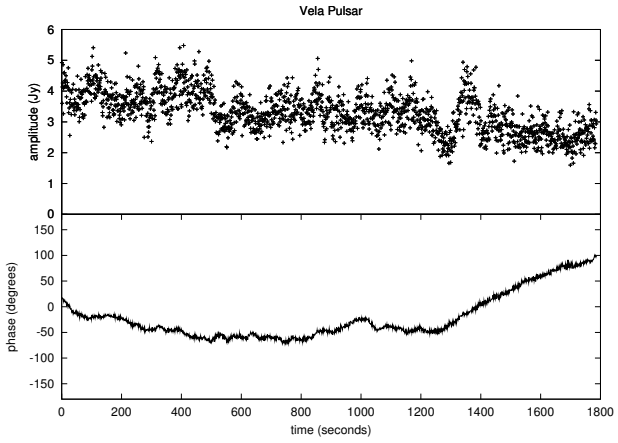


Figure 4.11: Vela Pulsar 1 second integrations of 1 ms wide on pulse region using LCP.

MRO recorders in operation (see Chapter 2). At Tidbinbilla a similar system developed as a collaboration between Swinburne and Prof. Frank Briggs of the Australian National University will be used. At Ceduna and Mopra, systems developed as a collaboration between Swinburne and Dr Richard Dodson of the Institute of Space and Astronautical Science in Japan, will be used. In Hobart a MkV recorder will be used. In South Africa (Hartebeestoek) and the USA (Pie Town) MkV recorders will also be used. In Japan at the Kashima antenna, the Japanese K5 recorder will be used.

On March 1st and 2nd 2004, Briggs, Dodson, Tingay and I gathered at the Parkes radio telescope to test our respective recorders and confirm compatibility. These tests were successful and some of the results are described in Chapter 2. Further tests are planned using the Swinburne-MRO recorders in conjunction with the MkV recorder in Hobart, to test compatibility between recorded data formats. Data recorded using the Japanese K5 recorder will be digitally filtered in software and provided in a pre-specified format to ensure compatibility. All of the Australian radio telescopes' recorders (except Hobart) will be using the same data format, and the same data recording software, albeit slightly different hardware.

Coordinated observations using all of these antennae are scheduled for two 3 hour slots on day of year 106 and 109, 2004, at the frequencies of 1.6 and 2.3 GHz, respectively. Strong continuum sources and masers will be the targets. The observation plan involves the transfer of a small amount of the data from each telescope

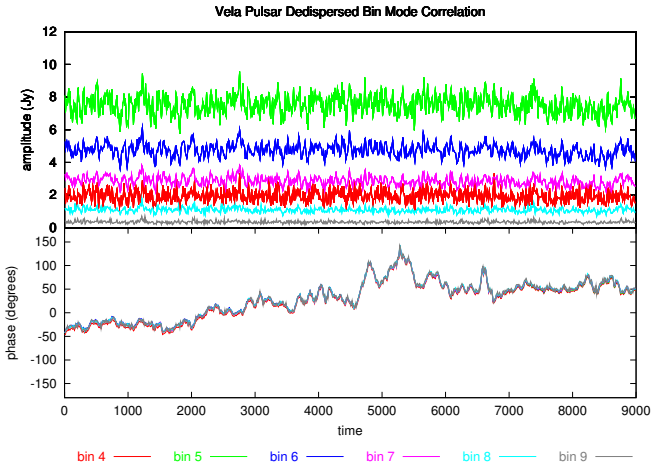


Figure 4.12: Binning mode of Swinburne software correlator. Six bins centred around the on-pulse region, with each bin being 1 ms of data. An integration time of ~10 seconds is used (112 pulses). The phases of the bins agree closely, and the amplitudes give a good indication of which part of the on-pulse region is being processed in each bin.

to a central software correlator that will form near-real-time fringes, performing the most extensive near-real-time e-VLBI experiment yet undertaken[3]. After the observation, the data will be transported back to Swinburne where it will be correlated using the Swinburne software correlator (see Chapter 3).

These experiments are being performed in anticipation that in 2004 or 2005, multi-gigabit connections will be available between ATNF telescopes in Australia and also between the telescopes and Swinburne University, making real-time e-VLBI using software correlation a realistic goal.

The fact that this work can be undertaken is a major outcome of this thesis.

### 4.3.2 Pulsar dedispersion and binning modes of software correlator.

The VLBI observations of Vela that were obtained using the S2 and Swinburne-MRO recorders have been used for testing of the Swinburne software correlator. This has included the testing of the pulsar observing mode, and the dedispersion

[3]e-VLBI efforts by the EVN - `http://www.evlbi.org/evlbi/evlbi.html` and VLBA - `http://web.haystack.mit.edu/e-vlbi/evlbi.html`

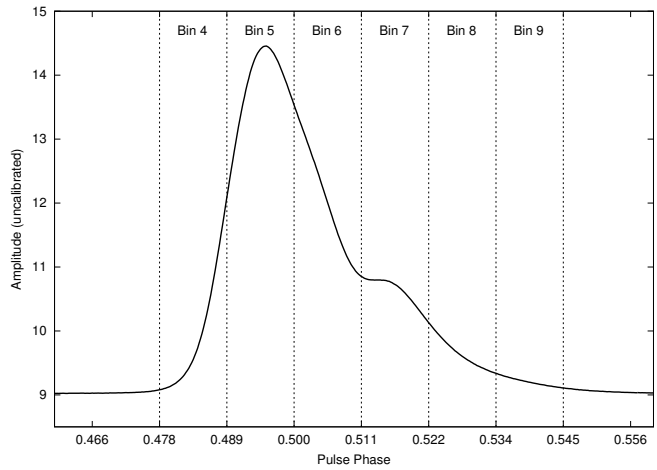


Figure 4.13: Vela pulsar bin locations used during software correlation. The locations of the bins as a function of pulse phase are shown in relation to the data from Figure 4.12.

of pulsar data. These data have also been used in an attempt to study the size of the Vela pulsar emission region (see section 4.3.3), which is an on-going collaboration with Drs Steven Ord and Steven Tingay and Prof. Matthew Bailes of Swinburne University as well as Prof. Jim Cordes of Cornell University.

Figure 4.12 illustrates the pulsar binning mode of the software correlator. It shows 6 on-pulse, 1 ms, bins of the data taken from the Swinburne-MRO recorders. The Vela pulsar has a ~4 ms wide on-pulse region, so the last two bins will not be discussed here as the amplitudes in these bins are low. It is however, interesting to note that some signal is detected, as is expected by the amplitudes shown in Figure 4.13. A comparison of the Vela pulsar's profile shows that the bins are appearing with the correct scale of amplitudes. The first bin (labeled bin 4) is lower in amplitude than the second, which in turn is stronger than the third and fourth bins. The maximum signal in the pulse profile is seen in the early part of the on pulse region, which is why the second bin is stronger. The data in Figure 4.12 have an integration time of 112 pulses, equating to approximately 10 seconds and are also dedispersed, but comparisons to dispersed data are not shown here as the differences are negligible at this frequency and bandwidth.

### 4.3.3 Investigating the Vela pulsar emission region size

The Vela pulsar (PSR J0835−4510) is one of the most well-known and well-studied pulsars. Recently, a number of treatments of the scintillation of the pulsar's emission, due to the ionised plasma that lies between the pulsar and the Earth, have sought to derive information on the size of the pulsar's emission region. To illustrate the flexibility of the software correlator developed in Chapter 3, it was used in an attempt to confirm an analysis of this type performed by Gwinn *et al.* (2000).

