Towards A Pulsar Virtual Observatory

A thesis submitted to The University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Engineering and Physical Sciences

2007

Michael John Keith

School of Physics and Astronomy

Contents

List of Figures

List of Tables

List of Code Examples

The University of Manchester

ABSTRACT OF THESIS submitted by Michael John Keith
for the Degree of Doctor of Philosophy and entitled
Towards A Pulsar Virtual Observatory. September 2007

In the 40 years since the serendipitous discovery of the first pulsar, there have been continuing advances in the technology and techniques of searching for pulsars. This thesis presents another iteration in the development of ever improving search techniques.

A new graphical candidate selection tool, JReaper, has been created, improving on a previous project. This tool includes advanced new features such as a pulsar candidate scoring system, which can be used to automatically distinguish between real pulsar signals and interference. Application of this new tool to the millions of candidates produced by the Parkes Multibeam Pulsar Survey has lead to the discovery of 30 previously unknown pulsars. Amongst these pulsars is the eccentric binary system J1753−2243, which has an orbital period of 13.6 days. The pulsar in this system has a rotational period of 95 ms and appears to be mildly recycled with a characteristic age in excess of 200 Myr. The companion's mass is constrained by geometry to lie between 0.48 and $\sim 2.5\ M_\odot$. Given the probable system evolution, it is likely that this system is the tenth known double neutron star system. J1849−0020, also discovered during this work, has a characteristic age of 67 kyr and is spatially coincident with a known supernova remnant. Measurement of the distance to the pulsar and remnant agree within the errors, and the separation of the pulsar from the remnant centre is consistent with the typical pulsar birth velocity.

The Perseus Arm Pulsar Survey is a new survey for pulsars that extends the successful Parkes Multibeam Pulsar Survey away from the Galactic centre. Anal-

ysis of this survey has been carried out at Jodrell Bank Observatory and the Osservatorio Astronomico di Cagliari, resulting in the discovery of 10 previously unknown pulsars. J0623+0341, discovered in this survey, exhibits a nulling phenomenon with bursts of pulses lasting ~ 1 to ~ 120 seconds, separated by nulls of length ~ 1 to ~ 60 minutes. This discovery starts to fill in the phase space gaps between intermittent pulsars, RRATS and classical nulling pulsars. Another pulsar, J0627+0705, shows an inter-pulse at a pulse phase of 180°. High resolution observations at 1400 MHz using the Parkes telescope show strong polarisation in both pulses. By modelling the polarisation using a rotating vector model, the geometry of the pulsar is determined. This is one of the very few pulsars where this is possible with high precision.

The pulsar search process is ideal for running in a grid environment, as each task is independent of parallel tasks. The Pulsar Virtual Observatory (PVO) project has been created with the aim of adapting pulsar search processing for a grid environment. To further this aim a queueing system for arbitrary grid systems, named the Abstract Queueing Architecture (AQuA) has been developed. The PVO system also aims to be able to estimate ahead of time the amount of time that is required to process the given tasks, therefore the AQuA system is designed to provide guaranteed resource even when the grid does not support it.

Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Data is a precious thing and will last longer than the systems themselves.

Tim Berners-Lee

Technology is a word that describes something that doesn't work yet.

Douglas Adams

Dedication

For my brother.

Acknowledgements

I would like to thank Andrew Lyne and John Brooke for giving me the opportunity to work on this project, and for their insight and guidance during the last three years. I also would like to thank Michael Kramer for his support in all aspects of this work.

I thank Bruno Harbulot for his work on the Pulsar Virtual Observatory project and Ralph Eatough for his helpful bug reports and numerous cups of tea. Thanks also to Duncan Lorimer and Maura McLauglin for their help with assorted pulsar problems and to Graham Smith for inspiration and proof reading.

I would also like to thank all the engineers and technical staff who have enabled the collection of high quality data both at Jodrell Bank and at the Parkes Observatory. In particular I would like to thank the following: Christine Jordan, for her help understanding the telescopes and data at Jodrell Bank, John Reynolds, for his excellent facilities at Parkes and John Sarkissian, for making the long nights at Parkes less boring. I would also like to thank the many pulsar astronomers that I have had the opportunity to learn from around the globe, to name just a few: Simon Johnston, Andrea Possenti and Marta Burgay. Finally, I would like to thank all my family and all the people who have made three years living at Cheshire Hunt bearable, in particular: Szymon Kozłowski, Ralph Eatough, David Champion and Sanhita Joshi. LGH.

This thesis was composed using vi and typeset with LaTeX.

The Author

Michael Keith was born in Dundee, Scotland on the 1st of June 1983. Having grown up in the north east of England, he was awarded the degree of Bachelor of Science in Physics with Computer Science from the University of Durham. Michael then eagerly took up the opportunity to work at Jodrell Bank Observatory and study for the PhD for which this thesis is written.

Figure 1: Radio Telescopes at Jodrell Bank Observatory

1

Introduction

1.1 Pulsars

Pulsars were first detected as localised periodic signals in radio observations. This pulsed emission is now believed to be caused by rotating neutron stars, with emission beamed along powerful magnetic field lines. Since the initial discovery, pulsar-like objects have been seen at nearly all electromagnetic wavelengths, from metres to picometres.

1.1.1 Background

In the summer of 1967, a survey for radio sources undergoing rapid scintillation from the interplanetary medium was carried out by Hewish et al. (1968) at the Mullard Radio Astronomy Laboratory, Cambridge. During these observations a signal was detected at a fixed celestial position, with a highly precise periodicity of 1.337 seconds. This was the first reported detection of a pulsar, and soon more such sources were discovered (Pilkington et al., 1968), including one with a periodicity of just 0.253 seconds.

The nature of these sources was linked to neutron stars when Gold (1968) wrote that the theoretically postulated neutron star was the only type of star with an equation of state capable of rotating or pulsing with such short periods. In fact, it was shown that periods of less than one second could not be the fundamental mode of any pulsing star. Since the periodicity detected was certainly from the fundamental period of the source, the physical manifestation of pulsars was tied to neutron stars. Since this initial discovery, pulsars have been detected with periodicities of the order of a few milliseconds. At the time of writing the pulsar with the shortest known period is located in the globular cluster Terzan 5, labelled PSR J1748-2446ad, with a rotational period of 1.39 ms (Hessels et al., 2006).

Observational pulsar studies are mainly split into two groups, namely searching and timing. Searching is the practise of discovering new pulsars, whereas timing is a means to precisely measure the rotational properties of a pulsar. This thesis mainly concentrates on the search for new pulsars, although all newly discovered sources undergo a nominal period of timing to accurately measure the basic parameters for population studies, as well as determine which individual pulsars are most useful. Pulsar timing is based upon the measurement of the time-of-arrival (TOA) of a pulse from the pulsar. By fitting a parameterised model to the measured TOAs, a very accurate model of the rotation of the pulsar can be computed.

These highly precise models of the rotational properties of pulsars allow us to study a multitude of different areas of fundamental physics and astrophysics. Due to their nature, pulsars tell us a lot about the extreme physics in the neutron star interior (e.g. Stairs et al., 2000), as well as in the pulsar magnetosphere (Ruderman and Sutherland, 1975). Binary pulsars can allow us to measure the masses of pulsars, which can put constraints on the equation of state of neutron stars. By constraining these models, we can get some idea of what kinds of particles are formed in the ultra-dense cores of these degenerate stars (Thorsett and Chakrabarty, 1999). Pulsars in binary systems can also provide precise tests of gravitational theories such as general relativity, as well as probe the atmosphere of the companion star. The double neutron star system B1913+16 shows orbital decay due to the loss of energy from gravitational wave radiation, providing direct evidence of the existence of gravitational waves and agrees with theoretical predictions better than 0.5% (Taylor and Weisberg, 1989). Even more precise tests of gravitational theories are provided by the double pulsar system J0737-3039A/B, for which so many orbital parameters are measurable that at the time of writing the masses of the system are constrained to better than $0.04M_\odot$ and the agreement with general relativity is better than 0.05% (Kramer et al., 2006).

Precise measurement of the pulsar position through timing allows us to measure the gravitational potential that the pulsar lies in, as well as put constraints on supernovae, the birth mechanism of pulsars. Measurement of the position of pulsars within globular clusters can be combined with the Doppler shift of the signal as it falls in the gravitational potential of the cluster. These measurements can allow us to obtain information about the masses contained in the cores of these clusters. The measurement of accurate proper motions of a statistical sample of pulsars allows us to constrain models of supernova explosions, which must be able to reproduce the observed velocity distribution (Hobbs et al., 2005). Associating young pulsars with remnants also helps to classify the types of supernova, as well as provide new information on formation mechanisms. The accurate measurement of positions through direct observation and timing can also provide information for linking different celestial reference frames.

1.2 Observational Properties of Radio Pulsars

For most of this thesis, the search for pulsars is the search for a characteristic signal. For search operations, a pulsar can be defined by its periodicity, its pulse profile, its dispersion measure and effects caused by binary motion. When performing long term timing of sources, further parameters are detectable, including the accurate position of the pulsar, its spin-down rate and any proper motion. Each of these parameters, and how they affect observations, are discussed below.

1.2.1 Periodicity

The most basic property of a pulsar is the period of rotation, and hence the periodicity of the signal. Pulsars are very precise rotators, the error in measurement of a typical pulsar period is less than 1 part in 10^{10}. This precision is an important indicator that any detected signal is indeed a pulsar, as man made radio frequency interference is often much less precise and the period could fluctuate over a number of hours. Unfortunately this fluctuation may not be significant over the length of a single observation.

1.2.2 Pulse Profile

The variation of flux as a function of phase is known as the pulse profile. These profiles vary considerably between pulsars, and even between different observation frequencies (Komesaroff, 1970). However for a given pulsar, the time averaged pulse profile is very stable (Helfand et al., 1975). This stability allows for precision timing by using the average profile as a template to precisely measure the pulse arrival time. Some pulsars show changes in profiles, switching between two or more stable profiles (Backer, 1970a). For these sources precise time-of-arrival measurements are more difficult to obtain, although generally these effects can be dealt with by using more complex pulsar models.

1.2.3 Dispersion Measure

As the signals from pulsars travel between the source and the earth they are affected by the interstellar medium they pass through. One of the effects of this is known as dispersion and causes the lower frequency signals to be retarded in time compared to higher frequency signals (Hewish et al., 1968). Therefore over a given band, the pulse arrives first in the high frequency part of the band and later at the low frequency part of the band. This time delay can be described by,

$$\Delta t = 4.148808\text{ms} \cdot \left(\frac{1}{f_{lo}^2} - \frac{1}{f_{hi}^2} \right) \cdot DM, \quad (1.1)$$

where f_{lo} and f_{hi} are in MHz and DM is the dispersion measure (Manchester and Taylor, 1972). The dispersion measure is a property of the source, and is defined as

$$DM = \int_0^d n_e dl, \quad (1.2)$$

where dl is in the direction of the pulsar, d is the distance to the pulsar and n_e is the electron density of the interstellar medium. Dispersion measure is usually given units of cm^{-3}pc.

Therefore, when observing a pulsar with a non-zero DM, the pulse is widened because of the finite bandwidth of the receiver system. This effect can be eliminated by coherent de-dispersion of the signal, or reduced by using a multichannel filter-bank. See section 1.5.2 for more details on de-dispersion.

1.2.4 Binary Motion

Whilst the majority of normal stars are in multiple systems, the formation mechanisms for pulsars are believed to produce a large kick that can eject the pulsar from any orbit. However a small fraction of pulsars are found in binary systems, either due to small and/or fortunately directed kick velocities, or possible capture by another companion in an active stellar environment.

Pulsars that remain in binary systems have additional complications to observation, due to the binary motion. For non-relativistic binaries, this motion can be characterised by the five Keplerian parameters, namely: Projected semi-major axis $x = a\sin(i)$, eccentricity e, orbital period P_b, longitude of periastron ω_0 and time of periastron T_0. Since the line-of-sight velocity to the pulsar is not constant, the pulse arrival times need to be corrected for the light-travel time across the orbit. The time in the Earth frame t_e can be related to the time in the pulsar frame t_p by adding the Römer delay:

$$t_p = t_e + a \sin i \left\{ (\cos E - e) \sin \omega + \sin E \sqrt{1 - e^2} \cos \omega \right\}, \tag{1.3}$$

where $\omega = 2\pi/P_b - \omega_0$ and E is the eccentric anomaly (Blandford and Teukolsky, 1976). This is related to time t by Kepler's equation:

$$E - e \sin E = \frac{2\pi}{P_b}(t - T_0). \tag{1.4}$$

This cannot be analytically solved for $e \neq 0$, although a solution can be obtained by numerical iteration. The correction in equation 1.3 is then used in the timing model in order to fit for these parameters.

For systems with short enough orbital periods and high enough masses, relativistic effects are also important, which are described by a series of 'Post-Keplerian' (PK) parameters (Robertson, 1938; Peters, 1964; Blandford and Teukolsky, 1976; Damour and Deruelle, 1985, 1986). The most significant PK parameter is usually $\dot{\omega}$ which describes the advance of periastron. This effect is measurable only in systems with non-zero eccentricity, as ω_0 is not defined for purely circular orbits. In general relativity, $\dot{\omega}$ is related to the total system mass by:

$$\dot{\omega} = 3T_\odot \left(\frac{P_b}{2\pi} \right)^{-5/3} \frac{1}{1 - e^2} \left(m_p + m_c \right)^{2/3}, \tag{1.5}$$

where $T_\odot = GM_\odot/c^3$. Care has to be taken when measuring this parameter for non-compact and non-rigid bodies as there may be periastron advance caused by tidal effects.

All of the PK parameters depend on the masses differently, therefore measurement of any two allows for a solution for the masses of the pulsar and companion. Discussion of other first and higher order PK parameters is outside the scope of this thesis, and so will not be included here.

When searching for new pulsars, none of the orbital parameters are known, and correcting for such a wide parameter space is impractical because of the amount of available computing power. However search observations that are less than approximately half of an orbit can approximate the changing line-of-sight velocity to a polynomial change in time. This simplifies the time adjustment to a simple polynomial expression.

$$t_p = t_e + \frac{at_e^2}{2c} - \frac{\dot{a}t_e^3}{6c}, \tag{1.6}$$

where a and $\dot{a}$ are the line-of-sight acceleration and derivative of acceleration of the pulsar at a given epoch.

This relationship can be used to transform the observed time-series into the pulsar rest frame for a given acceleration and acceleration derivative. More details on the methodology of this will be presented in Section 1.5.

1.2.5 Position and Proper Motion

Analogous to binary motion of the pulsar is the motion of the Earth in the solar system. The long orbital period of the Earth makes this motion insignificant for single observations, however for timing observations the yearly cycle is easily detectable. To correct for this motion, the observing time-frame is shifted to the barycentre of the solar system, using the Römer delay as discussed in Section 1.2.4. This model is highly sensitive to the position of the pulsar on the celestial sphere. Therefore by fitting these parameters, one obtains a pulsar position much more accurately than given by the telescope beam size.

1.3 Derived Physical Properties of Pulsars

Many physical properties of pulsars can be estimated by modelling pulsar spin down as due to energy loss by magnetic dipole braking (see for example Lyne and Smith, 2005).

The rotational energy loss of a pulsar is dependent on the moment of inertia, I, and the spin frequency, ν and frequency derivative, $\dot{\nu}$, and is given by

$$\dot{E} = -4\pi^2 I\nu\dot{\nu}. \tag{1.7}$$

If we assume that all the rotational energy loss is due to magnetic dipole braking, with a dipole moment $\mathbf{m}$, then we can write:

$$\dot{\nu} = -\left(\frac{8|\mathbf{m}|^2 \sin^2\alpha}{3\pi^2 I c^3}\right)\nu^3. \tag{1.8}$$

However there may be other causes of rotational energy loss, for example mass flowing out from the polar regions, therefore we can express this in a general form:

$$\dot{\nu} = -K\,\nu^n. \tag{1.9}$$

The parameter n is known as the braking index, and for pure magnetic dipole braking is 3.

1.3.1 Characteristic Magnetic Field Strength

By assuming a braking index of 3, the value for pure magnetic dipole braking, we can compute a 'characteristic magnetic field strength' at the pulsar surface. Since $B \approx |\mathbf{m}|/R^3$, we can use equation 1.8 to write an equation for the surface magnetic field strength as:

$$B_{surf} = \sqrt{\frac{-3c^3}{8\pi^2}\frac{I}{R^6\sin^2\alpha}\frac{\dot{\nu}}{\nu^3}}. \tag{1.10}$$

More commonly, this is written by substituting some typical parameters and using $\dot{\nu}/\nu^3 = -P\dot{P}$.

$$B_{surf} \simeq 3.2\times 10^{15}\sqrt{P\dot{P}}\ \text{Tesla}. \tag{1.11}$$

This measurement gives a good approximation to the magnetic field strength at the surface, however the radius, moment of inertia and the inclination angle of the field are not well known, so performing accurate measurements of the magnetic field strength this way is impossible.

1.3.2 Characteristic Spin Down Age

If one rearranges equation 1.9 and rewrites in terms of period, one can integrate from the pulsar birth time to the current time to give a measurement of the pulsar age.

$$\int_0^T dt = \frac{1}{K}\int_{P_0}^{P} P^{(n-2)} dP, \tag{1.12}$$

and therefore:

$$T = \frac{P}{(n-1)\dot{P}}\left[1 - \left(\frac{P_0}{P}\right)^{(n-1)}\right]. \tag{1.13}$$

If we assume that $n = 3$ and $(P_0/P)^2 \ll 1$, we can define a 'characteristic age' as:

$$\tau_c := P/2\dot{P}. \tag{1.14}$$

This value can be used to parameterise many of the characteristics of the pulsar, as well as give a simple estimate of the pulsar age.

1.4 Scientific Basis for Finding More Pulsars

Pulsars are valuable tools, both individually and as a population. Currently there are $\sim$ 2000 pulsars known, however this covers a wide variety of different types of sources. The discovery of new pulsars allows us to improve on the research already undertaken as well as give us the opportunity to explore new topics of research.

Outlined below are some of the different classes of pulsars, and a brief description of the science that can be done by identifying more of them. In addition, Figure 1.1 shows where these classes lie on the Period-Period Derivative, where pulsars are plotted according to their rotational period and the rate of change of period. In general however, it is impractical to search for a specific type of pulsar, and often the unique features of individual pulsars are only apparent after a period of timing.

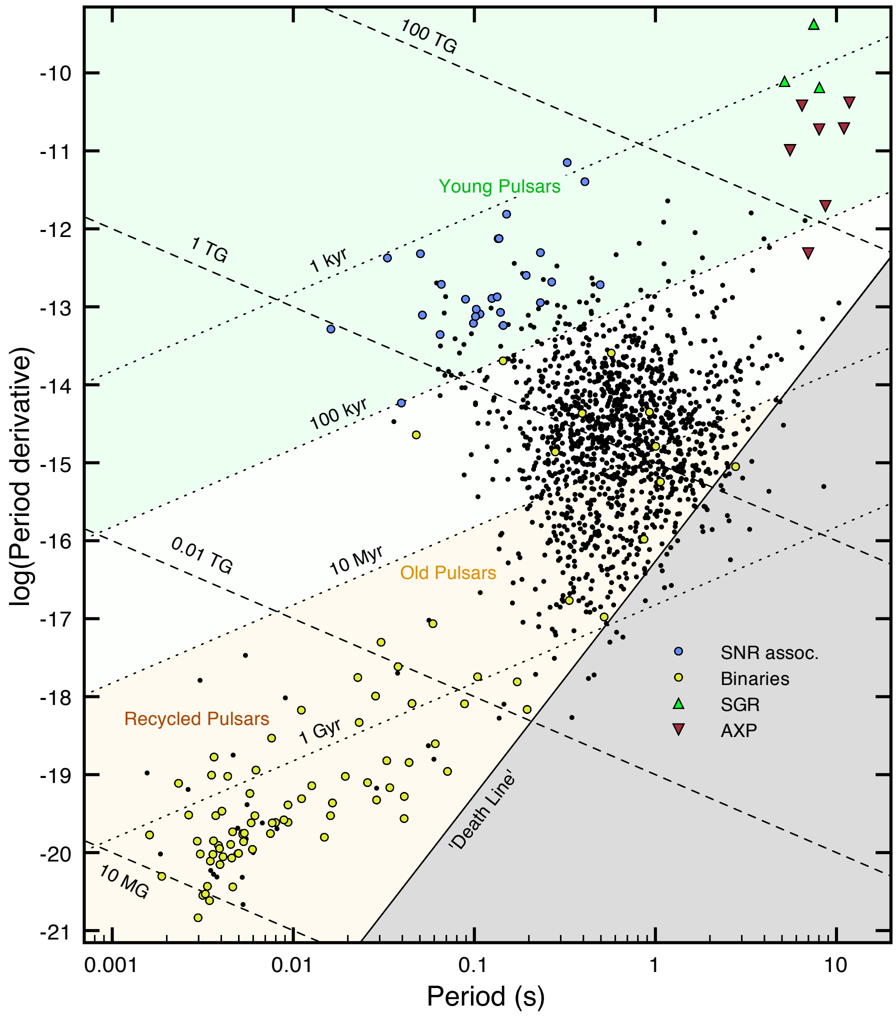

Figure 1.1: The Period-Period Derivative (P-$\dot{P}$) diagram. Each point on the diagram represents a known pulsar and is plotted according to its rotational period and the rate of change of period. The small dashed lines show lines of constant characteristic age, and the longer dashed lines show lines of constant characteristic surface magnetic field strength. The 'death line' is the boundary where the rotational energy loss is insufficient to power the standard radio emission mechanisms (Ruderman and Sutherland, 1975). The coloured regions are described in the text. Note that magnetic field strengths are marked in TG, where 1 TG = 1×10^8 Tesla.