Pulsar emission regions are extremely small (Gil & Kijak 1993) and resolving them directly through imaging is impossible with the limitations of current VLBI arrays. Angular resolutions of $2.4\times10^{-8}$ arcseconds would be required, translating into baseline lengths in excess of $2.7\times10^{9}$ km (~17 AU) at a frequency of 1 GHz.

A number of authors have explored the possibility that the observed scintillation of the pulsar emission can be used to probe the high resolution structure of the pulsar emission regions (Gwinn *et al.* 1997; Gwinn *et al.* 2000; Cordes 2000). In particular, Gwinn *et al.* (2000) used VLBI data collected on the Parkes to Tidbinbilla baseline at a frequency of 2.3 GHz to examine the scintillation of the pulsar via the observed distribution of correlated amplitudes. Since scintillation, as occurs under these circumstances, is a strong function of both frequency and time, the correlated amplitude must be measured well within the decorrelation time and bandwidth. Gwinn *et al.* (2000) estimated this to be 15 seconds and 39 $\pm$ 7 kHz for Vela at 2.3 GHz. Thus they correlated their data (using the Haystack Mark IIIB correlator [Whitney 1980]) in a binning mode after making a correction for dispersion of the pulsar signal in the interstellar medium (by adding an extra delay as a function of frequency to their correlator delay model). The data were correlated using 3 time bins across the pulsar's on-pulse region (~4 ms out of the total 89 ms period). The data correlated in these bins were eventually integrated over a period of 10 seconds, with a spectral bin width of 25 kHz. The distribution of amplitude measured in these averaging cells of 10 seconds $\times$ 25 kHz was compared to the theoretically expected distributions for both unresolved and resolved emission regions for the Vela pulsar, as calculated by Gwinn *et al.* (2000). Gwinn *et al.* (2000) claim that their model for a resolved pulsar emission region fits their observed amplitude distribution better than their model for an unresolved emission region, claiming an emission region size of 440 $\pm$ 90 km.

The software correlator was used to correlate data for the Vela pulsar on the

Tidbinbilla to Parkes baseline at a frequency of 2.3 GHz for a 2.5 hour observation using the S2 recorders. This is an almost identical observing setup to that of Gwinn *et al*. (2000). As an exercise in software correlation, an attempt to reproduce and confirm the results of Gwinn *et al*. (2000) has been made. There are a number of reasons for supposing that these new Vela VLBI observations, with software correlation, would give a more robust observational result than that obtained by Gwinn *et al*. (2000). Firstly the new observations, from March 2001, are more sensitive since new receivers have been installed at the telescopes since the 1992 observations of Gwinn *et al*. (2000). There is also a larger available bandwidth (12 MHz of usable bandwidth in 2 polarisations versus 6 MHz of usable bandwidth for Gwinn *et al*. 2000). The instrumental passband is more robust due to the use of digital filter banks in the new observation. The software correlator, and S2 recorders, use 2-bit sampled data, as opposed to the 1-bit samples used by Gwinn *et al*. (2000) with the Mark III recorders (see Chapter 2 - Rogers *et al*. 1983). The software correlator also employs more flexible correlations, including: exact de-dispersion of the pulsar signal in software prior to correlation; arbitrary on-pulse bin widths; arbitrary averaging time (with a minimum of one bin width); and arbitrary spectral resolution.

These improvements over the observations and correlations of the Gwinn *et al*. (2000) work should convey a better understanding of the effects of noise on the scintillation measurements, a better understanding of effects of averaging data over different frequency and time intervals, and ultimately a better analysis of the pulsar emission region size as a function of pulse phase.

Thus, in the first instance, 1 ms bins across the on-pulse region of the pulse phase, 31.25 kHz channels, ten seconds of averaging time and coherent software dedispersion (using a DM of 67.87) were chosen to match as closely as possible the correlation of the Gwinn *et al*. (2000) data. The frequency resolution of 31.25 kHz is still within the coherence time calculated by Gwinn *et al*. (2000), which was calculated for the Tidbinbilla to Hartebeestoek baseline. Four bins were correlated across the on-pulse region for Vela with the areas chosen representing a close match to the areas used by Gwinn *et al*. (2000). Even though four bins were created, adding the last two bins together allows the three bin windows used in the Gwinn *et al*. (2000) data to be created. The third bin in the Gwinn *et al*. (2000) is approximately twice as wide as the first two bins. The observed distributions

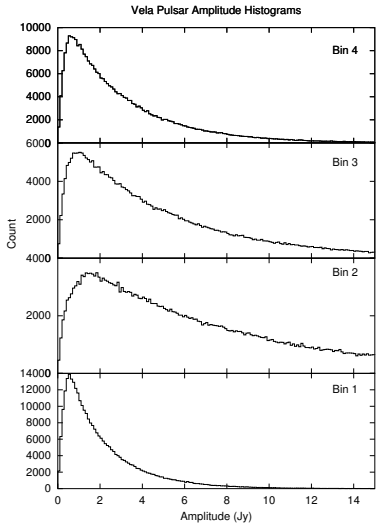


Figure 4.14: Gated Vela pulsar histograms of observed amplitudes from Swinburne. The four panels show the four separate pulse bins that are processed.

of amplitude for the new observations are shown in Figure 4.14. The amplitude distribution from separate areas of the on-pulse window are different, as was found in the Gwinn *et al.* (2000) data. There are 512 channels (giving channel widths of 31.25 kHz) and an integration length of 112 pulses (approximately 10 seconds), to match the correlation mode of Gwinn *et al.* (2000).

Fits to both the Gwinn *et al.* (2000) data and the Swinburne data have been attempted, with the fit to the Gwinn *et al.* (2000) being successful, while a good fit to the Swinburne data was not possible (see Figure 4.15). These fits have been made using the Gwinn *et al.* (2000) models for an unresolved source in the presence of noise (equation 15 of Gwinn *et al.* 2000)

The Gwinn *et al.* (2000) models include calculations that are intended to remove effects that are known to occur when observing pulsars. The Swinburne software correlator uses 2-bit correction techniques (Jenet & Anderson 1998) which assume the digitisation levels are optimally set. However, these techniques do not implement scattered power corrections and as such significant corrections to the raw data are required. The corrections applied to the 2-bit data mean that many of the noise corrections applied by Gwinn *et al.* (2000) are unnecessary. However, these corrections may introduce other effects, that are difficult to quantify,

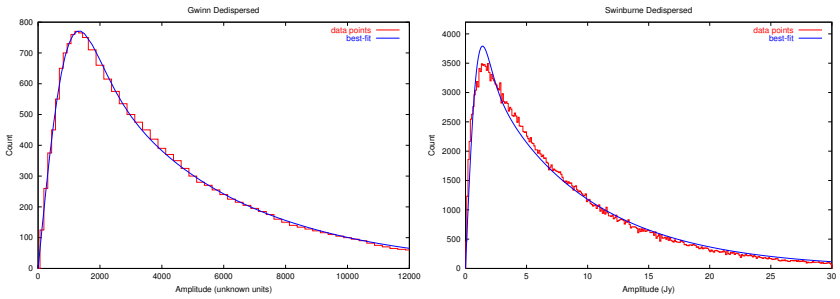


Figure 4.15: Fits to the Gwinn *et al.* (2000) and Swinburne histogram plots. The left panel shows the Gwinn *et al.* (2000) data with a good fit, and the right panel shows the Swinburne data with a poorer fit.

that may contribute to the poor fit of the Gwinn *et al.* (2000) model with the data shown in Figure 4.15.

It is possible that other factors also contribute to the poor fit in Figure 4.15. For example it was noted during an analysis of the Gwinn *et al.* (2000) results and data, that a number of inconsistencies, errors, and omissions exist in this paper. For example:

- Equation 23 does not follow from the preceeding text;
- The caption for Figure 4 is inconsistent;
- The widths of the gates in sections 3.1 and 3.2.3 are inconsistent; and
- Table 4 shows a size parameter that is not labeled correctly.

Most importantly the size of the pulsar emission region shown in Table 4 appears to have been calculated incorrectly. Instead of the quoted 440km size, the true size appears to be approximately 1330km.

In an endeavour to determine possible differences in the analysis Dr Carl Gwinn was contacted and a correspondence established. For instance a comparison of the computer codes that generated the theoretical curves detailed in Gwinn *et al.* (2000) was agreed to but Dr Gwinn could not locate the code he used to generate the theoretical curves for his paper. The comparison could not therefore be made. The number of typographical errors and inconsistencies in the Gwinn *et al.* (2000) paper instill a level of skepticism that the detailed theoretical expressions, as written in the paper, are accurate and usable in a analysis of the new data.

A soon-to-be-published paper by Prof. Jim Cordes reveals many of the errors made by Gwinn *et al*. (2000). In particular, Cordes claims that the Vela pulsar emission region region is unresolved, contrary to the claim of Gwinn *et al*. (2000). Cordes refutes Gwinn's claims on the basis of the underlying theoretical treatment of scintillation. Also, similar claims have been made by Jean-Paul Macquart (private communication). It is beyond the scope of this masters thesis to consider several competing theories of scintillation and the implications for the Vela pulsar emission region, especially when the expert theoreticians cannot agree amongst themselves! At first inspection what seemed to be a straight forward comparison of results has become very complicated on both the theoretical and observational fronts. Prof Jim Cordes has agreed to provide the code he used in his paper for the analysis of the new Vela VLBI data. This project is therefore on-going and it is hoped that within the next 12 months the Vela data correlated on the software correlator will make a useful contribution to the debate surrounding the Vela pulsar emission region.

## 4.4 Conclusions

This chapter has detailed observations made with both the Swinburne-MRO disk-based baseband recorders and the S2 tape-based baseband recorders, all correlated in software. Studies of masers using single dish telescopes have been produced with high spectral resolution with the use of software correlation techniques. Further studies of masers are planned, and will be undertaken in collaboration with other investigators. Pulsar studies have shown that the recorders have stable time recording abilities, and that they compare well with the faster and wider bandwidth CPSR2 recorder. Studies of RFI have also been made to show that these recorders are capable of recording enough information that mitigation techniques can be developed, as well as monitoring of RFI at various observatories.

Correlation of VLBI data sets has shown that the software correlator is capable of producing results that are directly comparable to the LBA correlator at the ATNF. Also the software correlator has shown its flexibility in being able to process pulsar data, including the dedispersion of the pulsars, and will be used in an on-going investigation of the study of emission region size of the Vela pulsar. The software correlator has also been shown to work well with the Swinburne-MRO

recorders, integrating the two main aims of this thesis.

The VLBI observations have also made use of near-real-time fringe checking software and it is expected that this will be developed further in the near future, with more extensive tests planned for April 2004. With the inclusion of faster speed connections between the radio telescopes in Australia it is possible that real-time fringe checking, and even real-time VLBI will be available in the not too distant future.

# Chapter 5

# Conclusions and future directions

## 5.1 Conclusions

The previous chapters (Chapters 2, 3 and 4) have detailed the work carried out for this thesis. The aims of the thesis were to demonstrate software solutions for specific data processing problems in radio astronomy and to demonstrate that these solutions can be used to obtain useful scientific data. Chapters 2 and 3 describe disk-based recorders and software developed to implement software solutions for radio astronomy. Chapter 4 outlines the acquisition of radio astronomy data that demonstrate the scientific usefulness of the software solution. In particular, the software solution is being used in two on-going, collaborative projects. Also the work undertaken in this thesis contributed greatly to the work program of the Swinburne SKA project, funded by the Major National Research Facility Program "Gemini and SKA: Australia's Astronomy Future". The aims of this thesis have therefore been fulfilled.

Chapter 2 detailed the work carried out in the development and testing of disk-based baseband recorders. The development of a disk-based baseband recorder (called Swinburne-MRO) for Australian radio telescopes was successfully undertaken. Data were recorded at both the Parkes radio telescope and the Australia Telescope Compact Array (using the tied array to form a synthesised radio telescope). These recorders have been tested at double the recording rate of the existing tape-based S2 recorders that are currently in use for VLBI recording throughout Australia. The data from these recorders have been made easily available for further processing by modifying the existing *BasebandDSP* library already in use

at Swinburne University.

Chapter 3 covered the work carried out in designing and building a prototype software correlator for VLBI purposes, with the specific intent of using the method of XF correlation that is currently used by the LBA correlator at the ATNF. The software correlator implemented post-correlation fringe-rotation and fractional sample error correction. This leads to a trade-off due to decorrelation of the signal. By using small correlation timescales (~1ms) this trade-off is kept to a minimum. This approach is ideal for wide bandwidth continuum observations, and is not suited to narrow bandwidth, high frequency observations. The correlator performance was verified through a direct comparison to the existing LBA correlator. Further development of the software correlator, including areas of pre-correlation fringe rotation; development of an FX style correlator; and an enhanced data distribution model are planned.

Chapter 4 made use of the Swinburne-MRO recorders developed in Chapter 2, and the software described in Chapters 2 and 3, for astronomical observations of radio sources. Observations of masers show that the hardware and software are capable of producing high spectral resolution results. Pulsar observations show the time stability of the recorders, and demonstrate the possibility of observing pulsars at radio telescopes not usually fitted with pulsar processing equipment. Studies of RFI have shown that these recorders would also be ideal for the monitoring and analysis of RFI mitigation tests. The Swinburne-MRO recorders have also been used in single baseline VLBI observations between the Parkes radio telescope and the ATCA. These observations have helped to show the flexibility of the software correlator that was developed in Chapter 3. VLBI observations of the Vela pulsar have been made in an attempt to understand the emission methods of pulsars. This work is being continued in collaboration with other investigators.

The use of the Swinburne-MRO baseband recorders to record data for single dish and VLBI observations, as well as the use of the custom software to process the data taken during these observations, has tied together the two main aims of this thesis.

## 5.2 Future directions for baseband recorders and software correlation

The development of disk-based recorders is currently a hot topic in radio astronomy with the development of MkV, K5, PCEVN and other recordings systems. The Swinburne-MRO recorders, while not yet as advanced as some existing systems have been shown to record at double the maximum data rate of the existing S2 recorders used in Australian VLBI observations . These recorders can improve the Australian LBA network of telescopes.

With the introduction of the Swinburne-MRO type recorders at Australian telescopes it will be possible to carry out near-real-time fringe checking of VLBI observations. This can lead to a faster turn around time for VLBI data as well an improvement of the effectiveness of the LBA as errors in the observation can be rectified as the observation is being undertaken. Installation of upgraded networks between the radio telescopes can lead to real-time fringe checking as well as the possibility of real-time VLBI observations. An attempt at near-real-time fringe checking will be made during the April 2004 VLBI observations where nine telescopes world-wide will be used in the largest to date disk-based recording observations.

The Swinburne software correlator has already been shown to function correctly when compared with the LBA correlator. Further development of this correlator will lead to improvements in the processing time of data, as well as the methods of processing. Different processing techniques will be tested with this correlator, including FX correlation and pre-correlation fringe rotation.

The *BasebandDSP* library is under continual development, and one of the areas that is planned to be enhanced is the ability to process data recorded in the MkV, K5 and Swinburne-MRO recorder formats. Another area of development is the conversion of data from any of the recorded formats to any other format (where possible). This means that systems such as the JIVE correlator and MkV correlators can be used to process data from any disk-based recorder. It will also be necessary to develop software filter-banks to convert high bandwidth recorded data to a lower bandwidth to allow data from different systems to be correlated together.