1.4.1 Young Pulsars

Pulsars with ages less than about 100 kyr are usually considered to be young pulsars. These pulsars are rapidly spinning down and tend to suffer from erratic behaviour such as glitches and timing noise. Glitching is characterised by a sudden discontinuity in period and period derivative, and can provide insight into the internal structure of neutron stars.

Discovery of more young pulsars can add more constraints to the models for the super-fluid nuclear matter that is predicted to make up the interior region of neutron stars (e.g. Anderson and Itoh, 1975).

1.4.2 Middle Aged and Old Pulsars

Middle aged pulsars pulsars have characteristic ages between 100 kyr and 10 Myr, old pulsars are considered to have ages greater than about 10 Myr. These pulsars make up the major component of the discovered pulsar population. Old pulsars have long periods of the order of seconds, and are approaching the 'death line' region where the radio emission mechanism breaks down. The 'death line' marks the boundary of the region where the rotational energy loss is insufficient to explain the radio emission using simple models, although the brightness of most pulsars would fall below detectable thresholds before reaching this region. Detection of pulsars near to, or past the 'death line' can provide new information about the complex physics behind the radio emission of pulsars.

1.4.3 Recycled Pulsars

Dead pulsars that go through a period of accretion powered spin up are known as recycled pulsars (Bisnovatyi-Kogan and Komberg, 1974; Shibazaki et al., 1989). These pulsars can reach periods of the order of a few milliseconds, although if the accretion process is interrupted they may have longer periods. Because of their formation mechanisms, these systems are usually found in binary systems, with a white dwarf or neutron star companion. Typically the faster millisecond pulsars are found with white dwarf companions as these have slow evolution processes that allow for a long period of accretion. These pulsars have a low magnetic field strength, and spin down slowly. They are also very stable rotators, and so can be timed precisely, which makes them useful for many gravity experiments (see Section 1.4.5 on binary pulsars), as well as a pulsar time-frame.

By measuring how pulsar clocks deviate from Earth clocks, one can get a measurement of the stability of the Earth time-frame. By using the celestial positions of the sources, one can also provide data to correct the solar system barycentre model, as well as potentially measure a stochastic gravitational wave background (Foster and Backer, 1990; Jaffe and Backer, 2003). The errors in these calculations depend on the number of pulsars that are used, so the discovery of new recycled pulsars will allow for more accurate measurements (Jenet et al., 2005).

1.4.4 Intermittent Pulsars, Nullers and RRATS

Many pulsars show nulling-type phenomena where the radio emission stops for a length of time. The timescales of these nulls can range from a few pulse periods in the case of nullers (Backer, 1970b), to years in the case of intermittent pulsars. Nullers and intermittent pulsars tend to have emitting times of the same order as the nulling times, however RRATS emit for less than one pulse period, with nulls for as long as hours (McLaughlin et al., 2006).

These pulsars show some of the strangest emission characteristics, and by studying these sources we can discover more about the underling processes behind the radio emission of pulsars (Kramer et al., 2006).

1.4.5 Binaries

Pulsars in binary systems can open many new windows into different areas of physics. For compact binary systems, such as double neutron star systems like PSR B1913+16 (Hulse and Taylor, 1975) and the double pulsar PSR J0737-3039 (Burgay et al., 2003; Lyne et al., 2004), pulsars can provide unique tests of general relativity and other gravitational theories (Taylor and Weisberg, 1989; Kramer et al., 2006). This can be achieved by observing the changes in the pulsar signal as the pulsar moves in the gravitational field of the other star, as well as the effects of the photons passing the companion star.

Many of these effects can also be used to measure the masses in the system, which can put constraints on the equation of state of the nuclear matter that forms the cores of neutron stars (Podsiadlowski et al., 2005).

Binary systems can also be useful when the pulsar, or its emission, interacts with the atmosphere of the companion star. This can be used to probe the outer regions of giant main sequence stars (Johnston et al., 1992).

1.5 The Pulsar Search Data Analysis Work-Flow

Searching for pulsars is a complex and time consuming task, however it is necessary to continue to benefit from the often unique science that can be achieved with pulsars. This section outlines the basics of pulsar searching, discussing each element in the search work-flow, which is illustrated graphically in Figure 1.2. The remainder of this thesis discusses enhancements to this search process.

1.5.1 Extraction

Extraction is a catch-all name for the initial preparation of the raw data files. The data files as recorded directly from the telescope may be stored in impractical data formats. For example, data files recorded on a telescope in real time have the data from multiple telescope beams interleaved. However when processing the data it is convenient to split the data streams into different files, so that they can be processed separately.

Interference mitigation processes can also be performed during this step. For example, removing known bad frequency channels and checking for strong un-dispersed signals.

1.5.2 De-dispersion

As the pulsar signals are dispersed by the interstellar medium, the pulse arrival time is delayed by different amounts at each frequency and the delay is proportional to ν^{-2}. To counteract this, the telescope signals are recorded in many narrow frequency channels, and the effects of dispersion are reversed digitally.

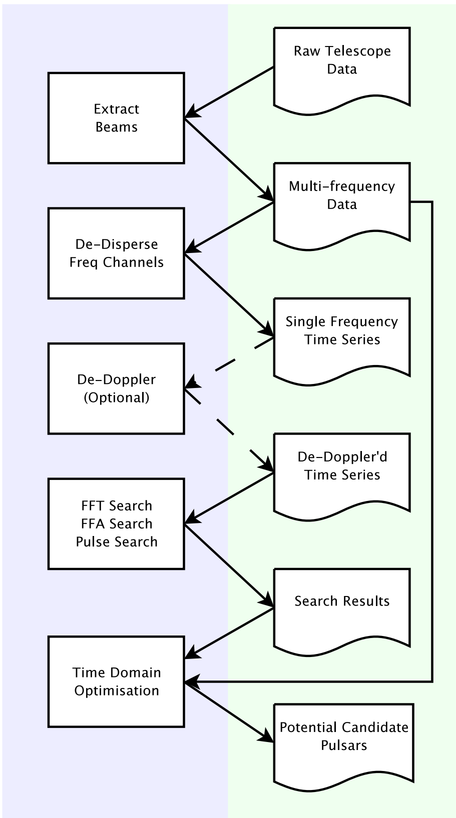

Figure 1.2: A typical work-flow for processing pulsar search data.

This process, known as de-dispersion, is simply a matter of applying a correcting time offset to each frequency channel and summing the channels.

However for search observations the amount of dispersion suffered by the potential source is unknown, and this process must be repeated with differing amounts of trial dispersion. This produces a large number of data streams, for which the subsequent processing must be performed independently.

It is possible to perform this de-dispersion in a highly efficient manner using a 'treeing algorithm' (Taylor, 1974). However this algorithm assumes that the time delay in each channel is linearly proportional to frequency. This approximation works for data where the total bandwidth is small, however for larger bandwidths the assumption is invalid and the sensitivity is reduced.

Future processing will most likely not be able to rely on this assumption as the bandwidth increases, but must de-disperse the data using more precise algorithms. In general, as the search algorithms become more complex, the de-dispersion step becomes less significant in terms of the total processing time. Therefore for future processing, the use of the treeing algorithm is less important.

1.5.3 Fourier Analysis

The Fourier transform is an ideal way to detect the constant periodic signals produced by the majority of discovered pulsars. The Fourier transform is a decomposition of a data series into a sum of sine and cosine waves. In this way an infinite sinusoidal wave will form a single infinitely narrow spike at the frequency corresponding to the period of the signal. More generally any periodic signal will form a spike at the fundamental frequency, plus a series of harmonics, with magnitudes depending on the shape of the input signal. By transforming the input time-series in this way, any pulsar signals in the data should be among the strongest signals in the Fourier spectrum, even if they are undetectable in the untransformed time-series.

Whilst the fundamental frequency of the pulsar is the strongest signal produced in the Fourier spectrum, pulsars with narrow pulses will have a lot of power 'lost' in the harmonics. By using a technique known as harmonic summing, it is possible to regain some of this 'lost' power. This is achieved by stretching the power spectrum by a factor of two, and summing it with the un-stretched power spectrum. In this way, the power from the second harmonic will be added on top of the fundamental power, increasing the signal by more than the $\sqrt{2}$ increase in noise (The increase depends on the amount of power in the pulse). One can then select the best signals from this new spectrum. This can be repeated to add more harmonics. However since the noise is also increased upon every addition, the signal-to-noise will reach a maximum at a harmonic fold depending on the pulse width. Therefore it is important to keep the results from each harmonic fold used, and the harmonic fold that a pulsar was detected strongest in can give an indication of the width of the pulsations.

Since the harmonic summing uses the power spectrum, it is possible that the process will boost the signal-to-noise ratio for signals that are not truly harmonically related. To prevent this, a process called profile reconstruction is used. This method takes the original Fourier spectra and performs a short transform of just the harmonic bins. For truly related signals, the harmonics, when transformed in this fashion will align in phase and give a large amplitude peak value. Noise signals that were in harmonically related bins of the power spectrum by chance will not align in this way and so the signal strength will be low.

The principal algorithm used to perform Fourier domain searches is the Fast Fourier Transform (Cooley and Tukey, 1965). This algorithm computes a discrete Fourier transform in $\mathcal{O}(N \log_2 N)$ operations, for transform length N. The improvement over a traditional DFT algorithm is immense, especially when the number of data points is large. For example, with the number of data points at 2^{23}, a typical observation, the computation time is reduced by a factor of 10^5.

1.5.4 Fast Folding Algorithm

The search data are dominated by red noise due to fluctuations in the background levels and the receiver system. This reduces the effectiveness of FFT based search methods at long periods, therefore most FFT searches enforce a period cutoff at around 5 seconds.

To search for longer periods, one can perform repeated folding of the time series at a range of trial periods. Traditional folding methods are highly inefficient, however it is possible to perform many such folds efficiently using a 'Fast Folding Algorithm' (FFA) (Staelin, 1969). The FFA uses a similar treeing approach to the FFT, reducing the number of operations from $\mathcal{O}(N^2/n)$ to $\mathcal{O}(N \log_2 N/n)$, where N is the number of data samples in the time-series and n is the number of data samples in the output profile.

1.5.5 Single Pulse Searching

The flux density of individual pulses can vary considerably, and it is possible to detect many pulsars by the detection of individual strong pulses. In addition to this, there are classes of pulsar-like objects that have been shown to show so few total pulses that standard periodicity searches cannot detect them.

The detection of single pulses relies on the pulses being highly dispersed relative to terrestrial burst emission. Pulsars can be detected by one or more dispersed pulses that are significant above the general noise, and not associated with any un-dispersed terrestrial interference (Cordes and McLaughlin, 2003).

1.5.6 Binary Motion Compensation

Binary pulsars have an apparent period drift due to the Doppler effect. If the binary period is of the same order as the observation time, the changing period will mask the signal in the Fourier domain. This is due to the power being distributed in different frequency bins.

These effects can be effectively removed from a time-series before processing, however to completely remove all the effects the binary parameters must be known. There are unfortunately too many parameters to perform a search for all possible orbits in a reasonable time scale. By expanding the orbital motion as a polynomial, it is possible to approximate the orbit by using acceleration and jerk[1] (or $\dot{P}$ and $\ddot{P}$ where P is the pulsar period). By using this approximation, searching for binary pulsars can be done with only one or two extra parameters. There are three typical methods for removing this acceleration from the search process, either by re-sampling the time-series (Middleditch and Kristian, 1984), using incoherent methods (Faulkner et al., 2004), or by using matched filters in the Fourier domain (Ransom et al., 2002).

For the time domain re-sampling At each trial acceleration and jerk, the time-series must be re-sampled with a new sample rate that varies with time. This adjusted time-series must then be searched in the standard way, for example with FFTs or FFAs. This method is relatively simple and once the time-series is adjusted, all the remaining processes operate as normal. Unfortunately this process is computationally expensive as the full FFT must be computed at each acceleration and jerk step.

For the incoherent search The number of bins which the pulse power is spread across for a given orbital acceleration is proportional to the observation time. Therefore this spread can be reduced by splitting the observation into n smaller segments. The signal strength in each part is however reduced by a factor of $1/n$, and due to the binary motion it is not possible to phase align the Fourier spectra. Therefore the signal can only be partly recovered by incoherently adding the power spectra. This reduces the overall signal strength by $1/\sqrt{n}$, and precludes any profile reconstruction. Even though this methods loses a significant

[1] Jerk is used in this document as a convenient contraction for ‘rate of change of acceleration’.

amount of signal, it is very fast, and so has been used many times in the past.

For the matched filters When using matched filters, the time-series is Fourier transformed as normal, however the resulting spectra are then analysed by convolving with the known response from each trial acceleration and jerk. This methods does not require repeated Fourier transformations, and so can be quicker than the time domain methods, however it is only suitable for the FFT search routines and requires generation of large numbers of filters.

The above approximations are only valid when the orbital period is greater than the observation length by a factor of two or more. Therefore when the observation length is long, or when looking for very fast binaries, another approach must be taken.

1.5.7 Time Domain Optimisation

In general each of the search algorithms simply produces a list of signal-to-noise ratios and periods at each trial dispersion measure, acceleration or jerk step. To manage such a large parameter space, signals are grouped together by period, keeping information about which dispersion measure, acceleration or jerks it was detected in. Unfortunately further processing is required to distinguish suspects that are pulsars from those that are noise and interference. This additional processing usually comes in the form of optimisation of suspects in the time domain.

The most basic time domain function is folding of data. This process steps through the time-series adding each sample to an appropriate output bin depending on the computed pulse phase at that moment. To fold data correctly requires more accuracy in the period than is provided by the finite width bins of a Fourier analysis. Therefore it is common to split the time-series into a number of sub-integrations, and fold those separately. With the shorter folds, the pulse will drift less in phase and therefore have less broadening. This also allows for the measurement of the variation of signal over the observation, providing useful clues for identifying pulsars.

Summing the sub-integrations is equivalent to folding the entire time-series. Shifting each sub-integration in phase before summing can compensate for errors in the initial folding period. By shifting and adding the sub-integrations in this way, one can perform an approximate folding of the time-series at a number of different periods in a computationally efficient manner. In a similar fashion, using quadratic or cubic shifts with time can compensate for orbital acceleration and jerk. When the offsets correspond to the correct pulse period, etc. the power from the pulse should all fall in a minimum set of bins. The correction of the sub-integrations increases the effective signal strength, and therefore the signal-to-noise ration can be maximised. The way that the signal strength varies with period shows how well aligned in phase the pulses are, and so this information is useful when distinguishing highly accurate pulsar signals from man-made interference.

1.6 Problem Specification

1.6.1 Present System for Access to Data Archives

There is a large set of pulsar data that has been released to the public domain, however much of this data set is not used outside the institutions that recorded the data initially.

Whilst the data are publicly available, the current methods of data access do not lend themselves to easy access.

The main issues with the current data access mechanisms are:

- No clear centralised meta-data archive
- Data files are large and difficult to manage

- Data is in proprietary formats and software is difficult to configure and install

The lack of a centralised archive means that there is no obvious starting point in searching for data, and that individuals at institutions must be contacted when attempting to access data. The large size of data files has been a problem, however this is now becoming more and more redundant with the increasing network bandwidths. Unfortunately the software that analyses the data are generally written 'on-demand' for specific tasks, and therefore it is likely that the software remains difficult to use.

It is my belief that to increase use of pulsar search data archives, and therefore to increase the amount of scientific collaboration, the following is necessary:

- Development of a centralised meta-data catalogue. This will allow prospective users to search for the data they require and determine which archives the data are present in, and the data access policies for this data.

- Development of a standard format for pulsar search meta-data. It seems that it is unlikely that all search software will decide upon a single data format in the near future, however it should be trivial to store information about the data in a standard format. This should allow users to easily interpret the data.

- Development of a Pulsar Virtual Observatory, which allows users to process data without configuring and installing search software locally.

1.6.2 Aims for The Project: Towards a Pulsar Virtual Observatory

It is outside the scope of this thesis to deal with all the issues outlined in the previous section, however this project aims to make progress in one or more of these areas.

The aims for this project are outlined below:

- Investigate current methods of searching for new pulsars.
- Determine means for applying pulsar search work-flows to grid computer systems.
- Develop a prototype system for scheduling pulsar tasks in a grid environment.

1.7 Document Outline

The remainder of this document is split into the following chapters:

- **Chapter 2** presents JReaper, a new graphical tool for analysing candidate pulsars.
- In **Chapter 3**, JReaper is applied to candidates produced during a previous processing of the Parkes Multibeam Pulsar Survey. Analysis of the effectiveness of the results and discussion of the discovered pulsars is included.
- **Chapter 4** discusses the Perseus Arm Pulsar Survey, a new survey for pulsars in the galactic plane. Describes the observations and analysis methodology.
- **Chapter 5** describes the results from the Perseus Arm Pulsar Survey, including a discussion of two interesting pulsars.
- **Chapter 6** discusses searching for pulsars using grid computing. A new architecture for scheduling and queueing tasks on an abstract grid system is introduced. Prediction of the run time and scheduling delay of jobs is also discussed.
- **Chapter 7** introduces PulsarHunter, a new software package for the time-domain optimisation of pulsar candidates.

- **Chapter 8** concludes the document and considers how well the project has achieved its aims.

2

New Methods for Analysing Pulsar Candidates

2.1 Background

The final stage of pulsar searching involves manual inspection of the detections from the automated search procedure. Unfortunately, however, there may be more than 200 candidates from each beam of the survey, meaning that for a large survey such as the Parkes Multibeam Pulsar Survey (hereafter PMPS, see section 3.1.1 and Manchester et al., 2001) there are around seven million candidate detections from the whole survey. It is therefore highly impractical to manually look at most of these results, and there must be a scheme for selecting which results to view. An obvious way of selecting the candidates that should be viewed is to threshold the date based on some sensible significance factor, such as the signal-to-noise ratio. This can be improved by designing a set of filters to remove the most common interference signals from the data. Whilst this will select many of the strongest pulsars in the data, many pulsar signals may fall below the threshold or be removed by interference filters.