# References

Bailes M., Manchester R. N., Kesteven M. J., Norris R. P., Reynolds J. E., 1990, *The Proper-Motion of Six Southern Radio Pulsars*, MNRAS, 247, 322

Bailes M., 2003, in Bailes M., Nice D., Thorsett S., eds, Radio Pulsars. Astronomical Society of the Pacific, p. 57

Bare C., Clark B., Kellermann K., Cohen M., Jauncey D., 1967, *Interferometer experiments with independent local oscillators*, Science, 157, 189

Bell J. F. *et al*., 2001, *Base Band Data for Testing Interference Mitigation Algorithms*, Publications of the Astronomical Society of Australia, 18, 105

Brisken W. F., Fruchter A. S., Goss W. M., Herrnstein R. M., Thorsett S. E., 2003, *Proper-Motion Measurements with the VLA. II. Observations of 28 Pulsars*, AJ, 126, 3090

Burrows A., Hayes J., 1996, *Pulsar Recoil and Gravitational Radiation Due to Asymmetrical Stellar Collapse and Explosion*, Physical Review Letters, 76, 352

Cannon W. H. *et al*., 1997, *The S2 VLBI system*, Vistas in Astronomy, 41, 297

Carlson B. R., Dewdney P. E., Burgess T. A., Casorso R. V., Petrachenko W. T., Cannon W. H., 1999, *The S2 VLBI Correlator: A Correlator for Space VLBI and Geodetic Signal Processing*, PASP, 111, 1025

Casse J. L., 1999, *The European VLBI Network MkIV Data Processor*, New Astronomy Review, 43, 503

Caswell J. L., 1999, *OH masers at 1612 and 1720MHz in star-forming regions*, MNRAS, 308, 683

Chuprikov A., Likhachev S., Molotov I., Jauncey D., Reynolds J., Roberts P. Wilson W., 1996, in Proceedings of the Technical Workshop for APT and APSG. p. 289

Clark B. G., Cohen M. H., Jauncey D. L., 1967, *Angular Size of 3C 273B*, ApJ, 149, L151+

Clark B., 1973, *The NRAO tape recorder interferometer system*, Proc. IEEE, 61, 1242

Cohen M. H. *et al*., 1975, *Observations with a VLB array. I - Introduction and procedures*, ApJ, 201, 249

Cordes J., Lazio T., 2002. *Ne2001. i. a new model for the galactic distribution of free electrons and its fluctuations*, astro-ph/0207156

Cordes J., 2000. *Interstellar seeing. ii. the case of the vela pulsar: Source unresolved*, astro-ph/0007233

Costa E., Loyola P., 1999, *CCD astrometry of faint compact extragalactic radio sources . I. First results and description of the program*, AAPS, 139, 297

Crane P. C., Napier P. J., 1989, in ASP Conf. Ser. 6: Synthesis Imaging in Radio Astronomy. p. 139

Dodson R., Legge D., Reynolds J. E., McCulloch P. M., 2003, *The Vela Pulsar's Proper Motion and Parallax Derived from VLBI Observations*, ApJ, 596, 1137

Emmering R. T., Watson W. D., 1994, *Apparent sizes and spectral line profiles for spherical ('three-dimensional') astrophysical masers*, ApJ, 424, 991

Garrett M. A., Porcas R. W., Pedlar A., Muxlow T. W. B., Garrington S. T., 1999, *Wide-field VLBI imaging*, New Astronomy Review, 43, 519

Gil J. A., Kijak J., 1993, *Period Dependence of Radio Emission Altitudes in the Pulsar Magnetosphere*, A&AS, 273, 563

Goddard D., Haynes R., 1994, *Pioneering a new astronomy; John G. Bolton Memorial Symposium, Parkes, New South Wales, Australia, Dec. 9-10, 1993*, Australian Journal of Physics, 47

Gold T., 1968, *Rotating neutron stars as the origin of the pulsating radio sources*, Nature, 218, 731

Gwinn C. R. *et al*., 1997, *Size of the VELA Pulsar's Radio Emission Region: 500 Kilometers*, ApJ Letters, 483, L53+

Gwinn C. R. *et al*., 2000, *Size Of the Vela Pulsar's Emission Region at 13 Centimeter Wavelength*, ApJ, 531, 902

Hankins T. H., 1971, *Microsecond Intensity Variations in the Radio Emissions from CP 0950*, ApJ, 169, 487

Harp G. R., 2002, in Advanced Telescope and Instrumentation Control Software II. Edited by Lewis, Hilton. Proceedings of the SPIE, Volume 4848, pp. 1-11 (2002). p. 1

Harrison P. A., Lyne A. G., Anderson B., 1993, *New determinations of the proper motions of 44 pulsars*, MNRAS, 261, 113

Hecht E. Optics, p. 422, Addison-Wesley, 2nd edition, 1987

Hewish A., Bell S. J., Pilkington J. D. H., Scott P. F., Collins R. A., 1968, *Observation of a rapidly pulsating radio source*, Nature, 217, 709

Horiuchi S. *et al*., 2000, *Imaging Capability of the Mitaka VSOP Correlator*, Advances in Space Research, 26, 625

Hulse R. A., Taylor H. J., 1974, *Discovery of a Pulsar in a Close Binary System.*, Bulletin of the American Astronomical Society, 6, 453

Jansky K., 1933, *Electrical disturbances apparently of extraterrestrial origin*, Proc.IRE, 21, 1387

Jenet F. A., Anderson S. B., 1998, *The effects of digitization on non-stationary stochastic signals with applications to pulsar signal baseband recording*, PASP, 110, 1467

Johnston K. J. *et al*., 1995, *A Radio Reference Frame*, AJ, 110, 880

Karastergiou A., Kramer M., Lyne A. G., Johnston S., Bhat R., Gupta Y., 2003, in ASP Conf. Ser. 302: Radio Pulsars. p. 195

Kohri K., Nagataki S., 2000, *Pulsar Kick and Asymmetric Iron Velocity Distribution in SN 1987A*, Progress of Theoretical Physics, 103, 713

Kondo T., Koyama Y., Nakajima J., Sekido M., Osaki H., 2003, in Minh Y., ed, New Technologies in VLBI. p. 205

Kraus J. D., 1986, Radio astronomy. Powell, Ohio: Cygnus-Quasar Books, 1986

Krishnamohan S., Downs G. S., 1983, *Intensity dependence of the pulse profile and polarization of the VELA pulsar*, ApJ, 265, 372

Leshem A., van der Veen A., Boonstra A., 2000, *Multichannel Interference Mitigation Techniques in Radio Astronomy*, ApJ Supplement, 131, 355

Lipson S., Lipson H. Optical Physics, chapter 9, p. 269, Cambridge Univserity Press, 2nd edition, 1981

Ma C. *et al.*, 1998, *The International Celestial Reference Frame as Realized by Very Long Baseline Interferometry*, AJ, 116, 516

Michelson A., 1890, *On the application of interference methods to astronomical measurements*, Philosophical Magazine, 30, 1

Minh Y., ed, New technologies in VLBI, International VLBI Service for Geodesy and Astrometry, The Astronomical Society of the Pacific, 2002

Moellenbrock G. A. *et al.*, 1996, *A 22 GHz VLBI Survey of 140 Compact Extragalactic Radio Sources*, AJ, 111, 2174

Moore G., 1965, *Cramming more components onto integrated circuits*, Electronics, 38(8)

Moran J. M., Crowther P. P., Burke B. F., Barrett A. H., Rogers A. E. E., Ball J. A., Carter J. C., Bare C., 1967, *Spectral Line Interferometry with Independent Time Standards at Stations Separated by 845 Kilometers*, Science, 157, 676

Moran J. Very long baseline interferometer systems, chapter 5.3, p. 174, 12,, Academic Press, 1976

Moscadelli L., Menten K. M., Walmsley C. M., Reid M. J., 2003, *A High Spectral Resolution VLBI Study of the 12 GHz Methanol Masers in W3(OH): Their Submilliarcsecond Structure and Clues on Saturation*, ApJ, 583, 776

Nyquist H., 1924, *Certain factors affecting telegraph speed.*, Bell System Technical Journal,, 3, 324

Pogrebenko S., Gurvits L., Campbell R., Avruch I., Lebreton J., van't Klooster C.,

2003, in Wilson A., ed, Workshop on Planetary Probe Atmospheric Entry and Descent Trajectory Analysis and Science. European Space Agency

Reber G., 1944, *Cosmic Static.*, ApJ, 100, 279

Reynolds J. E., Jauncey D. L., Russell J. L., King E. A., McCulloch P. M., Fey A. L., Johnston K. J., 1994, *A radio optical reference frame. 7: Additional source positions from a Southern hemisphere short baseline survey*, AJ, 108, 725

Ritakari J., Mujunen A., 2002, in International VLBI Service for Geodesy and Astrometry. p. 128

Rogers A. E. E. *et al.*, 1983, *Very-long-baseline radio interferometry - The Mark III system for geodesy, astrometry, and aperture synthesis*, Science, 219, 51

Romani R. W., 2003, in ASP Conf. Ser. 302: Radio Pulsars. p. 331

Romney J. D., 1995, *The VLBA Correlator*, Bulletin of the American Astronomical Society, 27, 815

Russell J. L. *et al.*, 1994, *A radio/optical reference frame. 5: Additional source positions in the mid-latitude southern hemisphere*, AJ, 107, 379

Stairs I. H., Splaver E. M., Thorsett S. E., Nice D. J., Taylor J. H., 2000, *A baseband recorder for radio pulsar observations*, MNRAS, 314, 459

Taylor G. B., Carilli C. L., Perley R. A., eds, Synthesis Imaging in Radio Astronomy II, 1999

Taylor J. H., Weisberg J. M., 1982, *A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16*, ApJ, 253, 908

Taylor J. H., Manchester R. N., Lyne A. G., 1993, *Catalog of 558 pulsars*, ApJ Supplement, 88, 529

Thompson A. R., Moran J. M., Swenson G. W., 1994a, Interferometry and synthesis in radio astronomy. Kreiger Publishing Company

Thompson A., Moran J., Swenson G. Interferometry and synthesis in radio astronomy, chapter 9, p. 252, Kreiger Publishing Company, 1994b

Tzioumis A. K., 1997, *The Australian LBA–status and developments*, Vistas in Astronomy, 41, 311

van Straten W., Bailes M., Britton M., Kulkarni S. R., Anderson S. B., Manchester R. N., Sarkissian J., 2001, *A test of general relativity from the three-dimensional orbital geometry of a binary pulsar*, Nature, 412, 158

Watson W. D., Wyld H. W., 2003, *Apparent Sizes and Spectral Line Profiles for Spherical and Disk Masers: Solutions to the Full Equations*, ApJ, 598, 357

Watson W. D., Sarma A. P., Singleton M. S., 2002, *Gaussian Spectral Line Profiles of Astrophysical Masers*, ApJ Letters, 570, L37

Whitney A. R., 1980, in Radio Interferometry Techniques for Geodesy. p. 317

Whitney A. R., 1993, in IAU Symp. 156: Developments in Astrometry and their Impact on Astrophysics and Geodynamics. p. 151

Whitney A. R., 2002, in Proceedings of the 6th EVN Symposium. p. 41

Wietfeldt R., Baer D., Cannon W., Feil G., Jakovina R., Leone P., Newby P., Tan H., 1996, *The S2 very long baseline interferometry tape recorder*, IEEE Transactions on Instrumentation and Measurement, 45(6), 923

Wietfeldt R., van Straten W., del Rizzo D., Bartel N., Cannon W., Novikov A., 1998, *The S2 baseband processing system for phase-coherent pulsar applications*, A&AS, 131, 549

Wilson W., Roberts P., Davis E., 1996, in King E., ed, Proceedings of 4th APT Workshop. p. 16

# Appendix A

# User Manual for disk-based recording at Australian Radio Telescopes

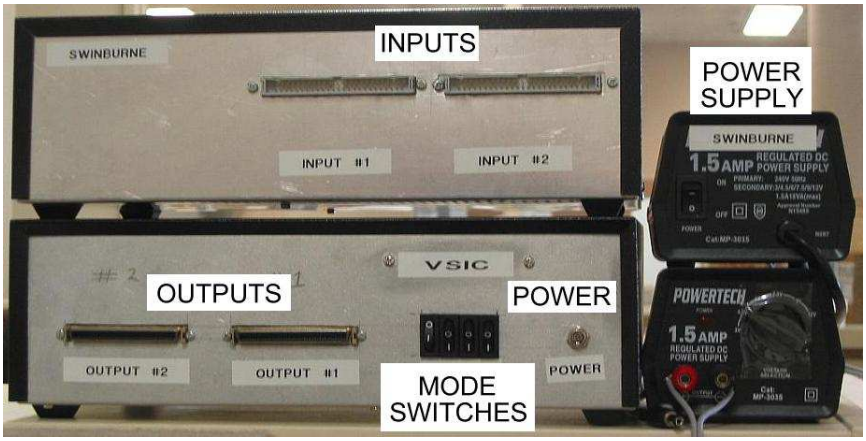


Figure A.1: Front and rear views of the VSIC Unit and Power Supply.

This document details the setup and connection of VSIB/VSIC hardware, computers and storage media to the DAS units at Australian radio telescopes. It assumes that the computer being used is a Dell PowerEdge 1600SC. Apple XRaid units are assumed as the storage medium.

**WARNING: If you don't understand all of this document then don't proceed without advice. Please read it through before attempting the setup.**

# VSIB and VSIC Hardware

The VSIC is a converter card that is housed in an external box. The VSIC will be directly connected to both the DAS and the VSIB card and requires a separate power supply. This power supply must be run at 4.5 Volts and support up to 1.5 Amps.

The VSIB is located inside the computer and has one external connector. Power for the VSIB is taken from the PCI bus in the computer.

## Plugging in the VSIC and VSIB

Figure A.1 shows the front and rear of the VSIC units. The important connectors have been labeled (#1 and #2); if input connector #1 is connected to the DAS, output connector #1 should be connected to the VSIB (Figure A.2). The VSIC is capable of converting the data from two DAS units, hence the two sets of connectors, #1 and #2.

The power plug is pre-wired and should not be modified. The voltage on the power supply should be set to 4.5 Volts. The inner pin is positive, with the outside of the plug being ground. The mode switches are to configure the VSIC board to the correct data conversion mode. The switches are connected to pins 7, 5, 3 and 1 on the VSIC board, see Table A.1 for details of different modes. Pin 7 corresponds to the left most switch, followed by 5, 3 and 1 on the right. If the 0 on the button is pressed in the switch is off, pressing 1 in means the switch is on. There is a switch on the rear of the power supply to allow you to turn it on and off without the need to unplug the power. If the VSIC is not going to be used for a long time it is suggested that this switch be turned off. The power supply tends to get quite warm. It will draw around 1 Amp if only one output is being used, and 1.4 Amps for both outputs.

The DAS is plugged into the input port using a 50 pin ribbon cable that is available at the telescope. The output plug is a 40 pin rainbow cable that plugs into the VSIB connector located on the computer.