To attempt to overcome some of the issues of this simple solution, a graphical candidate selection tool, REAPER, was developed (Faulkner et al., 2004), based on earlier work by Edwards et al. (2001) on the Swinburne intermediate-latitude pulsar survey. This allows for all the processing results to be displayed as points on a interactive scatter plot, Typically this will display period on the x-axis and signal-to-noise ratio on the y-axis. The user can select data points from this chart to view the full processing output from that result. This gives a number of benefits over traditional methods of candidate selection, principally in de-selecting against interference. Since interference signals are not sky-localised, they can appear in many beams, usually with similar periods, but differing signal strength. This causes interference signals to form vertical columns on the period/signal-to-noise ratio plot, and so they can be more easily avoided by the user. The graphical display also makes it easy to judge the floor level at which the detections are mainly noise, and so the approximate significance of a detection can be determined 'by eye'.

REAPER also has a number of other features, including automatically matching periods and positions of known pulsars to detections in the processing. It also performs basic bookkeeping, by allowing the user to log results that are classified as candidate pulsars, writing to a text log file and creating a HTML[1] formated version of the log.

2.2 JReaper Overview

Following on the success of REAPER, a new software tool called JReaper (a contraction of Java and Reaper) has been developed as part of this work. Whilst the concept of JReaper is firmly based on REAPER, it has been designed and coded entirely from scratch. This is because the linear design of REAPER did not promote easy adaption and extension. JReaper uses a modular design, with the majority of the code developed using generic interfaces that are entirely survey

[1]HyperText Markup Language - http://w3.org/TR/REC-html40/

and data type independent. This means that by using different modules for interfacing with the data, one can easily use a range of different data types on different surveys without any major changes to the core code base. JReaper includes the advanced functions of REAPER, and also introduces a number of new features, outlined below.

2.2.1 '3-D' Plotting

As well as the standard two axis plotting, JReaper allows for colouring of the data points to produce a '3-D' style plot. This can be used effectively to distinguish points that are close to each other in the standard view. For example, when plotting period and signal-to-noise ratio, showing detected dispersion measure on the colour axis can help pulsars at higher DM values to be distinguished from nearby interference.

2.2.2 Scoring

Whilst the manual aspects of pulsar searching cannot easily be replicated by a machine, it is possible to detect some of the basic indicators used by human observers to distinguish real pulsars from other sources. JReaper features a scoring algorithm that attempts to do this by assigning a numerical score to each source based on the plots that are produced by the search algorithms. The scoring routine uses a number of tests that are then combined with weights to give a final score. The tests currently used are as follows:

Sub-Integration Scoring This looks at how the signal changes in each sub-integration. Typical pulsars have a period and profile that is completely stable over the observation time, whereas most man made signals vary over time. Traditionally this can be seen by looking for straight lines running down a plot of pulse phase vs observation time, known as the sub-integration plot. There are a number of ways to automate this process.

1. 2-Dimensional Kolmogorov-Smirnov Test - The Kolmogorov-Smirnov test provides a statistical way to determine if two data sets differ from one another significantly. This can be applied to the sub-integrations by comparing the observed values with a set generated from noise. Any significant deviation implies that the signal is not due simply to noise. Whilst effective at detecting signals from noise, this test is not effective at differentiating between interference signals and real pulsar signals. It is also ineffective when the pulse is weak enough that there is not a significant signal in each sub-integration. For these reasons the Kolmogorov-Smirnov Test was not used in the final version of JReaper.

2. Hough Transform - Under the Hough transform the input data is mapped in terms of any set of parameterised curves (Hough, 1960; Duda and Hart, 1972). For the purpose of finding straight lines this means showing the data as a plot of how much power lies on each line with given gradient and intercept. This can be used to simplify the problem of finding a straight line down the sub-integrations to that of finding a distinct significant peak in the Hough transformed plane. This can be generalised to any polynomial to fit accelerated sources, however the dimensionality of the Hough plane increases with each extra power. This method is effective at removing signals that are not caused by strongly periodic sources, and is not dependent on the strength of a pulse in each sub-integration. However the Hough transform is computationally expensive and is effectively repeating work that was already done as part of the time domain optimisation. Computing the Hough transform is computationally equivalent to the period adjustment performed by sliding and adding the sub-integrations. Therefore if the time domain optimisation keeps each output profile, the Hough method can be used without separately performing the calculation.

3. Period curve scoring - This method considers the variation of signal-to-noise ratio with trial period as a marker of the periodicity in the sub-integrations. This works using a similar concept to the Hough transform, except the phase information is removed by computing the signal-to-noise ratio of each output profile. This one dimensional output is much simpler to score, as one can model the expected curve for a realistic source by:

$$W_{eff} = \sqrt{W_{int}^2 + (T_{obs} \cdot \Delta Period/Period)^2}, \tag{2.1}$$

$$\frac{Signal}{Noise} = \sqrt{\frac{Period - W_{eff}}{W_{eff}}}. \tag{2.2}$$

   Where $\Delta Period$ is the offset in period, W_{int} is the intrinsic pulse width, assumed to be the pulse width at the optimised period. Note that if W_{eff} is greater than the period, it is assumed that the pulse is completely smeared, and the signal-to-noise ratio is 0.

   By computing the RMS deviation from the model curve, the observed curve can be scored, with low RMS scoring highly. This is the preferred method for scoring the sub-integrations as it a simple calculation and does not rely on any complex 2-dimensional analysis.

DM Curve Scoring This scoring method considers how the detected signal-to-noise ratio varied with trial dispersion measure. Given a period, pulse width and observational frequency and bandwidth it is possible to compute a theoretical variation of signal-to-noise ratio with DM, similar to that described for period offset above:

$$W_{eff} = \sqrt{W_{int}^2 + (k_{DM} \cdot \Delta DM \cdot \Delta\nu/\nu^3)^2}. \tag{2.3}$$

Where ΔDM is the DM offset from the true DM, $\Delta\nu$ is the observation bandwidth, ν is the observation frequency and the constant $k_{DM} = 8.298 \times 10^3\ \mathrm{s}^{-1} \cdot \mathrm{cm}^3 \cdot \mathrm{pc}^{-1}$. W_{eff} can then be fed into equation 2.2 above. This model is then compared to the observed values by taking the RMS deviation from the theoretical values. To prevent bad scores from noise being detected at the same period, values with a signal-to-noise ratio less than 20% of the peak signal-to-noise ratio are ignored. Candidates score highly if their RMS deviation is low.

Zero DM comparison Some un-dispersed signals can be mis-detected at a high DM, caused when a signal with low variation with DM is mixed with a noisy background. To determine these sources, the DM curve scoring is re-performed assuming that the source was at DM of zero. If the scoring is the same or better assuming zero DM then there is a large chance that the DM value has been overestimated. This routine returns a high score if the difference between the zero DM score and the optimised DM score is large.

Profile Scoring Profile scoring detects what fraction of the pulse is above 60% of the peak value. This is designed to select against 'pulses' that have troughs in power rather than peaks. In general the processing algorithms treat this type of pulse the same as its inverse, however no known pulsars behave in this way. This is effective at selecting against interference with pulse widths that are more than 50% of the period, although it may select against some pulsars.

Interference often is detected with a high signal-to-noise ratio, however they typically do not show the basic characteristics of a pulsar signal and therefore score relatively badly compared to a real pulsar signal. The scoring functions are tuned to select signals that have typical pulsar properties, however this means that it is only likely to detect typical pulsars. For example, due to their changing period over time, binary pulsars often score lower in the sub-integration scoring than solitary pulsars, however this effect is limited by taking an average of all the tests.

The result of each score is stored with each candidate, and the final score value that is used for plotting is computed based on user specified weights. These weights allow the user to display the results in multiple different views. For example, weighting the sub integration scoring highly removes a lot of interference, but also affects binary sources, so the user may choose to view the data twice, once with high weighting and once with a low weighting to the sub integration scoring.

The advantage of scoring candidates is that it can be used in conjunction with other candidate selection methods. Generally, the more ways of viewing the data, the less likely a pulsar in the data set will be missed.

2.2.3 Harmonic Detection

In surveys covering very pulsar rich regions of the sky, such as the Parkes Multibeam Pulsar Survey (PMPS) (Manchester et al., 2001), one of the biggest sources of unwanted signals are previously known pulsars. These signals are useful for confirming that the survey is working as expected, however when looking for new pulsars, re-detections of existing pulsars waste much time and resources. Whilst it is simple to remove candidates with period and position of known pulsars, they often have many harmonics, including non-integer fractions of the period. Unfortunately pulsars can easily have fundamental periods that overlap to one of these harmonics, therefore removal of all sources that share these periods is likely to remove potential new pulsars.

To prevent this problem, JReaper has an additional check for harmonics that uses a pulse counting method. This determines the likelihood of a detection being a given harmonic by comparing the expected number of pulses with the number that can be counted in the folded profile of the detection. For example, a signal with one pulse cannot be a 2nd harmonic of a pulsar, similarly a signal with 5/7 of the period of a nearby known pulsar, which shows 5 pulses is likely to indeed be due to that source. For safety, a comment is made on any detection that matches a harmonic period, even if it is not actually marked as a known signal.

2.2.4 Plot Area Selection

Often a user wants to view all candidates from a given region of the JReaper screen. For example, the user has filtered the candidates and wants to select all the points down to a given threshold. JReaper therefore allows the user to select an area of the screen and sequentially display all candidates from that area, excluding any that have been marked as interference, or have already been viewed. This allows users to quickly perform a 'blind' search without repetitive clicking on data points. This can be seen as akin to the thresholding type of search discussed in section 2.1, although this also allows for easy adjustment of the selection criteria. Using this option, the best candidates can be searched in a less subjective manner and so reduce the number of user based selection effects.

2.3 JReaper Design

JReaper is entirely written in the Java programming language.

2.3.1 Data Organisation

In JReaper all 'detections' from the search process are called candidates, and are represented by the `Cand` class. This class defines accessors for all the values JReaper needs to draw the candidate on the main plot. It also has methods to get JReaper specific parameters, such as the candidate 'name' and the class of the candidate. The candidate class is an integer from -2 to 3, where 1 to 3 are classes that represent candidate pulsars of decreasing rank, 0 is a re-detection of a known pulsar, -1 is unclassified and -2 is classified as interference. Users will only be presented with the class 1-3 and textual descriptions of the other classifications. Finally the candidate has a method to get at a class representing the file that stores the details for this candidate. This is represented by an object that implements the `CandidateFile` interface, the actual type of the returned object depends on the underlying data type of the file.

Listing 2.1: `Cand` Accessors

```
// The basic params that can be plotted
double  getAccel();
double  getDM();
double  getJerk();
double  getMJD();
double  getPeriod();
double  getScore();
double  getSNR();
// Some params specified by JReaper
int     getCandClass();
String  getName();
String  getComment()
// Get the CandidateFile for this cand
CandidateFile getPhfile();
```

The `CandidateFile` interface is mainly used to present the full candidate details to the user. The core of JReaper does not have knowledge about the details of all the search results, therefore there is a custom instance of this class for every type of file that contains a candidate. This interface defines methods to create an image representation[2] of the candidate and creation of a Swing JFrame, that displays the candidate details. In both cases exactly what is to be displayed is left to the specific candidate file, as it depends very much on what is available.

Listing 2.2: `CandidateFile` Methods

```
java.io.File        createImage(Cand c, String path);
javax.swing.JFrame  getCandDisplayFrame(Cand c, MainView main);
java.io.File        getFile();
String              getName();
String              getUniqueIdentifier();
```

In JReaper, all candidates belong to a candidate list. The candidate list is a natural grouping of candidates, such as all the candidates from a particular beam.

[2]the actual image made and IO routines are entirely up to the candidate file, this allows for the use of ‘images’ which are not Java Image objects, e.g. postscript files

Candidate lists are stored as gzipped[3] text files that contain a header and the basic details for each candidate, including the specific candidate file Java class for re-instantiation. The file headers use a simple key=value syntax that makes it easy to update if new header elements are introduced.

Listing 2.3: Example candidate list file

```
CANDLIST=PA0011_008D1
FCH1=1374.0
BAND=288.0
TOBS=2100.0
TELESCOPE=PARKES
RA=114.047
DEC=-11.976
BEAMID=PA0011_008D1
##BEGIN_CANDS##
<!-- Here follows a large list of candidates-->
##END_CANDS##
```

Candidate lists are represented in JReaper with the `CandList` object.

Data libraries are the highest level data structure in JReaper. A data library contains a number of candidates lists, generally there is one data library for all candidates from a given survey or processing run. The data library class can be overridden to provide extra functionality, such as offline copies of the classified candidates or survey specific ways to load and sort candidate lists. For example JReaper provides a subclass of `DataLibrary`, `DatabaseDataLibrary` which will synchronise (in both directions) the candidates in JReaper with an external MySQL database. This can then be combined with a PHP [4] interface to create a dynamically updated website that shows the current status of all candidates. On disk, the data library consists of a simple text file that has a key=value format, containing the path to the root of the candidate list directories and the Java class of the data library, as well as any additional parameters specified by any overriding class. The file also contains a cache of the header information from the

[3]gzip is the GNU zip utility, see http://www.gzip.org/ and Deutsch (1996)

[4]PHP HyperText Preprocessor, see http://www.php.net/

candidate lists, to prevent loading the header from each file on JReaper startup. This cache is automatically updated if new candidate lists are found in the search directory.

Listing 2.4: Example data library file

```
CLASS=DatabaseDataLibrary
ROOT=/data/candlists/pmtest/
##BEGIN_CLIST_HEADERS##
PT0275_0851.clist.gz PT0275_0851 191.1 -65.5 1516.5 -288.0 900.0 PARKES
##END_CLIST_HEADERS##
URL=jdbc:mysql://localhost:3306/pmtest
USERNAME=user
PASSWORD=password
WEBPATH=/var/www/htdoc/
WEBURL=http://www.pulsarastronomy.net/pmtest
```

2.3.2 JReaper User Interface Design

The JReaper user interface is developed using Swing, a platform independent GUI toolkit for Java.

There are two main parts to the JReaper user interface, the ‘data load’ window and the ‘main plot’ window. The data load window allows for users to open an existing data library, or create a new one, select the candidate lists to load and import new candidate lists into the data library. The main plot window allows the user to plot and manipulate the loaded candidates in multiple ways.

The first task when using JReaper is to load or create a new data library. This is accomplished with the window shown in Figure 2.1. Once connected to a data library, one can then select the candidate lists to load. By default, all available candidate lists are shown, however they can be filtered by textual filtering on the candidate list name, or filtering by celestial position. This can be achieved through the interface presented in Figure 2.2.

Once the required data sets have been loaded, the user can progress to the JReaper main plot. An annotated example is shown in Figure 2.3. The left

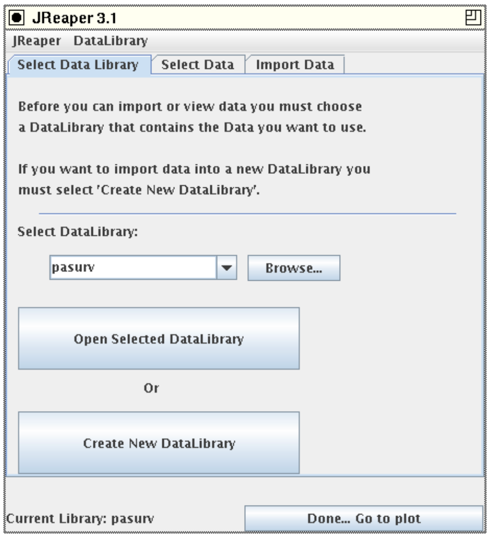

Figure 2.1: This window allows the user to create a new data library or connect to a pre-existing one.

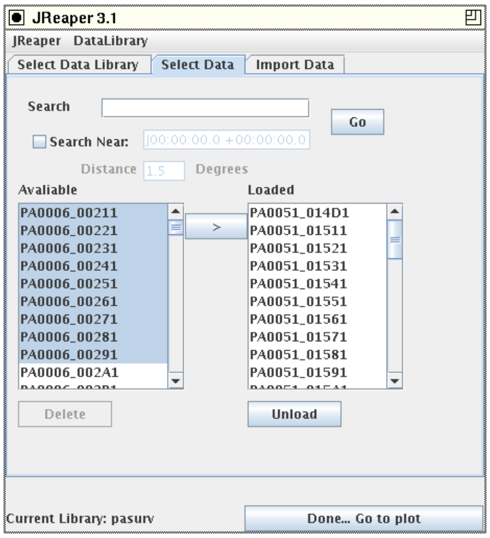

Figure 2.2: The JReaper data load screen. The left list shows the available candidate lists in the current data library, whilst the right list shows the currently loaded candidate lists. The list of available candidate lists can be filtered by text search or by sky position search.

part of the screen is filled with the plot of candidates, shown with one point per candidate, plotted according to the requested parameters. The user can zoom in or out by right clicking and dragging the cursor, and the zoom can be reset to the default value with the 'reset zoom' button. Clicking on any data point causes the plot for that data point to be drawn.

On the right side of the screen are a selection of options to adjust the way the plot is viewed. First is the field that can be plotted on the X, Y and Z (Colour) axis, the available options are:

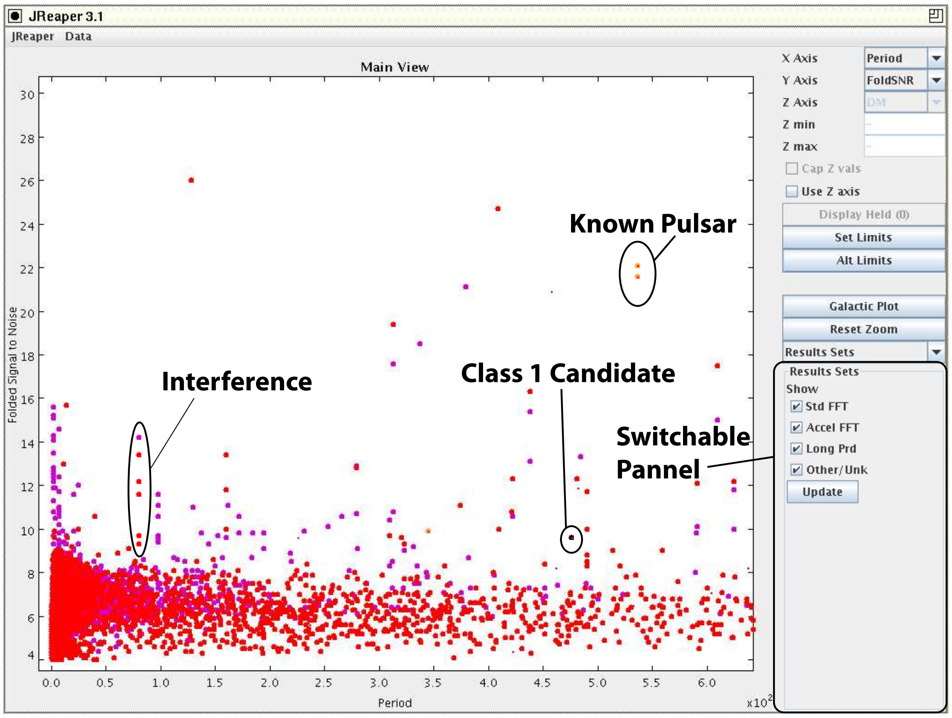

Figure 2.3: An annotated example of the main JReaper plot. The red and purple data points are candidates from two different search methods, the smaller points are those that have already been viewed. The points with a yellow halo are those that have been detected to be due to known pulsars. The point with a black halo is a candidate that has been classified as Class 1 (top likelihood candidate). In this plot, the columns of data points are likely to be due to interference signals as they are detected multiple times in different beams. The 'switchable panel' extends the options and can be switched between a number of different options.

- Spin period
- Dispersion measure
- Signal to noise of folded profile
- Signal to noise of spectrally reconstructed profile.
- Signal to noise in frequency spectrum
- Acceleration
- Jerk
- Observation modified Julian date
- Right Ascension
- Declination
- JReaper score

The next set of options allows for configuration of the Z-axis plotting. When enabled, this colours the data points depending on the chosen parameter value. To reduce the effects of outliers affecting the colour range, a maximum and minimum cap value can be set. By shift-selecting a number of data points, they are added to a queue of plots. Pressing the 'display held' button will cause each plot to be displayed in order. The 'set limits' button allows the user to restrict which data points are displayed by hiding points that lie outside of a set of given parameter ranges. To allow for efficient use of these limits, one can switch to an alternate set of limits by toggling the 'alt limits' button. The 'galactic plot' button switches the view to show galactic longitude against galactic latitude. Each beam that is represented is drawn onto the galactic view, with a nominal radius configured to be equivalent to a given beam size. This plot also shows all nearby known pulsars, helping the user to identify which known pulsars should be visible in the data set.

The select box allows the user to switch the lower right panel between a number of different displays. Currently the available panels are: Select which results set (i.e. search methods) to display, select which beams to display, select which harmonics to mark or hide.

2.4 Effectiveness of JReaper as a search tool

To assess the effectiveness of a tool such as JReaper is difficult as searching through pulsar suspects is still very dependent on the operator.

The most significant, and testable, enhancement provided by JReaper is the pulsar scoring system. A simple test of the score system's effectiveness can be carried out by performing a blind search for known pulsars. This can be implemented by sorting all suspects by their score and calculating how many pulsars are found per suspect searched. This can be compared to the same process, but sorting suspects by the more traditional signal-to-noise ratio. The suspects were separated into three period ranges; less than 20ms, between 20ms and 500ms and greater than 500ms. The results of this analysis are presented in Figure 2.4.

This analysis clearly shows that for all period ranges the pulsar scoring system is effective at selecting pulsars from other suspects, more effectively than by selecting based on signal-to-noise ratio. Unfortunately the relative scarcity of millisecond pulsars means that the statistics for Figure *a* makes the difference difficult to determine, however each pulsar is detected earlier when ordered by score than by signal-to-noise ratio. The most significant effect is for Figure *c*, the period range greater than 500ms. This is most likely to be due to the fact that for longer periods there is an abundance of impulsive interference, which the scoring can easily distinguish from pulsar like signals. In this region, to detect 90% of the detectable known pulsars, one must search ~ 4 times fewer suspects when a)

b)

c)

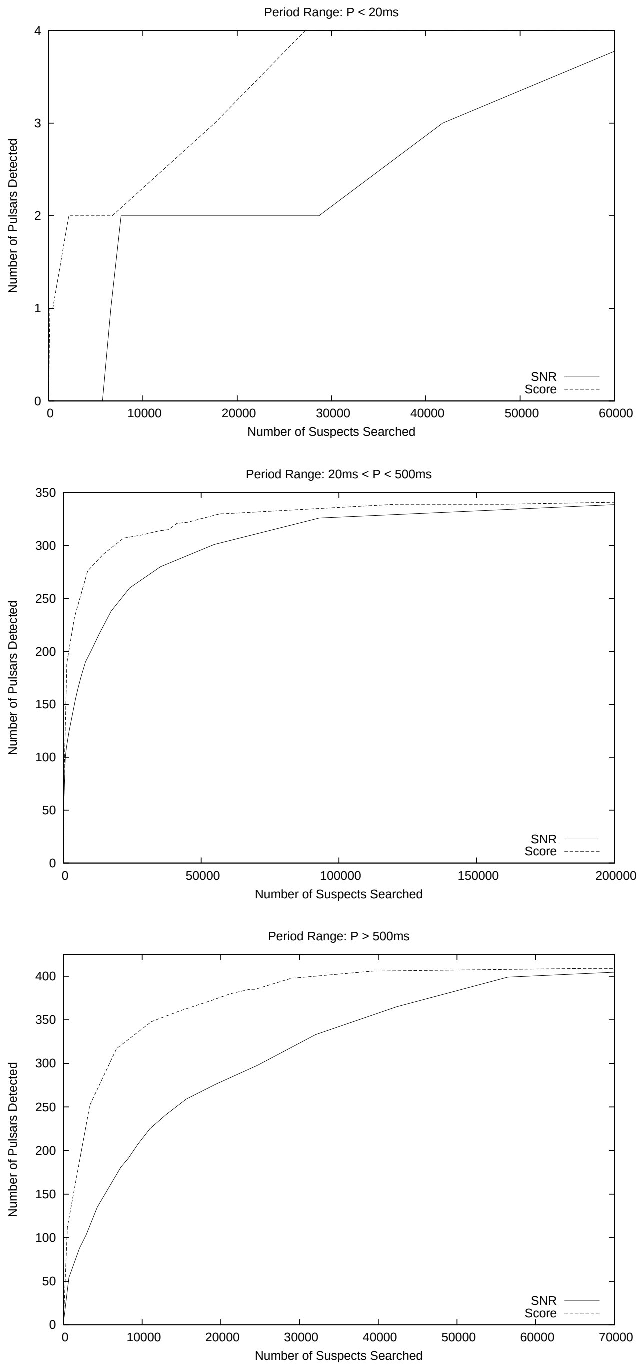

Figure 2.4: These figures show the number of pulsars detected per suspect searched when analysing candidates ordered by signal-to-noise ratio or JReaper score, for suspects in three period range groups.

ordered by score than signal-to-noise ratio. For Figure *b*, the 20-500ms region, this figure drops to ~ 3 times.

Clearly this blind search is not the most efficient means to search for pulsars, one would need to manually view of the order of 10^5 suspects to find 90% of the ~ 700 known pulsars in the data. Whilst this may be possible, it is not a sensible use of resource, especially if one can use JReaper to more intelligently select the candidates to view. However, this does show the relative effectiveness of the scoring system at preferentially selecting pulsars over other signals.

2.5 Uses of JReaper

JReaper was designed to operate using many different data types for different surveys. It is currently in use for the processing of the Methanol Multibeam Survey for Pulsars at Parkes, a survey of the galactic plane at 6.7 GHz using the new 'Methanol Multibeam' receiver[5]. During the initial phases it was using candidates produced by the Sigproc Tune software, however it later moved to use the PulsarHunter software, which produced a different candidate file type. The latest project to use JReaper is the current re-processing of the PMPS being carried out at Jodrell Bank. This processing uses more advanced acceleration searches, and uses JReaper to sort through the many millions of candidates. Again, this processing produces candidates in the PulsarHunter candidate format.

In this work the main use for JReaper has been a re-analysis of the candidates created in the last processing of the PMPS data, discussed in the next chapter. It has also been used in the processing of the Perseus Arm Pulsar Survey, discussed in full in chapter 4. Both of these uses have involved working with the candidates output from the pmsuspectfind software.

[5]So named because it was designed to observe the methanol maser line at 6.7 GHz.

3

Re-Analysis of Candidates from The Parkes Multibeam Pulsar Survey

3.1 Background

3.1.1 Parkes Multibeam Pulsar Survey

The Parkes Multibeam Pulsar Survey (PMPS) is the most successful pulsar survey to date. By utilising a new 13-beam HI receiver (Staveley-Smith et al., 1996), it was possible combine deep pointings with a large survey area. Processing and re-processing of the PMPS data set has lead to the discovery of more than 700 previously unknown pulsars (Manchester et al., 2001; Morris et al., 2002; Kramer et al., 2003; Hobbs et al., 2004; Faulkner et al., 2004; Lorimer et al., 2006a).

The survey targeted a 10 degree wide strip of the galactic plane covering the galactic longitude range $-100°$ to $50°$. This is a very pulsar rich area, prior to the survey commencing, $\sim$ 280 of the $\sim$ 750 known pulsars were in this area, even though it covers less than 4% of the sky. The target range was uniformly sampled, with pointings arranged such that the separation between adjacent beams was close to the half power width of the beam.

Each 13-beam pointing was observed for 2100 seconds, giving just over 2^{23} x 250 μs samples per beam. The dual-polarisation signal from each of the receivers is passed through a 288-MHz filter-bank, splitting the signal into 96 3-MHz wide channels. The 13 x 96 signals were then sampled in total power at 250 μs intervals with a single bit per sample per channel, and the data were then written to magnetic tape for offline processing. These tapes were sent to a number of institutions for initial processing, although the data set is now stored in its entirety at Jodrell Bank.

Initial processing took place at a number of sites, although there was a common processing work-flow for all data analysis. The data were processed using a work-flow similar to that described in section 1.5. The processing used a treeing algorithm for de-dispersion, compensating for the non-linear drift in the pulse arrival time by padding with extra frequency channels. In this original processing there was no acceleration searching, long-period searching, or single-pulse searching.

In 2002, the raw data were collected into a single archive and reprocessed using updated versions of the previous processing algorithms (Faulkner et al., 2004). The processing included additional acceleration searching (see Section 1.5.6), long-period searching (see Section 1.5.4), and single-pulse searching (see Section 1.5.5). The acceleration searching was done by means of splitting the time series into smaller segments and Fourier transforming them. The segments were then recombined, with a series of trial offsets to each segment. This incoherent method is not as sensitive as re-sampling of the time-series, although it requires much less computation and is a great improvement over standard FFT techniques when searching for binaries.

3.2 Application of JReaper to PMPS Candidates

The candidates from the 2002 reprocessing were read into JReaper, creating one candidate list per beam and collecting them together in one group per tape. These were then analysed one tape at a time, using multiple JReaper views including the new scoring system. Good candidates were then classified with a class of 1, 2 or 3, where 1 is the most likely candidates and 3 the least likely. These candidates were automatically added to the processing summary website, which allowed for confirmation observation schedules to be generated. Confirmation observations were scheduled initially to observe the class 1 candidates, then to observe the best from classes 2 and 3.

Using these methods 46 new candidate pulsars were selected, yielding a total of 30 confirmed new pulsars. The confirmation success rate for class 1 was 95%, for class 2 was 64% and for class 3 was 35%. Confirmation of candidates was carried out by performing a 'grid' observation where five short observations are carried out, one on the target position and four offset from the target position by half a beam width in each of the cardinal directions (Morris et al., 2002). The resulting detections, or non-detections and their relative signal strength can provide an improved position for the pulsar. If there are no detections in the griding observations, a longer survey length observation is carried out on the target position. Typically a non-detection in the longer observation implies that the candidate is not a real source, however there may be complications from intermittent pulsars or scintillation, therefore the best candidates will be re-observed multiple times.

3.3 New Discoveries

3.3.1 Basic Parameters

The basic parameters of each of the 30 discovered pulsars are outlined in Table 3.1, and the position of these sources on a P-$\dot{P}$ diagram is shown in Figure 3.1. These sources are all being regularly observed at either Jodrell Bank, or at the Parkes telescope. A full timing solution for each pulsar will be published once sufficient data has been collected.

Table 3.1: Basic parameters for the 30 newly discovered pulsars. Those that are marked with a '*' are discussed in full in section 3.4. Those marked with a '#' do not yet have a fully coherent timing solution. DMs quoted are measured across a 288 MHz band and are typically subject to a $\pm 5\mathrm{cm}^3\mathrm{pc}$ error.

Name	l (deg)	b (deg)	P (s)	DM ($\mathrm{cm}^3\mathrm{pc}$)
J1055-6032 #	-70.84	-0.79	0.099661	633
J1244-6530	-57.77	-2.66	1.546819	386
J1321-5926	-53.22	3.27	1.279058	367
J1432-6040	-44.91	-0.20	1.954430	421
J1518-5416	-36.62	2.67	0.214925	168
J1524-5817	-38.09	-1.20	0.961043	400
J1536-5913	-37.28	-2.73	0.557841	316
J1537-5519	-34.78	0.13	0.395731	611
J1635-4443 #	-20.89	1.80	0.206648	449
J1637-4818	-23.29	-0.83	0.837365	734
J1651-4921	-22.75	-3.11	0.156400	230
J1654-4238	-17.24	0.56	1.101555	962
J1702-4305	-16.60	-0.81	0.215507	537
J1715-3247 #	-6.75	3.23	1.260194	168
J1716-4108	-13.47	-1.80	1.036067	248
J1747-2644	2.05	0.76	0.500254	567
J1753-2243 *	6.30	1.66	0.095138	159
J1755-2027	8.48	2.42	0.322231	365
J1756-2619 #	3.56	-0.90	0.724514	535
J1818-1556	15.09	-0.22	0.952709	237
J1821-1429	16.63	-0.17	1.915131	627
J1830-1312	18.82	-1.49	0.747188	540
J1837-1332	19.35	-2.91	0.178756	155
J1840-0625	25.92	-0.47	1.893353	772
J1842-0738	25.09	-1.68	0.475392	315
J1845-0749	25.37	-2.39	0.861380	187
J1849-0020 *	32.41	0.07	0.166634	958
J1850-0003	32.76	0.10	2.191498	575
J1852-0026	32.54	-0.33	0.518721	512
J1855+0528 *	38.23	1.64	1.393482	353

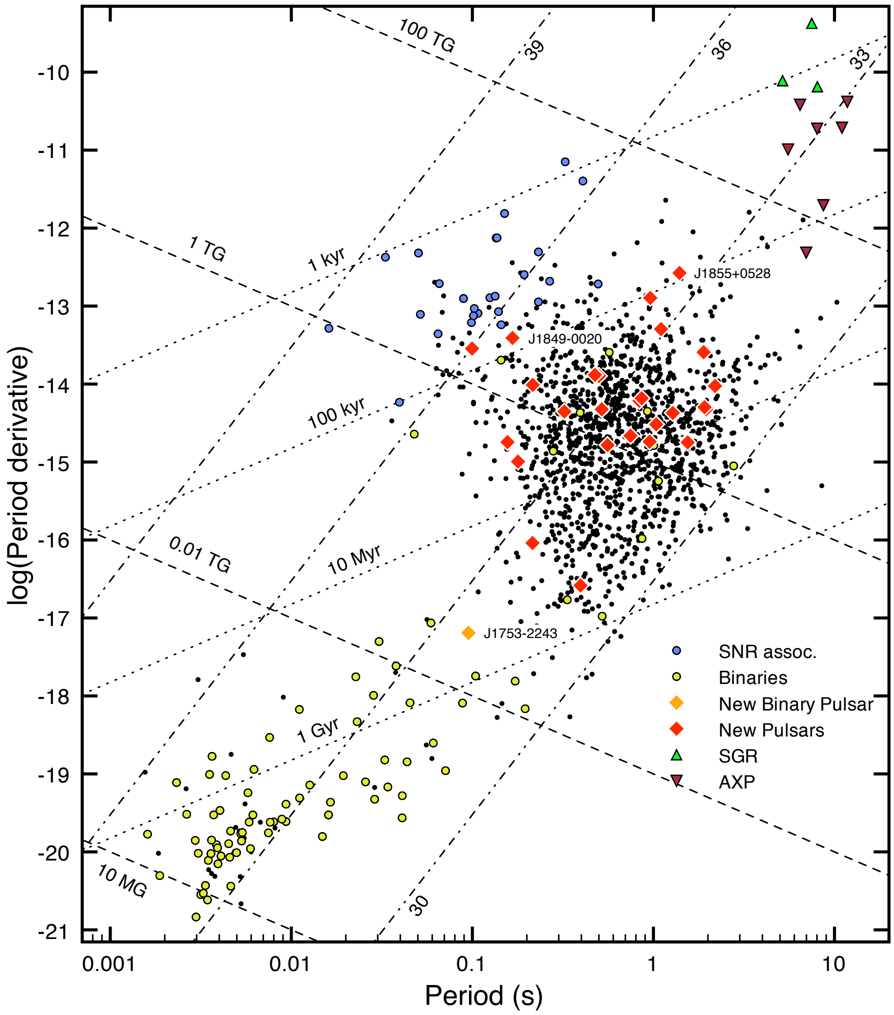

Figure 3.1: P-$\dot{\mathrm{P}}$ Diagram showing the new pulsar discoveries. Note that magnetic field strengths are marked in TG, where 1 TG = 1×10^8 Tesla.

3.3.2 Statistical Analysis

The following is a comparison of the basic parameters of these new pulsars to those discovered by Faulkner et al. (2004).

The distribution of pulsar periods is split into two populations, that of millisecond pulsars and that of the normal pulsars. The set of newly discovered pulsars contains no very short period pulsars, however the Faulkner, et al data set lists 11 pulsars with periods less than 20 ms. By scaling that number to the 30 pulsars in this work, we would expect to detect $\sim$ 2 or 3 new short period pulsars. The small numbers make it difficult to determine if the absence of detections is really significant. For the rest of this analysis, only pulsars with periods greater than 20 ms will be considered.

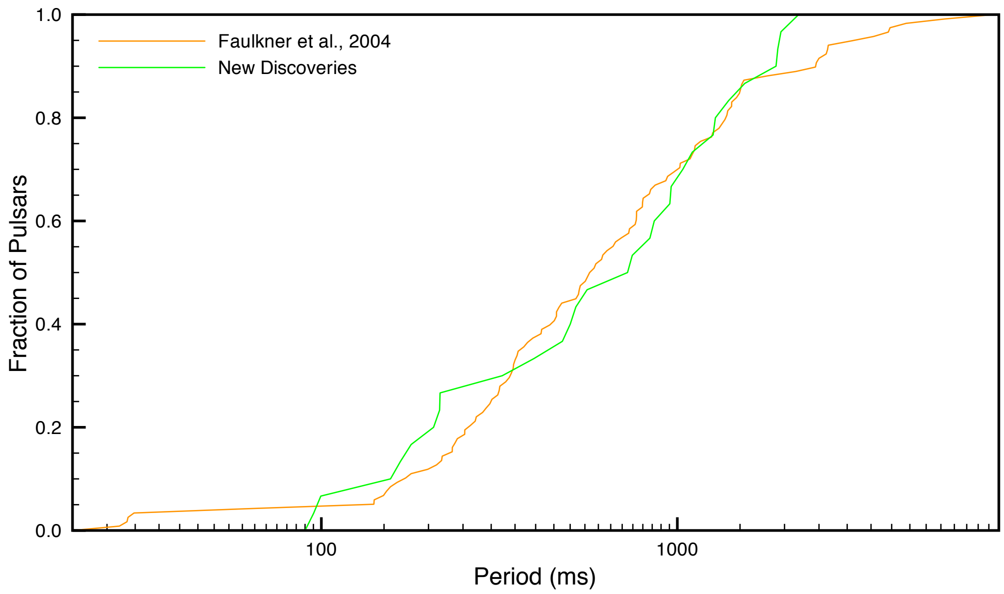

Figure 3.2: Cumulative distribution plot of periods for two samples of pulsars, using a logarithmic scale.

Figure 3.2 shows the cumulative distribution of the rotational periods of two samples. A two-sample Kolmogorov-Smirnov test (KS test) determines that these samples are statistically indistinguishable. This implies that although there may

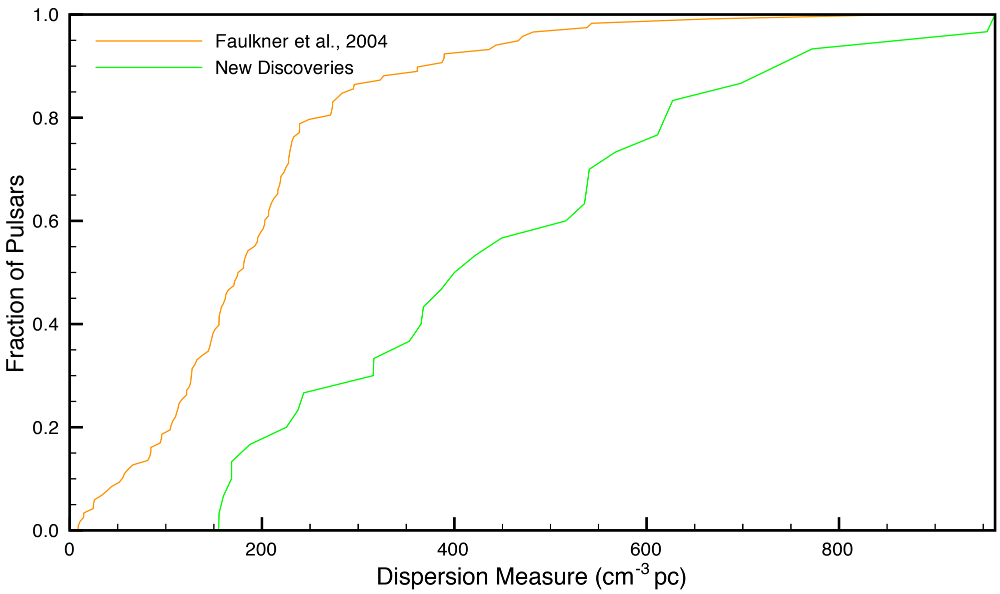

Figure 3.3: Cumulative distribution plot of dispersion measures for two samples of pulsars.

be a lack of short period pulsars, there are no difference in the periods of pulsars found in either search.