Table A.1: VSIC conversion modes

| Mode | 7 | 5 | 3 | 1 |
|---|---|---|---|---|
| Test | 0 | 1 | 1 | 1 |
| S2 | 1 | 1 | 1 | 0 |

# Dell 1600SC Computer

The computers used to record the data from the VSIB onto the external disks are Dell PowerEdge 1600SC servers. They have 3 separate PCI buses, enabling full recording of the data as well as the possibility of streaming data to remote machines.

Figure A.2 shows the rear of two Dell PowerEdge 1600SC machines. The left is the machine located permanently at Parkes, the right shows the portable system. The main difference between the two is that the Parkes version uses fibre ethernet for connection to the local area network as well as the CPSR2 cluster. The portable version uses unshielded twisted pair (UTP) gigabit ethernet. The Parkes machine also has redundant power supplies.

## Plugging in the computer

The important connectors on the computer have been labeled to make things clearer[1]. The port labeled network is where the telescope's local area network (lan) is plugged in. This should be an auto-negotiating port. It is preferable that the machines have permanent IP addresses, however they can be reconfigured to function as a DHCP machine if needed. The keyboard can be plugged into the purple port located just below the power supply. The power supply is located at the top of the machine.

The VSIB connection is located at the very bottom of the figure A.2. This is where the output from the VSIC is plugged in using the 40 pin rainbow coloured ribbon cable.

The Apple XRaid is plugged into both the C1 and C2 ports (labeled here on the Parkes machine). This port is located just above the VSIB connection. The C1 cable is plugged into the left side when looking from behind, and the C2 cable into the right. There should also be a cable run from the UTP port labeled XR-NET. This is to allow communication with the Apple XRaid so that it can be remotely powered and configured if needed.

# Apple XRaid

The rear of the Apple XRaid is shown in figure A.3 and this is where all of the important plugs are connected. The power button for the XRaid unit is located on

[1] The photos may not show all the labels, but the computers are fully labeled and this document can be used to locate the correct connectors if need be.

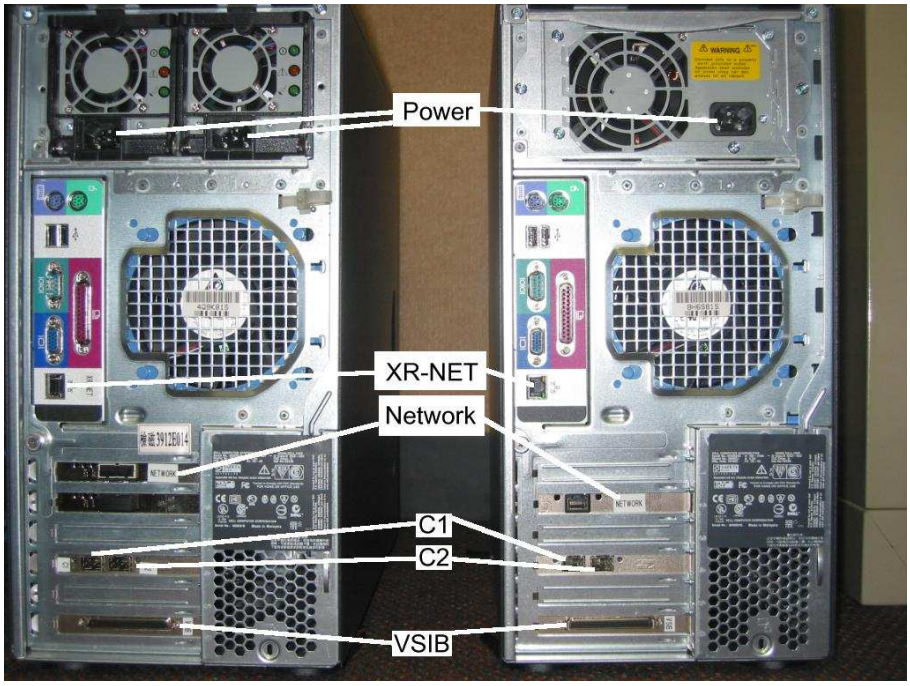


Figure A.2: Dell PowerEdge 1600SC computer. The machine on the left is the Parkes machine, and the machine on the right is the portable system. The Parkes machine has extra features such as redundant power supply and an extra ethernet. This is to allow direct connection to the CPSR2 cluster at Parkes.

the rear. The disk drives should be labeled in order from 1 on the left to 14 on the right. It is important that they be inserted into the cabinet in the correct order. However, there is only a need for each bank of 7 to be plugged in correctly. That is they should run from 1 to 7 on each side (or 8 to 14).

## Plugging in the Apple XRaid

Figure A.3 shows labels on all of the important plugs. As discussed in the computer section the C1 cable from the computer plugs into the top of the XRaid, and the C2 cable plugs into the bottom. The cable from the XR-NET UTP port on the back of the computer is plugged into the UTP port labeled XR-NET, this could be at the top or the bottom of the XRaid depending on which XRaid it is, simply plug in where the port is labeled. There are two power plugs that should both be connected. To power the unit up, press the power button. This should be done before booting the computer. To power the unit down, the disks need to be unmounted, press and hold the power button until the lights turn off. It is advised

that the computer is powered down first to ensure no loss of data.

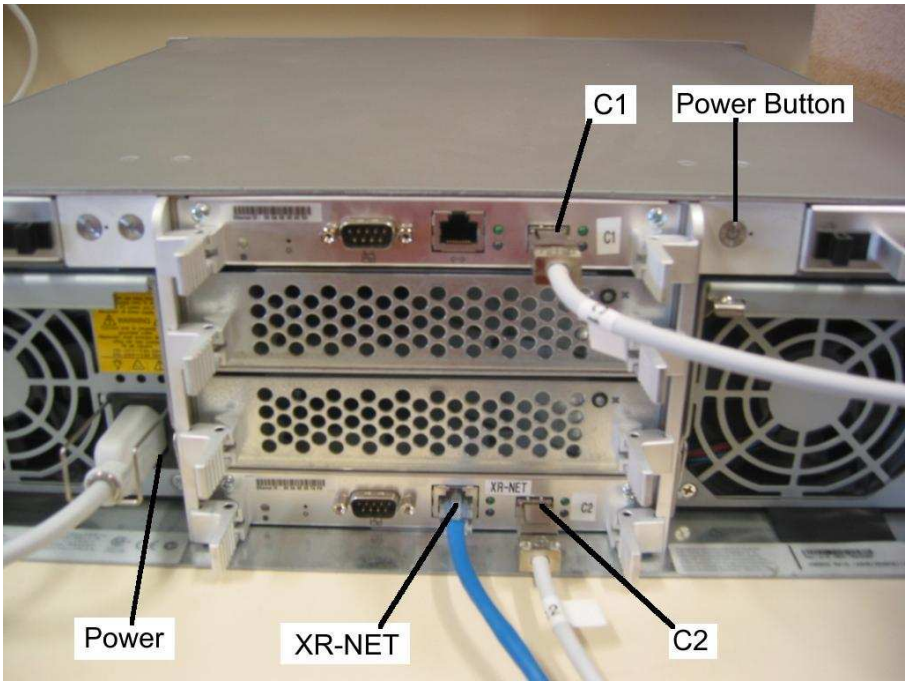


Figure A.3: Apple XRaid rear

Figure A.4: Apple XRaid front

## Software

There are two main types of software that runs the recorders. The first is the software to drive the hardware and the second is the software that initiates the recording and various configuration options. At present the recording software

is still in early stages of development and it is recommended that the user writes down all of the settings that are used at the telescope to ensure that the data are reduced correctly during post processing.

## Drivers

The drivers should already be configured and working on the machine. When booting the machine the `linux-vlbi` profile should be used (this should be the default). The version of linux running on the machine is SuSE 8.2, and it will be running a modified 2.4.20 kernel. The modifications allow for the big-phys-area requirements of the driver. Most of the setup information is available in the documentation that ships with the VSIB recorders.