Figure 3.3 shows a similar plot, displaying the cumulative distribution of the dispersion measures of two samples. In this case however it is clear that the distributions differ, and the KS test verifies that these distributions differ with a significance level of greater than 99%. The mean dispersion measure for the Faulkner et al. pulsars is 199, considerably lower than the mean value of 453 in the new discoveries data set. This may be due to the fact that larger DM values are likely to produce increased sensitivity of the algorithm which computes scores for the dispersion measure curve.

Figure 3.4 again shows a cumulative distribution plot, this time showing the signal-to-noise ratio for the detections in the two data sets. Again, there is a significant difference, the KS test indicating a difference between the samples at the 90% confidence level. This suggests that the new search methods improve

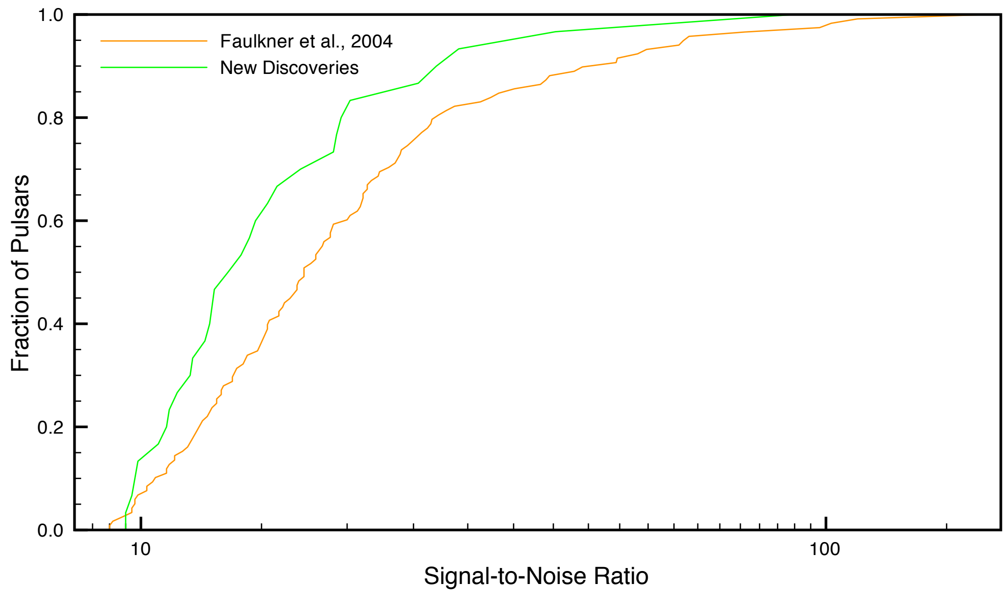

Figure 3.4: Cumulative distribution plot of signal-to-noise ratios for two samples of pulsars, using a logarithmic scale.

sensitivity to low signal-to-noise candidates. In general many of the brighter sources will have been detected in earlier works.

It is also worth noting that the two new detections with the highest signal-to-noise ratio have periodicities of 99.7 ms and 500.3 ms, i.e. very close to the 1/2 and 1/10 harmonics of 1 second, often a strong interference signal. It is probable therefore that these two sources were not detected in the original processing because they were close to interference signals. This suggests that whilst the new techniques are best suited for detecting weaker pulsars in the data, they are also effective at distinguishing stronger pulsars from interference signals.

3.4 Individual Pulsars of Interest

3.4.1 PSR J1753−2243: A Pulsar in an Eccentric Binary System

Initial confirmation observations of J1753−2243 failed to re-detect the pulsar, however it was soon discovered that this failure to re-detect the source was due to a large period change due to orbital motion.

Observations over the next month showed that the period did not drift in a sinusoidal fashion, and that the pulsar must be in an eccentric binary system. The orbit was eventually solved using the FITORBIT software at Jodrell Bank. The variation of the apparent rotation period with time can be seen in Figure 3.5.

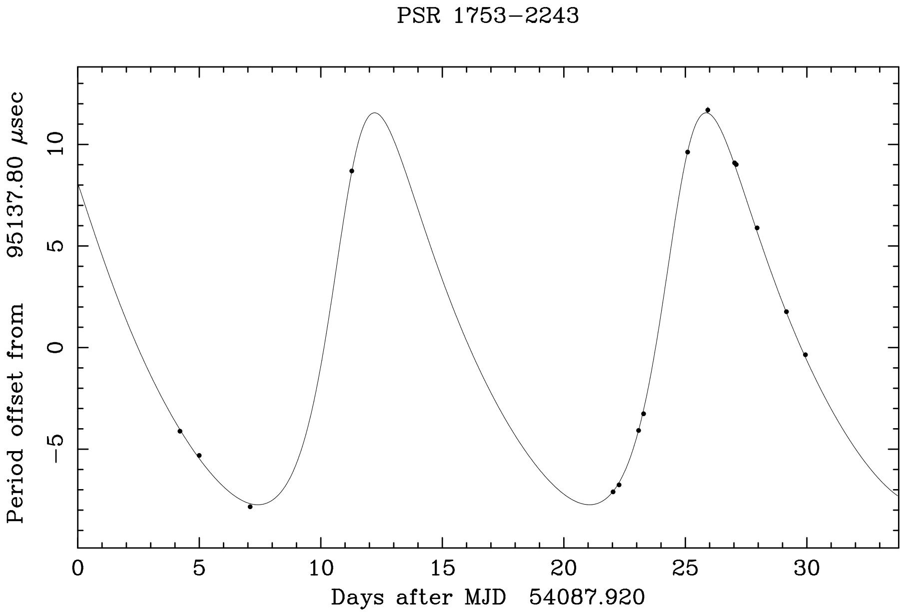

Figure 3.5: Output from FITORBIT, showing the drift of period of PSR J1753−2243 over 30 days.

Continued monitoring and timing of the pulsar gives an accurate model of the orbit of the system, as described in Table 3.2.

The pulse period of 95 ms is typical of that from a young pulsar, however the

Table 3.2: Basic measured and derived parameters of PSR J1753−2243. Computed from 38 TOAs with RMS error of 588 μseconds. Errors in the last digit are shown in parentheses

Parameter	Value	Units
Period	0.09513780859 (7)	Seconds
Period 1st derivative	0.0064 (18)	$\times 10^{-15}$
DM	159.9 (7)	pc/cm^3
RA (J2000)	17:53:40.0 (3)	Seconds
Dec (J2000)	-22:42:42 (249)	Arcseconds
Orbital period	13.63753 (1)	Days
Epoch of periastron passage	06/12/30 03:07:19 (14)	Seconds
Projected semi-major axis	18.1151 (3)	Light-Seconds
Longitude of periastron	-49.322 (5)	Degrees
Eccentricity	0.30357 (3)	
Characteristic Age	234 (66)	$\times 10^6$ Years
Characteristic Magnetic Field	2.58 (72)	$\times 10^6$ Tesla
Mass Function	0.0343155(7)	Solar Masses
Minimum Companion Mass	0.4875	Solar Masses

period derivative of $(0.006 \pm 0.002) \times 10^{-15}$ implies a characteristic age of around 230 Myr. With these rotational parameters the history of this pulsar is not entirely clear. It is possible that this pulsar was born with a very weak magnetic field and therefore evolved slowly, with the period changing little over the millions of years. If the pulsar evolved in this way, it must have had an atypical evolution path, as there are no obvious progenitors for a non-recycled pulsar with these characteristics. However it is also possible that this pulsar evolved as a recycled pulsar, where the binary companion provides a period of accretion induced spin-up. In this case, the accretion must have been interrupted before minimum spin period was reached, otherwise the pulsar would have evolved into the 'millisecond' pulsar regime. Although it seems much more likely that the pulsar is indeed a partially recycled pulsar, both cases will be discussed here.

Measurement of the mass of the companion can give a great deal of insight into the companion type and the history of the system. The measured companion mass depends on the inclination angle, however by assuming a random orientation, the probability of a given inclination angle i goes as $\cos(i)$. Therefore by assuming a nominal pulsar mass of $1.4M_{\odot}$ we can give a 95% confidence that the companion mass is less than $2.5M_{\odot}$ (for $i \approx 18°$), and use this to define a maximal system mass of around $4M_{\odot}$ and a maximal semi-major axis of approximately 36 light seconds.

Unlike superficially similar systems in globular clusters such as J1748−2246J (Ransom et al., 2005), the chance of creating binaries through exchange interactions in the Galactic disk is negligible. Therefore the evolution of the binary should be considered as a closed system that must produce an eccentric system with either an old pulsar with a small magnetic field or a partially recycled pulsar. Another constraint comes from the measurement of semi-major axis, in that the companion radius must be smaller than the closest approach, given by $r_{closest} = a(1 - e)$. For the 1753-2243 system, this value ranges from 5.6 to 18 $R_{\odot}$ for the inclination angles stated above. Neither of these can rule out a main sequence star, however this does exclude a giant star companion.

In the case of the non-recycled pulsar, the companion could be either a main sequence star, a neutron star or a white dwarf. If the companion is a main sequence star, it must be a low mass star with a long evolution path, so that the neutron star forms first. Alternatively if the companion was a very massive star it would evolve faster and form a neutron star before the pulsar is created. In these scenarios, the two stars evolve essentially independently. It is known however that there are eccentric binaries with a pulsar orbiting a white dwarf, such as PSR J1141−6545 (Kaspi et al., 2000). These systems are suggested to evolve by a process by which two stars with similar masses evolve together. The heavier star evolves faster and forms a giant star which collapses to a white dwarf, during this process matter is accreted onto the less massive star, which is then pushed above the threshold for forming a neutron star. The second star then undergoes a supernova explosion, which then causes the eccentricity of the system.

In the case where the pulsar has undergone a period of accretion induced spin-up, the companion must have evolved through a giant phase and therefore must be a degenerate object. Since the pulsar has undergone accretion, the orbit of the system would most likely have circularised. To explain the eccentricity it is therefore necessary that the companion star underwent a supernova explosion, the kick from which was great enough to cause the eccentricity but not disrupt the system. This also explains the fact that the pulsar is only partially recycled as the slower transformation to a white dwarf would almost certainly provide enough accretion to maximally recycle the pulsar. In this case, it is almost certain then that the companion must now be a neutron star. If the system is a double neutron star binary, assuming the masses lie between 1.2 and 1.8 $M_{\odot}$, the inclination angle range is $25^{\circ} \lesssim i \lesssim 29^{\circ}$, which has a nominal probability of just 3% of occurring in a randomly selected system.

The spin parameters strongly suggest that the pulsar underwent a partial recycling, therefore it appears then that this system is likely to be a double neutron star system. This would make PSR J1753−2243 the 10th discovery of a system of this type. Measurement of the companion mass, or a direct detection of the companion would allow for this hypotheses to be more thoroughly tested.

More constraints on the system mass can be computed by the detection of relativistic effects in the pulsar timing residuals. The most prominent measurement is usually the advance of periastron in the system, which, if due to relativistic effects, is dependent on the total system mass to the power of 2/3. By measuring the total system mass, one cannot directly measure the masses of both objects, however by assuming a sensible range of allowed masses for neutron stars, a double neutron star system can be ruled out. Unfortunately measuring masses cannot distinguish between a double neutron star system and a system with a heavy white dwarf.

The measurement error in the advance of periastron can be directly converted into an error in the total system mass measurement. Plotted in Figure 3.6 is the predicted error in the total system mass for different observation timescales. This plot was generated by simulating data using the TEMPO2[1] software package. Given the current residuals, it should be possible to measure the mass of the system to less than 0.1M$_\odot$, within approximately seven years.

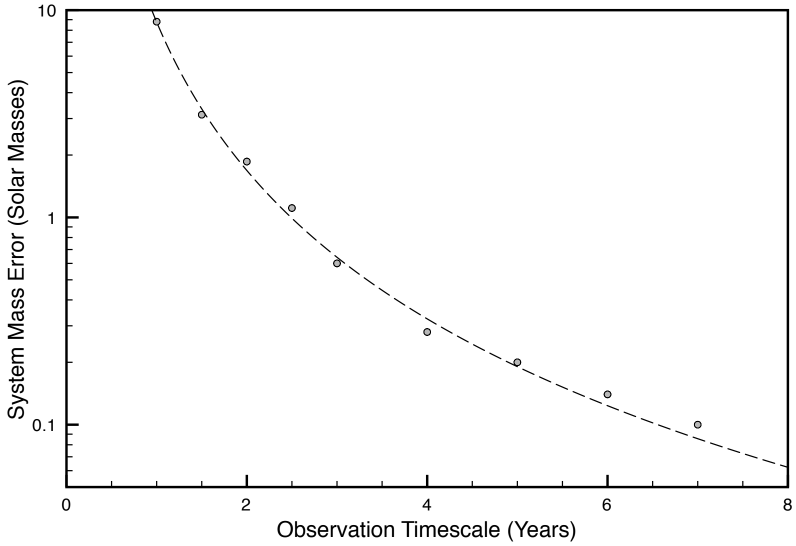

Figure 3.6: Simulated errors in the measurement of the total system mass for J1753−2243, with varying observation timescales. The dashed line is a smoothed curve through the points and drawn as a visual aid. This assumes two observations per month with a typical RMS error of 600 μseconds

Of course, the most constraining way to determine the companion type is a direct detection. If the companion is a main sequence star or a white dwarf this requires observations in the optical band. Detection of the companion star may not be easy however as the pulsar lies close to the galactic centre and therefore

[1]Tempo2: http://www.atnf.csiro.au/research/pulsar/tempo2/

the star field is crowded and the level of extinction is high. The distance to the pulsar can be estimated from the dispersion measure and the Cordes and Lazio (2002) NE2001 electron density model as $D = 3 \pm 0.5$ kpc. At this distance a typical white dwarf has a optical brightness of around 22 to 27 magnitudes. Given the likely extinction and possible blending issues this would be right at the limit of current optical telescope technology.

If the companion is a second neutron star, it is possible that the companion may also be a pulsar. Long observations of PSR J1753$-$2243 have been used to search for a companion, however no pulses have been detected down to a threshold of ~ 0.1 mJy. This lack of detection could be due to geometry (the pulsar beam does not sweep the earth), or that the neutron star does not emit radio pulsations.

If J1753$-$2243 is indeed a partially recycled pulsar in a double neutron star system, we can consider how it relates to other double neutron star systems. Although 9 double neutron star systems are known, two are excluded from this discussion, PSR J1906+0746 as this is a second-born young pulsar (Lorimer et al., 2006b), and B2127$-$11C as this is in a globular cluster and probably formed via an exchange interaction rather than by binary evolution (Prince et al., 1991). A comparison of the parameters for these 7 double neutron star systems, and J1753$-$2243 is outlined in Table 3.3.

Faulkner et al. (2005) suggest that double neutron star systems such as this form in a manner that produces a relationship between spin period and orbital eccentricity. It is suggested that the longer the accretion process, the shorter the spin period and the more time the orbit has to circularise. It has been suggested that the relation would still be detectable even given, or in fact due to, the effects of the second supernova explosion (Dewi et al., 2005). This relationship is plotted graphically in Figure 3.7.

Clearly PSR J1753$-$2243 does not follow the trend set out by Faulkner et al. (2005). Figure 3.8 and 3.9 show how spin period and eccentricity vary with orbital

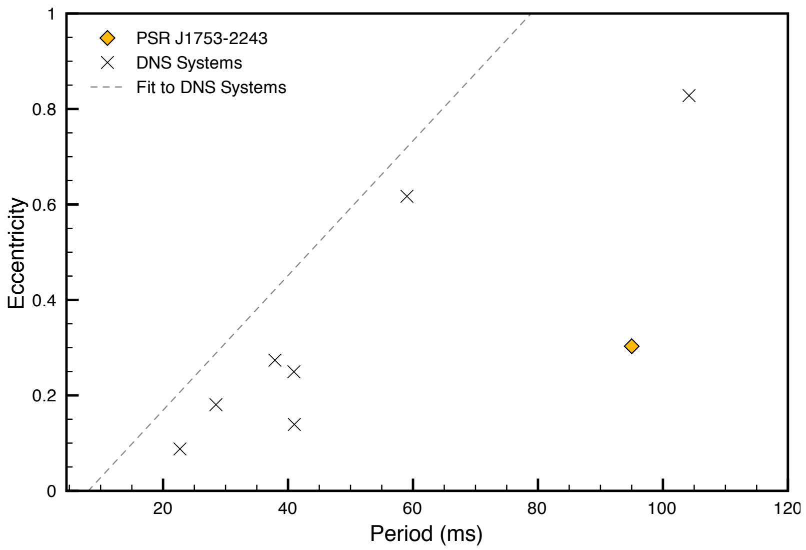

Figure 3.7: The spin period and eccentricity for seven double neutron star systems and PSR J1753−2243. The dashed line shows a linear fit to the seven double neutron star systems.

period, although there is no obvious relationship between these parameters.

In general PSR J1753−2243 is very similar in characteristics to PSR J1811−1736 (Lyne et al., 2000). The comparison between J1753−2243 and J1811−1736 is summarised in Table 3.3. Optical observations for J1811−1736 have not detected any coincident source, however the limiting sensitivity was not great enough to rule out anything other than a giant star. Timing of J1811−1736 by Corongiu et al. (2007) now constrains the total system mass to $2.57 \pm 0.10 M_{\odot}$, which is similar to that of other confirmed double neutron star systems. These authors, under the assumption that the pulsar is indeed mildly recycled, make a similar argument to that presented above for PSR J1753−2243, that PSR J1811−1736 is a double neutron star system.

In general it seems that PSR J1753-2243 shows many of the features of a double neutron star binary, however it also shows several dissimilarities. In the first instance, the mass function of the system is much lower than those of the other

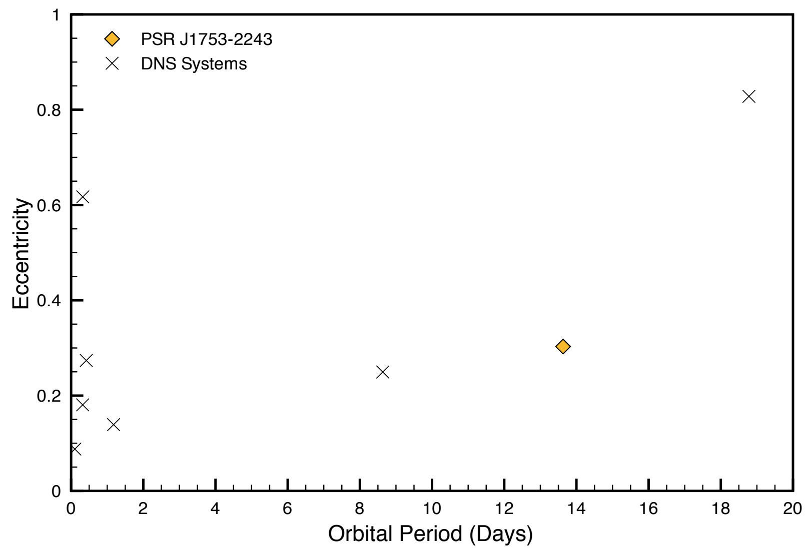

Figure 3.8: The Orbital Period and Eccentricity for the 7 double neutron star systems and PSR J1753−2243.