## Recording

Once the machine has been booted, you need to login. The username is `vlbi` and the password can be obtained from observatory staff. Once logged in you need to manually mount the XRaid disks. This is done by typing `mount /export/xraid0; mount /export/xraid1`. This assumes that both xraid units are powered and connected to the PC. The recording software is located in the `/lfs/data0/root/vsib` directory. The best way to record data is to change directory to where the data are going to be written, that is /export/xraid0/<directory name>, and then run the software in this directory.

To record data type something like:

```
/lfs/data0/root/vsib/wr 80000 24000 1 outfile%04d 4000 < /dev/vsib
```

- The first argument (80000) is the number of bytes to write in each block. A block is written to disk each time it is filled. It is strongly advised not to change this number.

- The second argument (24000) is the number of blocks to write in total. This number dictates the length of the recording. 24000 will record for 60 seconds. The total recording time is 4000 times the number of seconds you wish to record. This assumes a data rate of 128 or 256 Mbits/second.

- The third argument should be left as 1 - further details are found in the documentation that accompanied the boards (/lfs/data0/root/vsib/docs/).

- The fourth argument details the filename, outfile is the name of the file, %04d tells it to use four numbers for the filenames, e.g. outfile0000 outfile0001 outfile0002 outfile0003, etc. If recording for a long duration you may need to replace the %04d with a larger number, e.g. %10d.
- The fifth argument is the number of blocks to write for each file. 4000 is 10 seconds at 80000 bytes per block, again assuming a data rate of 128 or 256Mbits/second.

Apart from the name of the filename and the length of recording, it is advised that you use the default parameters to ensure that the recording works properly.

Note: The recording modes assume a recording speed of 128 or 256 MBits/s. The reason that it can be 128 or 256 is that the recorder cannot record a 4 bit mode. The only mode that suits 128MBits is the 8 bit mode, which would give the same data rate as 256Mbits/second. There is software available which will reduce the data back to 128Mbits/second after it is recorded. The recorder software will be replaced by better versions to remove this problem.

# Further help

Please contact the person who sent you this document if you are seeking further help. Also see the documentation that ships with the VSIB hardware, located on the computer in /lfs/data0/root/vsib/docs.

# Appendix B

# Running the Software Correlator

This document details the use of the prototype software correlator that has been developed at the Swinburne University of Technology. It formed part of the Masters thesis for Craig West at the Centre for Astrophysics and Supercomputing.

## Correlator Details

The software correlator is built to run on the Supercomputer located at Swinburne University of Technology. However, it should run in any cluster computing environment. The software correlator is designed to use the Message Parsing Interface (MPI) as the means of communication between the multiple processors. The preferred version of MPI (`http://www-unix.mcs.anl.gov/mpi/`) is LAM (`http://www.lam-mpi.org/`) as it is more flexible than MPICH (`http://www-unix.mcs.anl.gov/mpi/mpich/`), although the software should function with any complete version of MPI.

## Running the correlator

Running the correlator requires having an MPI environment already configured. It is assumed that the user understands how to use their own local version of MPI, and none of those details will be discussed here.

As of version 0.69 of the `mpi_swc` correlator, the usage of the correlator is:

```
Usage: mpi_swc <options> config_file
Options:
        -B nodes    Number of Baseline nodes to initalise [1]
        -D dm       Dedispersion (dm=-1 for automatic)
        -h          This help screen
        -P name     Pulsar mode, name is the from pulsar catalogue []
        -s source   Source number to process [0]
        -S segments Number of contigous segments [0]
        -T secs     Start processing secs into source [0.0]
        -v          Version information
        -V          Verbose mode (lots of output)
```

The items in the square brackets indicate the default values.

**-B nodes** is used to control the number of baselines to process at a time. It should not be set higher than the total number of baselines as this will not increase speed, but actually limit the number of machines being used as processing nodes. The baseline nodes are automatically defined as processors 1 to the number defined here. Processor 0 is the head node processor. A number smaller than the total number of processors can be used and this will mean the correlator makes extra runs to complete all the baselines.

**-D dm** is for controlling the dispersion measure (dm) to be used when working in pulsar mode. It is inactive (even if supplied) for any modes that are not pulsar (use -P to activate the pulsar mode). If set to -1 the dm of the given pulsar will be used automatically. The dm should not be set to 0, instead, if you require a 0 dm, use a small number like 0.001.

**-h** print the help screen

**-P name** this is the name in the catalogue of the pulsar. This is used to obtain information such as the period and dm of the pulsar. Pulsar mode is activated when using this option. This will cause small bins to be generated across the pulsar period. These bins can then be added in phase at a later stage with the correlator averaging program (`corrave` - see below).

**-s source** use this to process the sources (defined in the configuration file) as needed. It will only allow you to process one source at a time. The source numbers start at 0 and have no relationship to the source number defined in the configuration file.

**-S segments** is used to control the number of segments that are being processed when in pulsar mode (see -P). The size of the segment is defined by BUFFER_SIZE (see configuration file). The number of segments are automatically centred over the pulse phase 0 area. That is where the pulsar is timed. Use of pulsar software may be needed to move the pulse phase centre to the correct area.

**-T secs** tells the correlator to ignore the first `secs` number of seconds before processing the current source. Processing will continue till the end of the source.

**-v** gives the current correlator version (currently 0.69).

**-V** turns on the verbosity option for the correlator. This produces a very large amount of output and should be avoided, or at least piped into a log file. The log information is outputed only to the standard output of the computer and not standard error.

An example of running the correlator (using LAM) via the commandline on all the available processors is:

```
mpirun C -O mpi_swc -T 60 configs/v112b.config,
```

which tells the correlator to skip the first 60 seconds of data, and use the configuration detailed in the configs/v112b.config file. The details of this file (and the files it uses) are listed below.

The directory from where the correlator is executed is also the directory where the files are stored. There needs to be a directory called `output` and files for each of the baselines which detail the delays. The delay files are labelled `delay_BL0.dat`, `delay_BL1.dat`, etc for each baseline (see below for how to generate). The correlator will output the raw (small integration) files into the output directory, and will name each of them with a baseline number, job number and bin number (if running in pulsar mode). The directory will also contain a file called `timestamps` which has the details for each file in the output directory, including the observation time of the file.

Post-processing averaging is performed using the utility `corrave` which has the usage:

```
corrave file_average filelist
```

Corrave can only average the full number of channels given in the files, partial channels processing is not available, so selecting the correct setting in the correlator configuration file is important. The `file_average` is the number of files to average together, and the `filelist` is a file containing all the files that are to be averaged. It is recommended that these averages are performed in a subdirectory

of the directory the correlator was executed in. The output of corrave is two sets of files. The first file is called `averages` and this file contains the full average of the entire data set, it has the time, from the start of the average, for each point, plus the calibrated amplitude and phase information. An example of the `averages` file is (Note: the first line normally doesn't exist, it is shown to aid explanation):

```
Time  Amplitude Phase(rad) Phase(deg)  Real    Imaginary
56.5  2.27923   -0.343608  -19.6873    2.146   -0.767842
168.5 2.29782   -0.340269  -19.496     2.16607 -0.766874
280.5 2.32135   -0.359679  -20.6081    2.17281 -0.817054
```

The other set of files is linked to each line of the `averages` file. It includes the information for each time averaged file, before it is also frequency averaged. This allows access to a bandpass for each of the lines in the `averages` file. The files are named `timeave_0`, `timeave_1`, etc and have the format listed below:

```
channel amplitude phase(rad) phase(deg) real imaginary
```

The last channel in the list is the Nyquist channel, and is not included in the `averages` file.