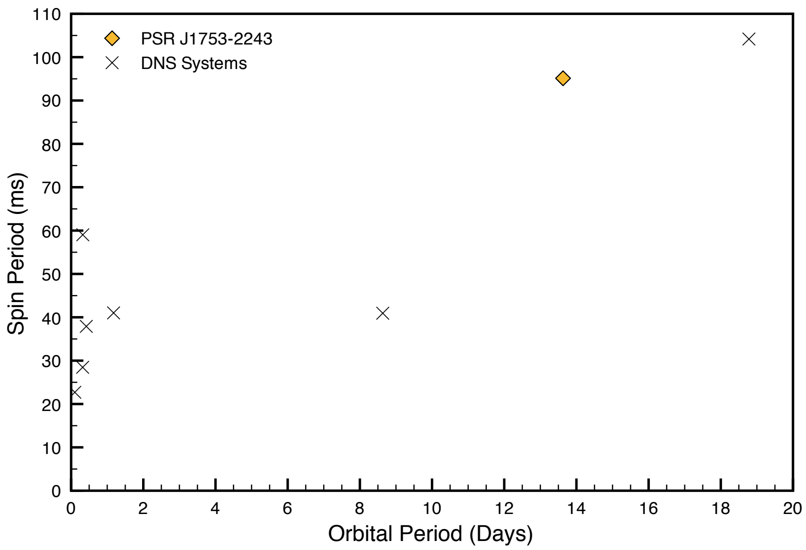

Figure 3.9: The Orbital Period and Spin Period for the 7 double neutron star systems and PSR J1753−2243.

Table 3.3: Outline of the parameters for 7 known DNS systems and J1753−2243. Pulsars are ordered by their spin period. The 7 DNS systems are selected by their probable formation mechanism where the pulsar is the firstborn neutron star. [1]Kramer et al. (2006) [2]Faulkner et al. (2005) [3]Konacki et al. (2003) [4]Hobbs et al. (2004) [5]Champion et al. (2005) [6]Weisberg and Taylor (2005) [7]Corongiu et al. (2007)

	J0737−3039A	J1756−2251	B1534+12	J1518+4904	J1829+2456	B1913+16	J1753−2243	J1811−1736	
P_{spin}	0.02269	0.02846	0.03790	0.04093	0.04101	0.05903	0.09513	0.10418	Seconds
$\dot{P}_{spin}$	0.0002	0.0001	0.002	0.00003	0.00005	0.008	∼0.006	0.0009	$\times 10^{-15}$
P_{binary}	0.10	0.32	0.42	8.63	1.17	0.32	13.63	18.77	Days
$A \sin(i)$	1.41	2.75	3.73	20.04	7.23	2.34	18.11	34.78	Light-Seconds
e	0.08	0.18	0.27	0.24	0.14	0.62	0.30	0.83	
τ_c	204	443	249	23000	12400	108	∼230	1830	$\times 10^6$ Years
B_{surf}	6.4	5.4	9.7	1.0	1.5	2.3	∼2.5	9.8	$\times 10^6$ Tesla
m_c	1.25	1.10	1.30	0.81	1.26	0.87	0.49	0.93	Solar Masses
Ref	[1]	[2]	[3]	[4]	[5]	[6]	-	[7]	

similar double neutrons star systems, and the chances of the required orientation is only 3%. The system also lies far from the relationship between spin period and eccentricity suggested for these types of systems. Regardless of the companion type, it seems that PSR J1753-2243 is an interesting system, lying in a poorly sampled region of the spin period/orbital period/eccentricity phase space. Therefore it will be important to continue observations, to further constrain the system mass and also to monitor the long term evolution of the system.

3.4.2 PSR J1849−0020: A Young Pulsar with a Possible Supernova Remnant Association

If one models pulsar spin-down as magnetic braking, one can define a characteristic spin down age. For the discovered pulsar J1849−0020, this age is of the order of 6.7×10^4 years. This age places the pulsar firmly in the category of young pulsars.

The traditional formation mechanism for pulsars involves supernova explosions, the remnants of which can be detectable for up to 10^5 years (e.g. Shull et al., 1989). Unfortunately however only a tiny fraction of all young pulsars are associated with a supernova remnant. The reasons for this are not entirely clear, however there are two main factors that can influence this. Firstly, pulsars often have a transverse velocity of the order of several hundred kilometres per hour (Hobbs et al., 2005). At this velocity, the pulsar can travel a long distance from the centre of the remnant in the thousands of years since its birth. Although the remnant initially expands faster than that, as the gas interacts with the interstellar medium it quickly slows down. This can mean that a pulsar can be detected well outside of the remnant making association difficult (Gaensler and Johnston, 1995b). The second issue is that many supernova remnants may not be visible, either because the flux density is too small or because the environment is not suited.

Even when a pulsar is detected near to a supernova remnant, determining association is not easy. The probability of a chance alignment goes as the square of the distance from the centre of the remnant and can be large for the crowded central areas of the galaxy (Gaensler and Johnston, 1995a).

For PSR J1849−0020 a spatially coincident source, SNR G32.45+0.1, is listed in the Green supernova catalogue (Green, 2006). This source has been observed as a shell like structure in the X-ray band using the XMM-Newton observatory (Yamaguchi et al., 2004), and is also coincident with a detection in the NVSS L-band VLA survey (Condon et al., 1998). The shell structure is somewhat irregular, however the angular radius is around 2 to 3 arc minutes. The position of the pulsar is well defined by timing measurements, and is 2.5 arc minutes south west of the nominal centre of the remnant. This position is shown in relation to the remnant in Figure 3.10.

By considering a random sample of positions in this area of the sky, one can get an estimate of the probability of a chance alignment. Positions were selected with $-50^\circ < l < 50^\circ$ and $|b| < 1^\circ$. At a separation of 2.5 arc minutes, the probability of a chance alignment is approximately 0.25%.

Hydrogen absorption features suggest that the distance to the remnant is 17 ± 11 kpc (Yamaguchi et al., 2004), which is compatible with the estimated distance to the pulsar of 10.8 ± 5 kpc, calculated using the measured dispersion measure and the NE2001 (Cordes and Lazio, 2002) electron density model. At the time of writing the ATNF Pulsar Catalogue (Manchester et al., 2005) lists 235 known radio pulsars with spin-down ages of less than 5×10^5 years, however only 26 of those are suspected to be associated with a supernova remnant.

Association of pulsars with supernova remnants is difficult, and is it impossible to totally rule out the possibility of a chance alignment. There are however several tests that can rule out an association. The most significant of these is a measurement of the proper motion of the pulsar. It would be expected that the pulsar would originate in the centre of the remnant, therefore the proper motion

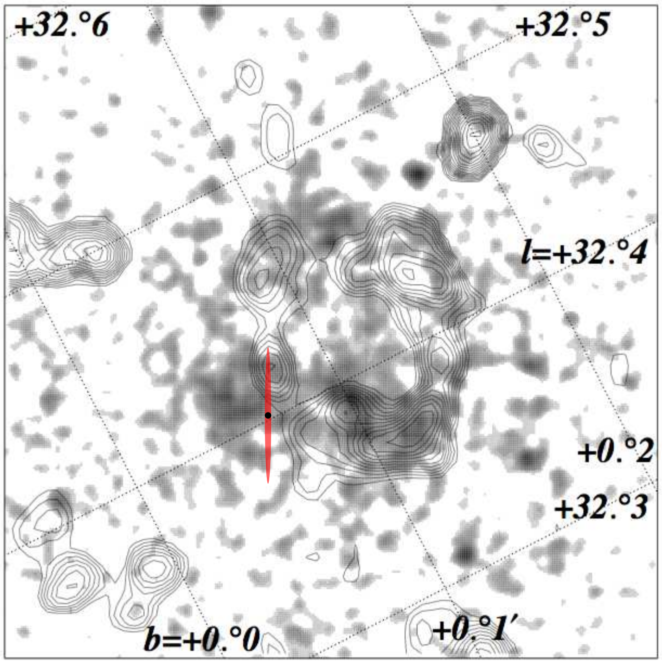

Figure 3.10: Map of the position of PSR J1849−0020 in relation SNR G32.45+0.1. The red ellipse shows the calculated 1-σ error in the position of the pulsar. Underlying map taken from Ueno et al. (2005), which shows XMM-Newton x-ray data in grey scale and a NVSS L-band radio map in contours.

should be directed away from this point. Measurement of proper motion requires several years of observations, however accurate measurements have been made on other pulsars. Given that we have estimates of the time since the supernova occurred (the pulsar characteristic age) and the distance the pulsar is from the centre of the explosion (from the distance estimates and the angular separation), we can estimate the velocity of the pulsar perpendicular to the line of sight. The distribution of pulsar velocities is well known (Hobbs et al., 2005), and therefore if the computed velocity lies outside of the range of measured velocities then association is unlikely. Unfortunately there are large systematic errors in both the distance and the age measurements, therefore the velocity measurement is highly uncertain. For PSR J1849−0020 the transverse velocity is approximately

Table 3.4: Basic measured and derived parameters of PSR J1849−0020. Computed from 47 TOAs with RMS error of 1000 μseconds. Errors in the last digit are shown in parentheses.

Parameter	Value	Units
Period	0.166633375305 (4)	Seconds
Period 1st derivative	39.075 (1)	$\times 10^{-15}$
DM	958.0 (3)	pc/cm^3
RA (J2000)	18:50:14.68 (5)	Seconds
Dec (J2000)	-00:26:04 (2)	Arcseconds
Characteristic Age	67.5	$\times 10^3$ Years
Characteristic Magnetic Field	0.26	$\times 10^9$ Tesla

140 km/s for a nominal distance of 11 kpc, age of 67 kyr and a separation of 3 arc minutes. The velocity scales proportionally to changes in any of the three input parameters, however the computed value is well within the range of measured velocities.

Given the above, it appears that there is a significant probability that PSR J1849−0020 is associated with SNR G32.45+0.1. Measurement of a proper motion will allow for the association to be made with greater certainty. Although it is believed that pulsars are formed by supernova explosions, the fraction of young pulsars associated with supernova remnants is only $\sim$ 1 in 10. The reasons for the lack of associations are not clear, possibly only a small fraction of supernova remnants are ever detectable. Detection of these types of associations give more data for modelling the complex pulsar formation mechanisms.

3.4.3 PSR J1855+0528: A young pulsar with a large magnetic field

The timing of PSR J1855+0528 shows a period of 1.38 s and a large period derivative of 267×10^{-15}. These values imply a characteristic surface magnetic field strength of 2×10^9 Tesla and an age of 83 kyr. Currently there are only 16 other radio pulsars known with magnetic field strengths greater than 2×10^9 Tesla.

Table 3.5: Basic measured and derived parameters of PSR J1855+0528. Computed from 66 TOAs with RMS error of 6200 μseconds. Errors in the last digit are shown in parentheses

Parameter	Value	Units
Period	1.393481704 (2)	Seconds
Period 1st derivative	267.0 (9)	$\times 10^{-15}$
DM	352.0 (5)	pc/cm^3
RA (J2000)	18:55:14.7 (6)	Seconds
Dec (J2000)	-05:22:36 (15)	Arcseconds
Characteristic Age	83.7	$\times 10^3$ Years
Characteristic Magnetic Field	2.1	$\times 10^9$ Tesla

Figure 3.11 shows where PSR J1855+0528 lies on the P-$\dot{\text{P}}$ diagram in relation to other high magnetic field pulsars and magnetars. Lying between the main body of the pulsar population and the magnetar region, these high magnetic field pulsars can provide valuable information about the evolution of neutron stars. Long term timing of PSR J1855+0528 may allow us to determine the direction of movement in the P-$\dot{\text{P}}$ diagram, and hence obtain insight into the evolutionary history and possible future of this source.

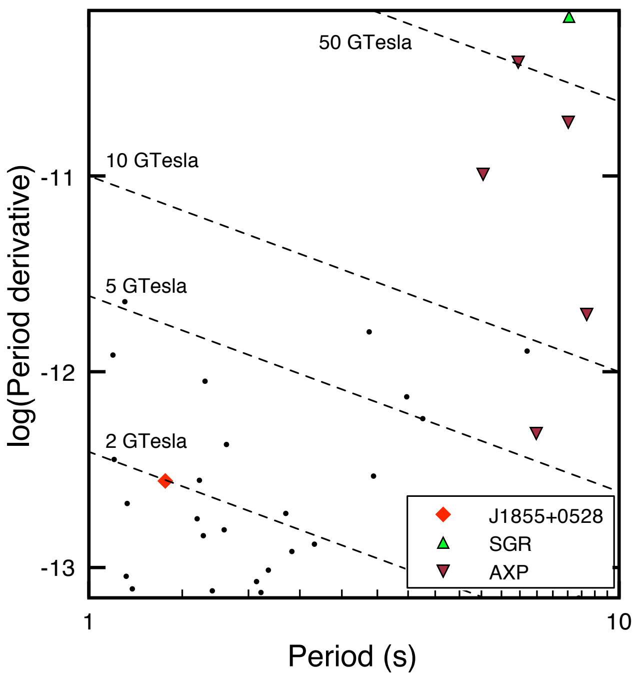

Figure 3.11: A crop of the region of the P-$\dot{\mathrm{P}}$ diagram for pulsars with large magnetic fields.

4

Processing The Perseus Arm Pulsar Survey

4.1 Background

Following on from the success of the Parkes Multibeam Pulsar Survey (PMPS, described in section 3.1.1), several similar surveys were carried out. The Parkes High Latitude Survey (PH Survey) performed a shallow, but wide survey of galactic latitudes $|b| < 60°$. Whilst this was not as deep as the PMPS, it discovered several interesting systems, most notably the Double Pulsar PSR J0737-3039A/B (Burgay et al., 2003). In a similar vein, surveys were carried out by the Swinburne pulsar group (Swinburne Surveys), which also used a smaller dwell time, but covering a more modest range of latitudes $|b| < 15°$, but over the same longitude range as the PMPS. This chapter will discuss the latest of these surveys, the Perseus Arm Survey for Pulsars (PA Survey), which continues the PMPS away from the galactic centre. An outline of these different surveys is given in Table 4.1 and Figure 4.1.

One of the main goals of the PA Survey is to complete the deep low latitude survey of the portion of the galactic plane visible from Parkes, using a uniform

Table 4.1: Basic parameters for four recent pulsar surveys with the HI receiver at the Parkes Radio Telescope. Note that total survey time does not include time allocated to follow up observations. Figure in parentheses shows the additional PMPS discoveries described in Chapter 3.

	PMPS	PH Survey	Swinburne Surveys	PA Survey
Latitude Range	$\|b\| < 5°$	$\|b\| < 60°$	$5° < \|b\| < 15°$	$\|b\| < 5°$
Longitude Range	$-100° < l < 50°$	$-140° < l < -100°$	$-100° < l < 50°$	$-160° < l < -100°$
Pointing Length (Mins)	35	4.4	4.4	35
Sample Time (μs)	250	125	125	125
Number Of Pointings	2670	6456	4702	1068
Total Survey Time (Days)	70	20	14	26
Total Detected Pulsars	1014 (+30)	42	170	27
New Discoveries	733 (+30)	18	69	10+

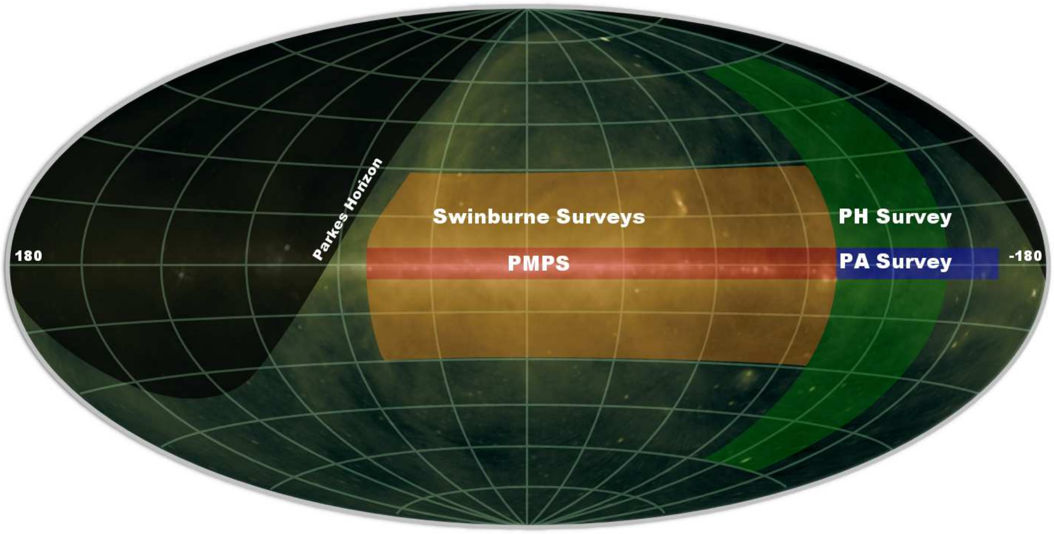

Figure 4.1: The coloured regions represent the sky coverage of recent surveys carried out at the Parkes Radio Telescope using the HI multi-beam receiver. The basic parameters of these surveys are outlined in Table 4.1. The underlying 408 Mhz radio map of the galaxy is used to show the general distribution of galactic matter (Max-Planck-Institut für Radioastronomie).

sampling over 210 degrees of the Galaxy disk. Prior to the survey commencing there were only 18 known radio pulsars in the survey region. Therefore although the numbers of pulsars that are likely to be detected is small, it will allow for statistical models to be extended further from the galactic centre. In particular the new pulsars from the PA Survey will give insight into the radial distribution of pulsars in the outer regions of the galaxy.

4.2 Observations

The PA survey region can be covered by 1068 13-beam pointings. Of these 1068 pointings, 107 of them were observed during the PMPS, which originally covered part of this region. At the time of writing, 811 of the remaining 961 pointings have been observed, leaving 150 pointings to be observed.

The survey began observations in May 2004, and will reach completion by 2008.

4.3 Survey Parameters

The PA survey extends the survey region covered by the PMPS towards the galactic anti-centre. Most parameters have been kept identical to that of the PMPS, so as to provide a consistent data set across the entire galactic plane as seen from Parkes. The only significant change is in decreasing the sampling interval from 250 μs to 125 μs. This change increases the sensitivity to millisecond pulsars, although it also doubles the amount of data storage and processing required. Table 4.2 outlines the parameters for this survey.

Table 4.2: Parameters for the PA survey and for the hardware used.

Survey Parameters		
Sky Coverage	$-160° < l < -100°$	
	$\|b\| < 5°$	
Number of Pointings	1068×13	
Total Survey Time	26	days
Pointing Length	35	minutes
Sampling Interval	125	microseconds
Number of Samples	2^{24}	
Data Size per Pointing	13×192	2^{20} Bytes
Back-end Hardware		
Centre Frequency	1374	MHz
Number of Channels	96	
Channel Bandwidth	3	MHz
Bits/Sample/Channel	1	
Receiver Hardware		
Number of Beams	13	per pointing
System Temperature	21	K
Av. Limiting Sensitivity	0.22	mJy
Beam Width	~ 14	arc minutes

4.4 Processing Methods

The PA survey processing has been split among two sites, at Jodrell Bank Observatory and Osservatorio Astronomico di Cagliari, however since the processing methods differ, it is intended that the data be replicated to both sites and all data be processed with both algorithms for maximum sensitivity and cross checking. The base software for processing the PA survey is mostly identical to that for the 2001 reprocessing of the PMPS, with minor adaptations to account for the change of sample rate and the subsequent doubling of the data length.

At Jodrell Bank, as part of this work, adjustments have been made to the time domain optimisation routines to increase the sensitivity to binary pulsars, the detection of which is one of the main aims of the survey. The basic processing routines allow for only optimisation of period/DM or period/acceleration, but not both. By integrating the CDM software (Faulkner, 2004), which allows for period, acceleration, jerk and DM optimisation, sensitivity to fast binary pulsars is dramatically increased.

Searching this extra parameter space also dramatically increases the processing time, however in general this effect is compensated by a number of factors. Firstly, the basic optimisation routine has been improved and runs faster than the original code. Secondly, the number of candidates has been reduced by combining results from the standard FFT search, the segmented acceleration FFT search and the long period FFA into a single candidate list with no replication of candidates. Finally, the amount of computation power is increased for the PA survey over the PMPS, principally due to less data and faster hardware than in 2001.

The Jodrell Bank processing also uses JReaper to analysis the results, as described in Section 2.2.

4.4.1 Management of Search Processing

To process a survey of this scale, a data and process management system was developed. This kind of management is important to ensure that no data files are missed should technical issues occur and processing fail to complete successfully. When designing Search Manager, particular attention was paid to the importance of recovery of interrupted tasks. This was particularly important in the processing at Jodrell Bank, as the COBRA[1] computer cluster which was used for much of the processing was frequently taken offline for use in real-time processing of astronomical signals.

Search Manager is designed on a centralised control architecture, where work is allocated to generic tasks by a centralised server. The physical design of the computer clusters at Jodrell Bank does not allow for cluster nodes to easily communicate with an external server, however they have access to a shared NFS file system, therefore the client-server communication is powered by a series of shared files. In general this is not a preferred method for interprocess communication, listed below are some of the issues with this method, as well as a description of how they can be minimised.

- Efficiency: processes must poll the file system to receive updates. If the polling interval is high resources are wasted performing many updates of the share, if the polling interval is too short then time is wasted between the message being issued and the message being polled.
  - Efficiency in the process management system is not important, as the fraction of the time spent performing the work for each job is far greater than the process management timescales.
- Reliability: Shares may not be updated immediately and there is no way to lock the file-system when performing updates on existing or new files.

[1] Coherent On-line Baseband Receiver for Astronomy.

– This problem can largely be overcome by ensuring that only the process that creates a file can change the contents of that file.

The advantages of using such a system are the general simplicity. For this method of processing, scripts can update files without executing any other software, and the state of the system is both shared between client and server and is not affected if the server process is prematurely terminated.

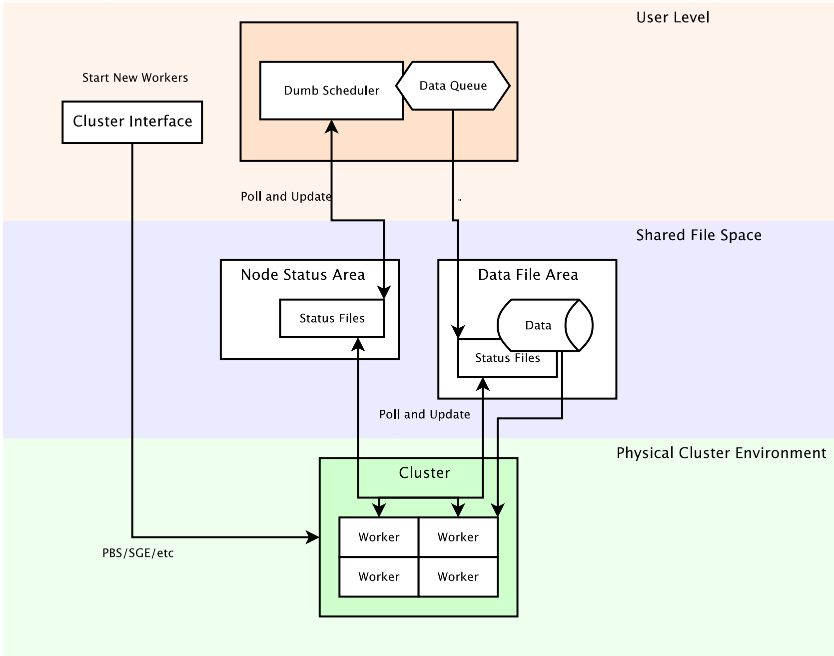

Figure 4.2: Schematic overview of the architecture of the SEARCH MANAGER system.

Figure 4.2 shows an overview of the architectural design of the system. This design forms the basis of the design that is later used in the Pulsar Virtual Observatory project, discussed further in section 6.3. The flow of control is outlined in Figure 4.3, and detailed here. Worker scripts are started using the cluster batch queue system, and when they are executed they create a *nodeid*.ready file in the shared space. The SEARCH MANAGER then creates a reply in the form of a *nodeid*.submit file. This file contains the name of the data file that the task should work on, and if the node was previously interrupted during processing,

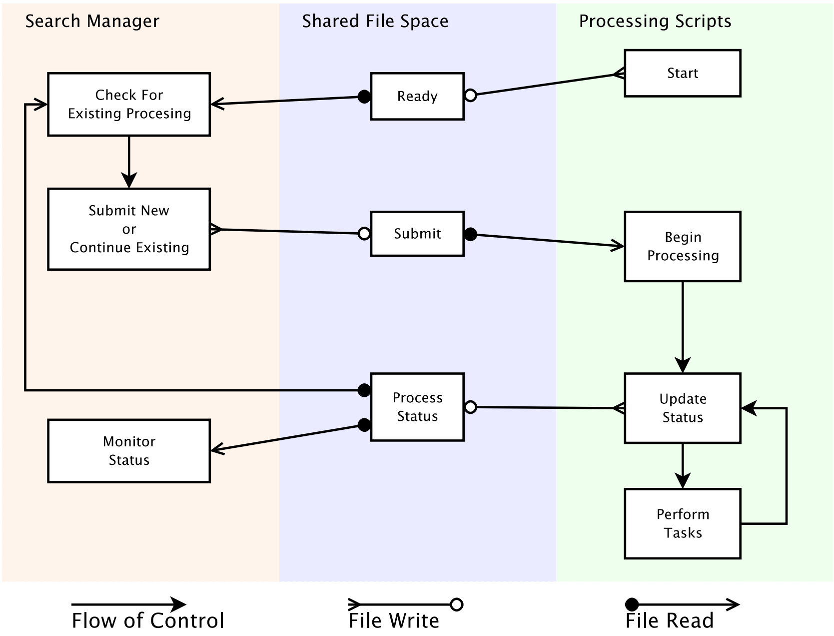

Figure 4.3: Schematic overview of the flow of control in the Search Manager system.

which stage of processing to begin. The *nodeid*.ready file is then removed. The worker script then fetches (if required) the target data file and begins processing, depending on what task was specified by the Search Manager. Whilst processing the worker writes to *nodeid*.process files which inform the Search Manager of the current status of the worker. When a worker has finished processing the data, a *data_file_name*.done file is written to the data location, allowing the Search Manager to be sure that the file was processed successfully.

There are also a number of features implemented to handle error cases. When a worker is running on a node a separate node daemon is launched which updates a node status file every 15 minutes. This allows the Search Manager to measure the CPU/memory load and disk space remaining, however if the file is not updated then it can be assumed that the node has failed silently. The user can then be informed that the node has failed, and that maintenance needs to be carried out. In the case of a software failure that is caught by the worker script, a *nodeid*.ABORT file is created containing the error code for the error that occurred. This status prevents the SEARCH MANAGER from starting new jobs on that node until the user has cleared the abort status.

The user can also request that the task terminates after completing the current process and stop when the processing reaches the next milestone. This allows the user to stop a job, and then re-start it with no waste of processing time. In this instance the SEARCH MANAGER creates a *nodeid*.die file, and the worker will reply with a *nodeid*.dead file when the process has successfully terminated.

5

Results of the Perseus Arm Pulsar Survey

5.1 Re-Detection of Previously Known Pulsars

There are 21 radio pulsars known to lie in the area covered by the PA Survey. The survey has re-detected 17 of these sources. PSR J0625+10 was not re-detected principally because it was below the sensitivity of the survey. PSR J0628+09 is classed as a RRAT, and emits bright single pulses with long nulls in between. As might be expected, it was not detected during the standard searches, however due to issues discussed in section 5.2 it was also not initially detected in the single pulse search. Importantly, the double pulsar PSR J0737−3039A was re-detected in a blind search, however the weaker companion PSR J0737−3039B was not detected. The non-detection of pulsar B is expected as it is weaker than A, and is not visible for much of the orbit. Two other pulsars, PSR J0838−3947 and J0733−2345 lie in areas of the survey area for which observations were taken during the PMPS, and therefore not re-observed.

Table 5.1 lists the 21 known pulsars, noting those that were re-detected and those that were not.

Table 5.1: The 21 known pulsars in the PA survey region. All pulsars for which detection was expected have been re-detected in the survey processing.

Name	l (deg)	b (deg)	Detected	Notes
J0621+1002	200.570	-2.013	Yes	
J0625+10	200.878	-0.956	No	Predicted flux ~ 0.13 mJy
J0628+09	202.203	-0.799	No	RRAT (Detected upon re-inspection)
J0631+1036	201.219	0.450	Yes	
J0725-1635	231.473	-0.330	Yes	
J0726-2612	240.079	-4.644	Yes	
B0727-18	233.760	-0.341	Yes	
J0729-1448	230.392	1.424	Yes	
J0733-2345	238.713	-2.030	-	Covered in PMPS
J0737-2202	237.685	-0.318	Yes	
J0737-3039A	245.236	-4.505	Yes	
J0737-3039B	245.236	-4.505	No	Visibility dependent on orbital phase
B0740-28	243.773	-2.444	Yes	
J0818-3232	251.358	1.875	Yes	
B0818-41	258.749	-2.735	Yes	
J0820-3921	257.261	-1.583	Yes	
J0821-4221	259.825	-3.137	Yes	
B0826-34	253.965	2.561	Yes	
J0835-3707	257.076	1.990	Yes	
J0838-3947	259.606	0.928	-	Covered in PMPS
B0844-35	257.190	4.710	Yes	

5.1.1 Re-detection of PSR J0737-3039A

Since many of the modifications to the search routines are focused on the detection of binary pulsars, the re-detection of PSR J0737-3039A is an important test. The 0737-3039 system is the most extreme binary pulsar system known, with high acceleration or rate of change of acceleration through the 2.3 hour orbit.

Although the pulsar is relatively strong, it is essentially un-detectable in a 35 minute observation without some kind of binary motion compensation. Figure 5.1 (left) shows a PA survey observation containing PSR J0737-3039A folded without any binary motion compensation. Although visual inspection of the sub integrations shows that the pulsar is there, the folded signal-to-noise ratio of just 6.9 is below the typical noise threshold in this type of data.

Figure 5.1 (right) shows the output of the same data when passed through the PA survey processing routines at Jodrell Bank. In this case, the data have been searched in acceleration and acceleration derivative phase space, and the signal-to-noise ratio has increased to 20.5, the strongest signal in the results for this beam.

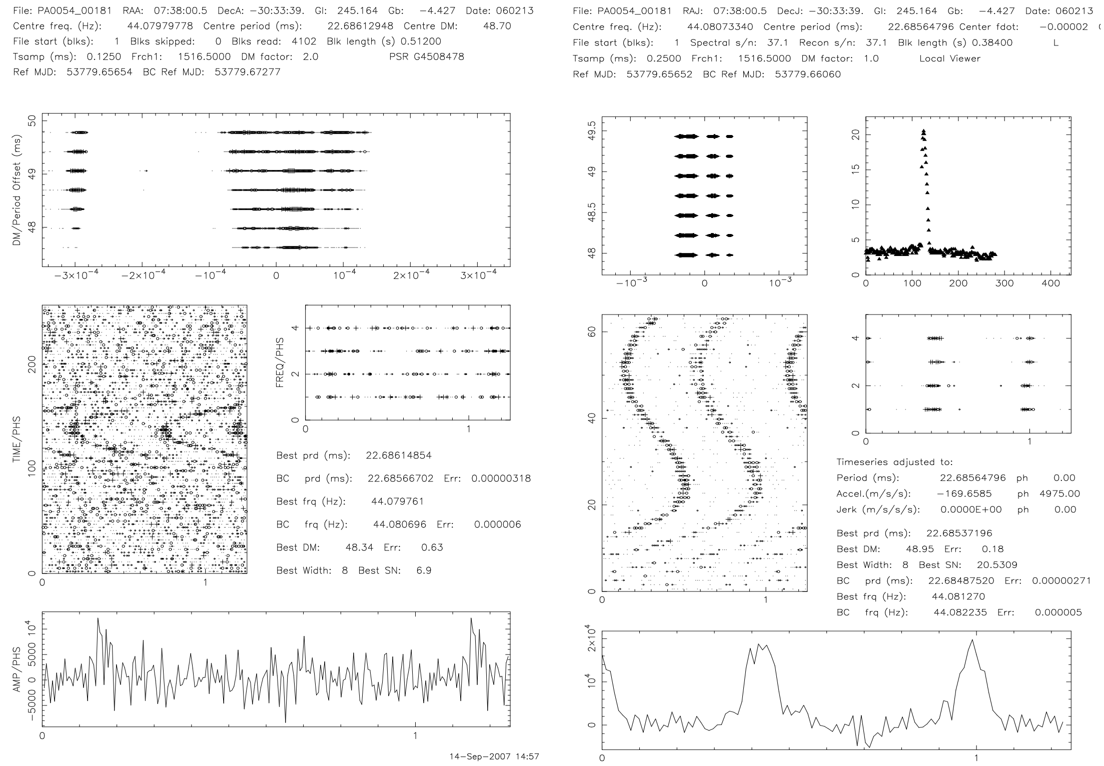

Figure 5.1: The PA survey pointing that covers PSR J0737-3039. The left figure shows the data folded without any binary motion correction, this plot has been generated post-detection. The right figure shows the same data folded with acceleration and acceleration derivative (jerk) adjustment, this plot is from the output from the PA survey processing routines at Jodrell Bank.

From top to bottom, the plots show: Variation of signal-to-noise ratio with trial period and trial DM in the time domain optimisation; For the right plot only, the variation of signal-to-noise ratio with trial DM in the spectral search; Folded intensity against phase, for sub-integrations on the left, and sub-bands on the right; Optimal folded amplitude against phase, the pulse profile.

5.2 Problems with the PA Survey Processing

Although the search algorithms used for the PA Survey processing are essentially identical to those in the PMPS processing, the change in sampling time caused a number of issues. The root cause of the problem is that the software developed for the PMPS processing assumed that there were at most 2^{23} samples in a time series, however in the PA Survey, there were 2^{24} samples. For the most part this was addressed by extending the fixed length arrays in the software to 2^{24} points, long enough for the new data.

Unfortunately this was missed for part of the code that handles the single pulse searching. This caused half of the data to be lost, although only for the low DM trials since the data is re-sampled above a given DM. This made results for the second half of each time-series difficult to interpret because missing low DM data makes distinguishing un-dispersed interference signals from dispersed signals difficult. Therefore it is likely that any bursting sources may be missed in the single pulse search because of this issue. Indeed the known pulsar J0628+09 was initially missed because of this problem, upon re-inspection of these data it is clear that the pulsar is visible.

The PA Survey data also appear to suffer from greater amounts of interference signals than the PMPS data. There are a number of possible causes of this, for example changes in the receiver hardware or back-end system, changes in the Parkes environment, or changes in the processing routines. As with any electronic system it is possible that the quality of the hardware may reduce over time, however this should be considered fairly unlikely as there is a high level of maintenance that is carried out on the hardware systems. The environment at Parkes may be contributing more or different interference signals than in the prior survey. The PA Survey processing uses interference filters that were designed for the initial PMPS survey, however it is possible that the interference environment could have changed since then and so the filters are no longer a good match.

Finally, the additional acceleration searching used in the Jodrell Bank part of the PA Survey could increase the amount of interference that is detected. This can be caused because an interference signal that arbitrarily drifts in period can corrected in period and acceleration several ways that produce similar signal-to-noise ratio measurements. This means that the same signal is detected multiple times with subtly different periods and accelerations, increasing the number of candidates that need to be searched.

There are plans to re-process these data in the future using advanced interference mitigating techniques and an updated single pulse searching routine.

5.3 New Discoveries

At the time of writing, ten confirmed new pulsars have been discovered in the PA survey. Six of the newly discovered pulsars now have coherent timing solutions from data obtained from Jodrell Bank Observatory and the Parkes telescope. The other three sources are currently being timed at Parkes. The basic properties of all sources are outlined in Table 5.2, listed in order of Right Ascension.

Most of the new pulsars show typical pulsar properties, however several of them show interesting phenomena and will be discussed in detail in section 5.4.

Table 5.2: Basic parameters for the 10 pulsars discovered in the PA Survey. Those that are marked with a '*' are discussed in full in section 5.4. Those marked with a '#' do not yet have a fully coherent timing solution.

Name	l (deg)	b (deg)	P (s)	DM (cm^{3}pc)
J0621+0342	206.25	-5.06	0.269954	72
J0623+0341 #*	206.48	-4.46	0.613759	56
J0627+0705 *	203.90	-1.99	0.475876	138
J0628+0653	204.17	-2.08	0.346522	86
J0630-0049 #	211.27	-5.05	0.680573	97
J0645+0854	204.27	3.05	0.903913	142
J0647+0906	204.23	3.28	1.234853	155
J0652-0140 #	214.57	-0.57	0.924959	116
J0658+0027 #	213.35	1.82	0.563292	122
J0821-3824	256.52	-1.03	0.124836	195

5.4 Individual Pulsars of Interest

5.4.1 PSR J0623+0341: A Pulsar Exhibiting Intermediate Length Nulling

The pulsar PSR J0623+0341, discovered in September 2006, emits in short bright bursts of pulses, with longer nulls between the bursts. In consecutive rotations, the pulsar can go from essentially quiescent to an extremely bright pulse and back to quiescence. Typically however the bursts last for 2 to 150 rotations, with a considerable variability in pulse brightness during the burst. Between these bright bursts, no pulsation is detected for around 30 to 5000 rotations. With the pulse period of 613 ms, this equates to bursts of 1-90 seconds followed by nulls of 3 to 50 minutes.

Figure 5.2 shows the pulse power in sequential single pulses for three observations. It can be seen that the pulses move from a mean value of 0, to a mean of around 1, however individual pulses can be very bright.

Figure 5.3 shows a grey-scale plot of data from 250 sequential pulses. Two bursts of emission can be seen, the first starting at around pulse 60 and ending at around pulse 130, and the second starting at around pulse 210 and ending around pulse 225. The main peak of the pulse is aligned to phase 0.0, and a hint of the secondary peak occurs at a phase of around 0.04. It seems that the pulsar can change from a bright state to quiescent state in a single pulse.

This behaviour is somewhat similar to that of the 'intermittent pulsars' such as PSR B1931+24 (Kramer et al., 2006), which shows normal pulsation for a number of days, followed by total nulls of around 30 days in length. This also shares similarities with the Rotating Radio Transients (RRATs) (McLaughlin et al., 2006), which show single bright pulses separated by minutes or hours of total nulling. Figure 5.4 shows these groups of sources plotted by the time-scale (the average time between bursts of pulsation) and the on fraction (the fraction

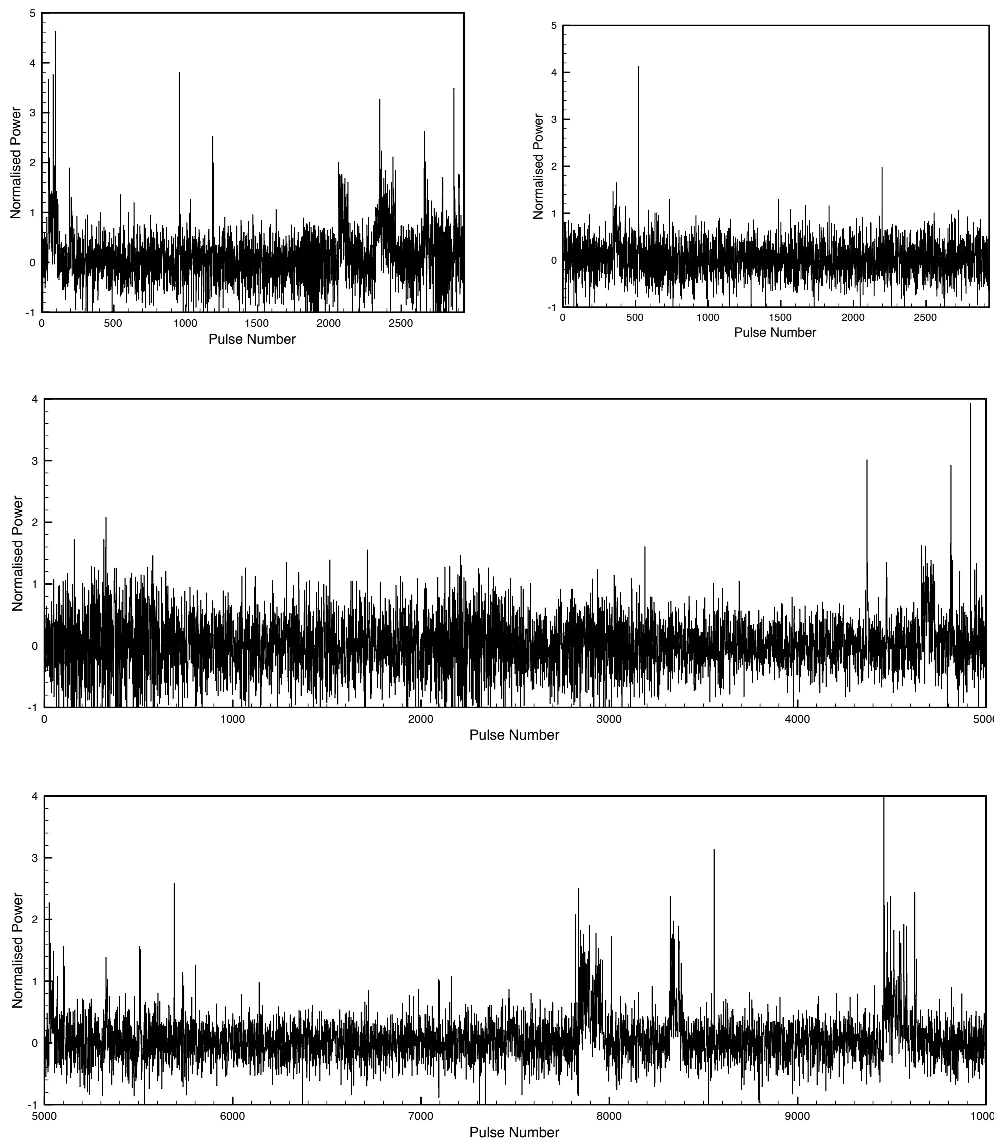

Figure 5.2: On pulse power measured in individual pulses in three time-series, normalised so that the off pulse segment of each pulse has a mean of zero and a variance of one. The first two observations are 35 minutes in length, containing 2934 rotations. The third observation, split over the bottom two panels is 1.7 hours, 10000 rotations in length. Data were collected with the Parkes telescope.