# Correlator configuration files

There are three main configuration files for the correlator. The first details the main settings the correlator is to use when processing data, with the other two defining where the data are and what the source positions are. Comments can be included by using # as the first character of a line.

## Main correlator configuration file

Listed below is an example correlator configuration file. It is derived from the configuration file that is used for the Australian Long Baseline Array (LBA) correlator. There have been a few additions and modifications made to enhance the flexibility of the software correlator. Most of the fields have been commented in the example below to help identify their purpose. The correlator does not yet use all these fields, but they have been included as they shall be used in future versions.

```
# Swinburne SWC EXPERIMENT FILE
EXPERIMENT V112B

# Reference Date, day number
DATE 2001 084

# sources file
SOURCES v112b.sources

# data files
DATA v112b.data

ANTENNAS 2
# NAME  TSYS    Delay(secs)   Delay(usecs)  at T0(secs)  rate(us/s)
PKS     25.0 0  0.0           0.0           0/00:00:00   0.0
DSS43   20.0 0  13.0          -2.338        0/08:30:00   3.639e-07

# Freq (MHz) Bandwidth (+'ve for USB, -'ve for LSB)
FREQ/BW  2290.00  +16

#Baselines (ALL, 1*, 2*, 1-2, 2-3, etc)
BASELINES ALL

#Channels (512 channels yields 1024 lags)
CHANNELS 64

#Products (AC = auto correlation, CC = cross correlation)
#         (L = left poln, R = right poln, none = off)
CC LL

#Polarizations / Frequency (2P = 2 poln, 2F = 2 freqs)
POLN/FREQ 2P

#Integration length (seconds)
INTEGRATION 1.0

#output filename (/dev/null for none, relative to starting directory)
OUTPUT /dev/null

#ADVANCED USE ONLY !!!!! Could cause unwanted results
BUFFERSIZE 32000
STEPSIZE 32000

#Scan Source TimeRange
   1 0826-373 0/08:00:00 - 0/08:15:00
   2 J0835-4510 0/08:15:00 - 0/11:00:00
```

**EXPERIMENT** is the name of the experiment being processed - currently unused.

**DATE** is used to determine the year and day number of that year that the observation occurred on. This field is given in universal time (UT). All other times are given relative to the start of this day.

**SOURCES** is the names the file that contains the information about the sources (see below).

**DATA** contains the name of the file that has a list of where to find the files to be used by the correlator (see below).

**ANTENNA** section covers the details for all of the antennas used in the observation. Firstly it has the number of antennas, followed by the details. Each line contains, the antenna name (as defined in a system wide correlator file), the telescopes system temperature (one for each polarisation), the gross delay to apply to each telescope (seconds). The delay to apply to each telescope due to offsets in the maser clocks (microseconds), and at what time in the observation (relative to the start of the day) the offset appears, followed by the rate at which the maser clocks are drifting (microseconds/second). The last three fields define the clock errors that are due to the local maser at each telescope.

**FREQ/BW** defines the frequency of the observation, and the bandwidth. For example, `2290.0 +16` is 2290 MHz centre frequency, with 16 MHz of bandwidth. The + before the 16 indicates the data is upper side band. That is the real local oscillator frequency (LO) is $2290.0 - \frac{16}{2} = 2282.0$MHz.

**BASELINES** is used to define which baselines of the observation are to be correlated. Currently this defaults to all and the use of anything else will not change the results. If less baselines are required the antennas that are not needed can be commented out, or the extra baselines can be correlated and not processed any further.

**CHANNELS** defines the number of frequency channels the correlator will produce. This number is half of the number of lags, that is 512 channels produces 1024 lags.

**PRODUCTS** define the multiplications to be performed on the data. AC is for Autocorrelation, and CC is Cross-correlation. L is for the left (or first) polarisation, and R for the right (second) polarisation in the data. Currently there is no method of joining these two polarisations together if the data was for the same polarisation but two different frequencies.

**POLN/FREQ** is used to indicate if the data is two separate polarisations, or two separate frequencies. Currently the only option is 2 polarisations (2P).

**INTEGRATION** is used to determine how long in time the output data should be integrated. This option is currently not used.

**OUTPUT** is to determine where the output data should be stored. It is not an active option, and the output data will automatically be stored in a directory called `output` relative to where the correlator is executed.

**SCAN** section details the observations that are in the data files. The scan number is to be used for future versions when the correlator outputs files in the FITS format. The source field is looked up in the SOURCES file to obtain the position of the source, and the time-range is used to locate the source in the data files.

**ADVANCED SECTION** deals with two features of the correlator. The first, BUFFERSIZE, is the size of the data buffer to be sent to each node for processing. This number MUST be larger than the delay between the telescopes. The STEPSIZE is for determining how far to move forward in the data to get the next section of data. If BUFFERSIZE and STEPSIZE are the same then the correlator does not use the same data for the correlation steps. If STEPSIZE was to be set at half the size of BUFFERSIZE, then the correlator would be performing twice the number of correlations, with each correlation including half of the data of the previous correlation. It is strongly recommended that these not be changed, they are set at 1ms of data for S2 16MHz data-rates.

## Source Configuration File

The source configuration file defines the positions of each source used in the main configuration file. Additional sources can be listed so that a single source file could be used for all observations.

```
0826-373    08:28:04.78026  -37:31:06.2812  J2000
J0835-4510  08:35:20.605    -45:10:34.83    J2000 ; VELA_PSR
```

## Data Configuration File

The data configuration file lists where the data can be found for each of the telescopes. The file is divided into columns, with each column defining the files for a single telescope. As the data rates need to be the same for each telescope, there must be the same number of files for each telescope. If this is not possible, then you can comment out the lines as needed to allow for different observations and different baselines. The first line in the file indicates which telescope the data file belongs to, with the other lines being the data file locations.

```
PKS     DSS43
# 0826-373
data/pks/01-084-08:12:00.psr   data/tid/01-084-08:12:00.psr
data/pks/01-084-08:13:00.psr   data/tid/01-084-08:13:00.psr
data/pks/01-084-08:14:00.psr   data/tid/01-084-08:14:00.psr
# VELA_PSR
data/pks/01-084-08:15:00.psr   data/tid/01-084-08:15:00.psr
data/pks/01-084-08:16:00.psr   data/tid/01-084-08:16:00.psr
```

## Generation of Delay files

Due to the correlator not yet having its own internal method for calculating the delays of given sources and baselines, it is necessary to generate a delay file that can be input into the correlator. This is because the generation of delays can be done using a programming library called CALC. At the time of writing CALC was not available for the linux architecture (which is what the software correlator is designed to run on), however it was under development. After it becomes available the software correlator will be modified to support this new method and this step will no longer be necessary. The current method is to login to the Australia Telescope National Facility (ATNF) correlator computer (login information changes and will not be detailed here, please contact someone at the ATNF for information). Once logged in you need to do type the following:

```
toca
run ctest
```

`ctest` is a modified version of wtest, which was written by Warwick Wilson from the ATNF. This program generates the delays that the correlator needs using CALC routine calls. The machine is a VAX/VMS machine that is used to control the Long Baseline Array Correlator at the ATNF.

`ctest` will ask for some information, and give examples, then generate the delay files. These delay files need to be put in the directory the correlator is executed in.

The information the `ctest` requires is:

source name,

station 1,

station 2,

date in the format YY MM DD HH MM (two or three digit years starting with 00 = 1900, 2003 = 103, etc),

time increment (in seconds)

number of intervals (of time increment)

The name of the output file is `FOR009.dat`. This file needs to be renamed to the baseline with which it is associated. That is if it is for the first baseline, that is baseline 0, call it `delay_BL0.dat`. Any other name and it will not be recognised or processed correctly.