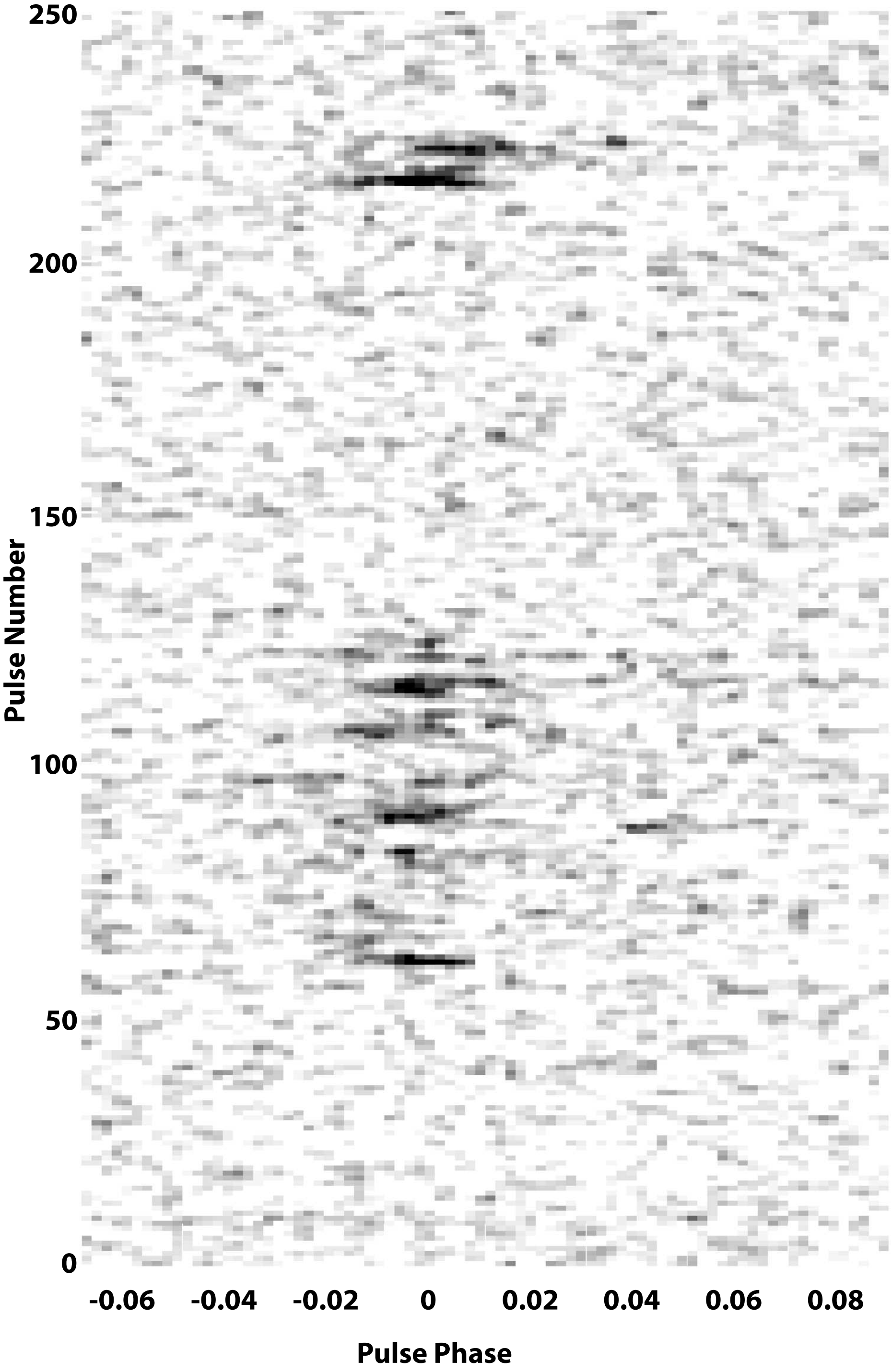

Figure 5.3: Grey-scale plot of single pulses from PSR J0623+0341.

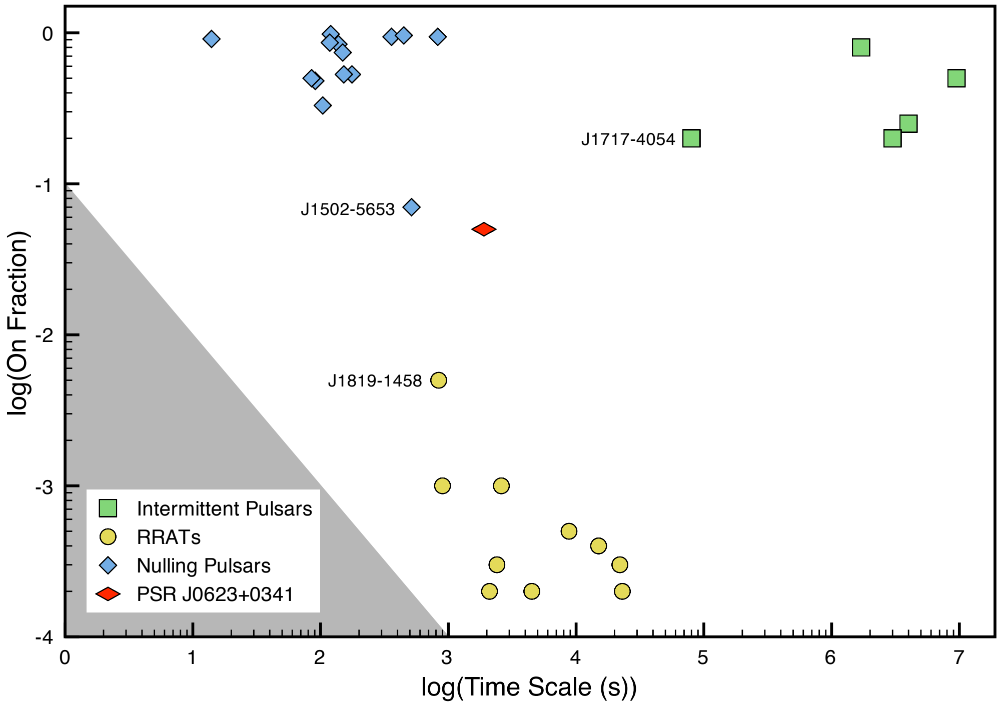

Figure 5.4: Plot showing the nulling timescale verses pulse fraction for different classes of nulling phenomenon. The grey region to the left is forbidden for pulsars with periods greater than 100ms, as the nulling would occur on timescales less than the period. Data for the nulling pulsars from Wang et al. (2007). Data for the intermittent pulsars are unpublished and are provided by M. Kramer. Data for the RRATS is from McLaughlin et al. (2006).

of the time spent in a pulsing state). Clearly, PSR J0623+0341 shares the space between the RRATs, the intermittent pulsars and nulling pulsars. This makes the discovery of sources like PSR J0623+0341 potentially valuable as they may provide a link between the many different forms of pulsar nulling phenomenon. The most similar previously known pulsar is PSR J1502−5653, which is considered to be a nulling pulsar, however this pulsar is not typical of nulling pulsars, and it is known to have the highest nulling fraction of all known pulsars.

The gaps in this phase space may be partly due to the difficulty in detecting pulsars of this type. With small pulse fractions, increasing time scale implies a reduction of the chance of pulses occurring during a survey observation, and also reduces the chance that confirmation observations will detect any pulses. For small pulse fractions, the greatest chance of detection is when the pulse timescale is of the same order as, or less than the survey observation time.

Detection of pulsars like PSR J0623+0341 using traditional Fourier methods is difficult due to both the reduced flux and the difficulty of distinguishing this type of emission from RFI signals. J0623+0341 was likely only found as the pulses are very bright and it was visible for about 20% of the discovery observation. For a similar pulsar with a lower brightness, or a smaller pulsing fraction, regular Fourier methods would be insufficient.

However, since the pulsar emits bright single pulses, pulsars such as this should be easily detected in the single pulse search which has previously been so successful for detecting RRATs.

5.4.2 PSR J0627+0705: A Pulsar with a Polarised Inter-Pulse

From the discovery of PSR J0627+0705 it was apparent that the pulsar showed an inter-pulse at a phase of approximately 180° from that of the main pulse. This strongly suggests a geometry where the inclination angle between the rotation axis and the magnetic axis (α) is approximately 90°, and that each pulse is caused by a separate magnetic pole. The alternative explanation that this is two edges of the same emission cone is possible, however the closeness to 180° and the overall narrowness of the pulses suggest that this is not the case.

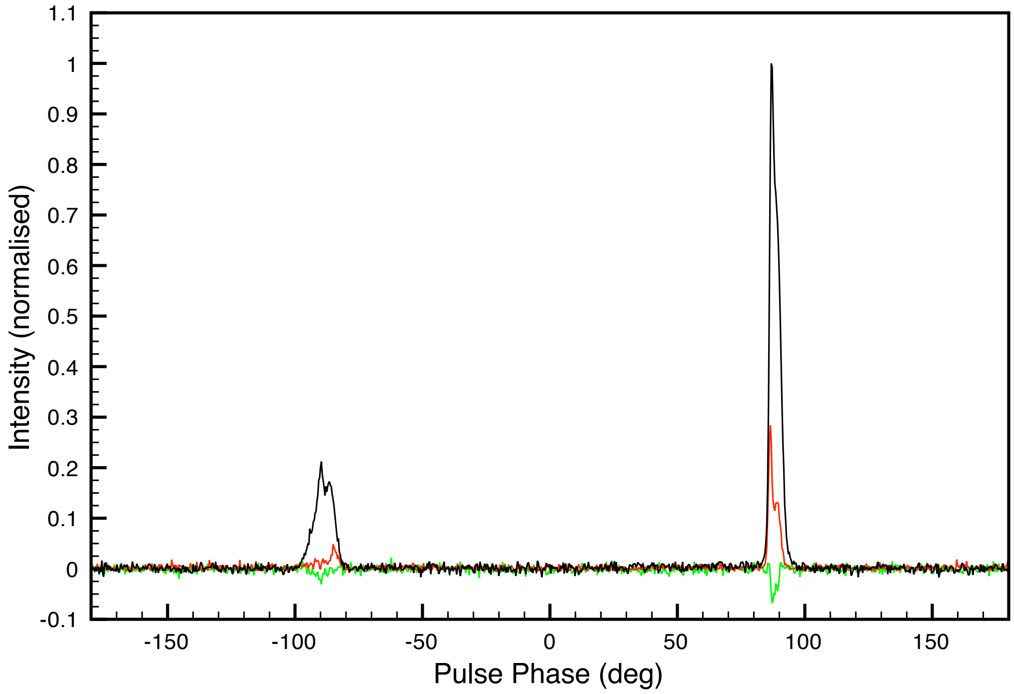

Figure 5.5: Folded profile of PSR J0627+0705 at 1400 MHz. Total intensity is shown in black, linear polarisation in red and circular polarisation in green. Values are normalised so that the maximal intensity is unity. Data provided by S. Johnston.

Constraints can be placed on the geometry of a pulsar if one assumes that the polarised emission from the pulsar is due to the so-called 'rotating vector model'

(RVM). Radhakrishnan and Cooke (1969) suggested that the characteristic S-shaped swing of polarisation position angle can be simply explained if the linear polarisation of the emission is defined by the direction of the magnetic field at the point of emission. In this model, as the cone of emission passes over the observer, the polarisation position angle changes as different parts of the magnetic field come into view. The line-of-sight direction can be visualised as an arc cut across the magnetic pole, and the measured swing in polarisation position angle dependent on the angle of the field lines that are crossed by the cut.

To determine the geometry of a pulsar we define two parameters, α, which is the inclination between the spin axis and magnetic axis, and β, the angle between the magnetic axis and the line of sight to the pulsar. The variation of polarisation position angle, Ψ, with pulse phase, ϕ, can be computed from α, β, Ψ_0 (a offset in PPA) and ϕ_0 (an offset in phase) as shown below:

$$\tan(\Psi + \Psi_0) = \frac{\sin\alpha\ \sin(\phi - \phi_0)}{\sin(\alpha+\beta)\ \cos\alpha - \cos(\alpha+\beta)\sin\alpha\cos(\phi-\phi_0)}. \tag{5.1}$$

In this geometry, the line-of-sight direction draws a circle of size $\alpha + \beta$ around the rotational pole of the pulsar. If the line-of-sight trajectory is equatorward of the magnetic pole, it is considered an 'outer line of sight', otherwise, when the line-of-sight trajectory is between the spin axis and the magnetic pole it is considered to be an 'inner line of sight' (Everett and Weisberg, 2001). Over a small range of pulse phase, the swing in polarisation position angle for the two different geometries is difficult to distinguish as α and β are degenerate in this range.

Therefore to truly get a measure of the geometry one must sample pulse phases far from $\phi = 0$. This can be achieved using pulsars with a large pulse fraction, however measurement of a polarised inter-pulse is much more constraining as it can more easily distinguish between inner and outer lines of sight as an outer line of sight will reach a stationary point and turn over, whereas an inner line of sight will not. Even though the inter-pulse may come from a different magnetic pole, the mathematics remains the same.

Because of this degeneracy there are very few pulsars for which the RVM fit can be well constrained. However for PSR J0627+0705, the inter-pulse is sufficiently polarised to be used to constrain the geometry. Figure 5.5 shows a folded profile of PSR J0627+0705 measured in intensity, linear polarisation and circular polarisation. Both the main pulse and the inter-pulse are up to 25% linearly polarised depending on phase. Variation of the polarisation position angle with pulse phase is shown in Figure 5.6, along with a fit to the data. The data were fitted using α, β and offsets in both phase (ϕ_0) and position angle (ψ_0).

The variation of χ^2 with trial α and β values is shown in Figures 5.7 and 5.8, with the latter being an enlargement of the region surrounding the best fit. As expected, the solution is bi-modal, with each solution representing the geometry for one of the magnetic poles. The solutions are $\alpha = 96.5°$, $\beta = -1.33°$ and $\alpha = 83.5°$, $\beta = 11.7°$. This complies with the expectation that $\alpha \sim 90°$, so that both magnetic poles could be visible. The rotating vector model only considers the dipole as a whole, rather than individual magnetic poles, however we can associate each solution to one part of the pulse. The main pulse is associated with the $\beta = 11.7°$ solution, whilst the inter-pulse is associated with the $\beta = -1.33°$ solution. Figure 5.9 shows this arrangement schematically.

Given the difficulty in fitting the RVM swing for most pulsars due to co-variance of α and β, the strong inter-pulse in PSR J0627+0705 provides an excellent opportunity to constrain the geometry of the pulsar. Although the fact that the pulsar has a clear inter-pulse is highly indicative of an orthogonal rotator, the fact that the RVM fit agrees well with this gives credence to the determined parameters.

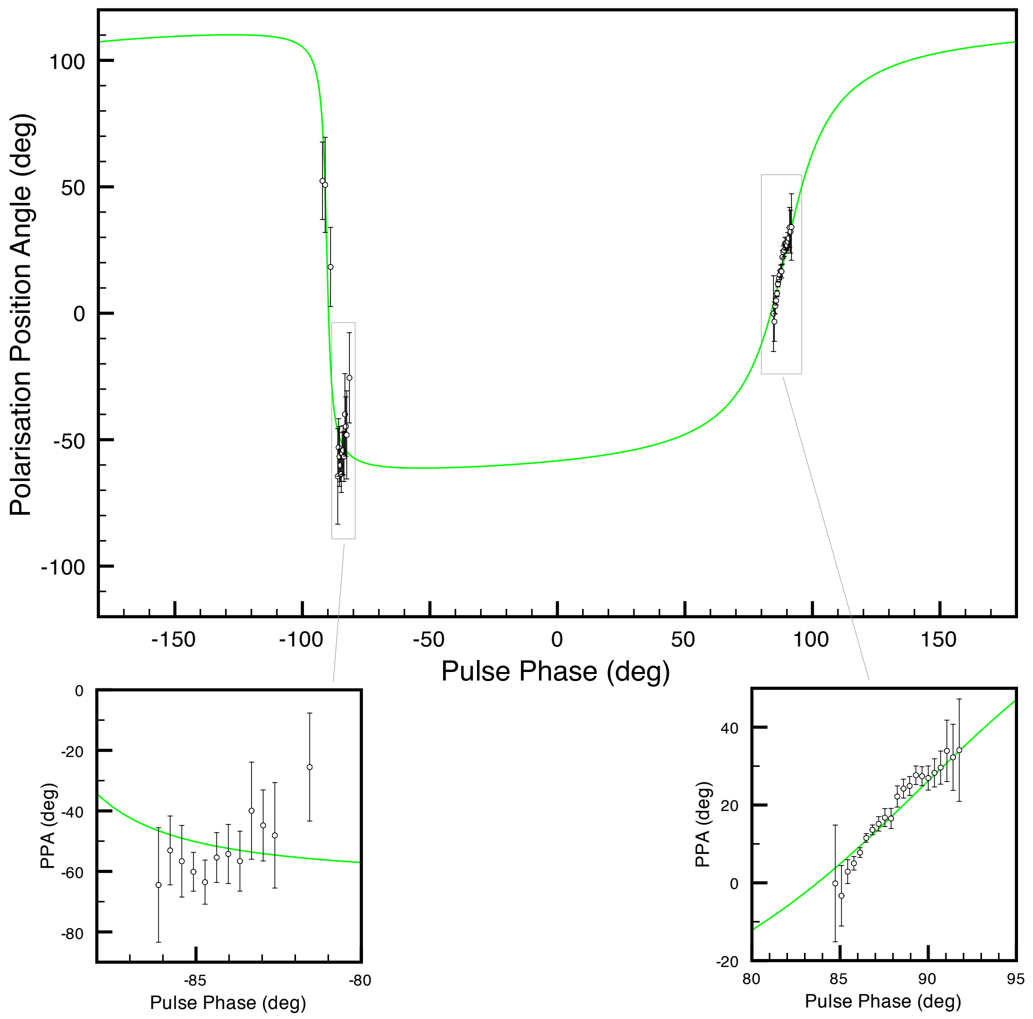

Figure 5.6: The best fit to the polarisation position angle swing PSR J0627+0705. Inserts show detail from two regions of the pulse where emission occurs. This fit shows $\alpha = 96.5°$ and $\beta = -1.33°$, with reduced χ^2 of 1.2.

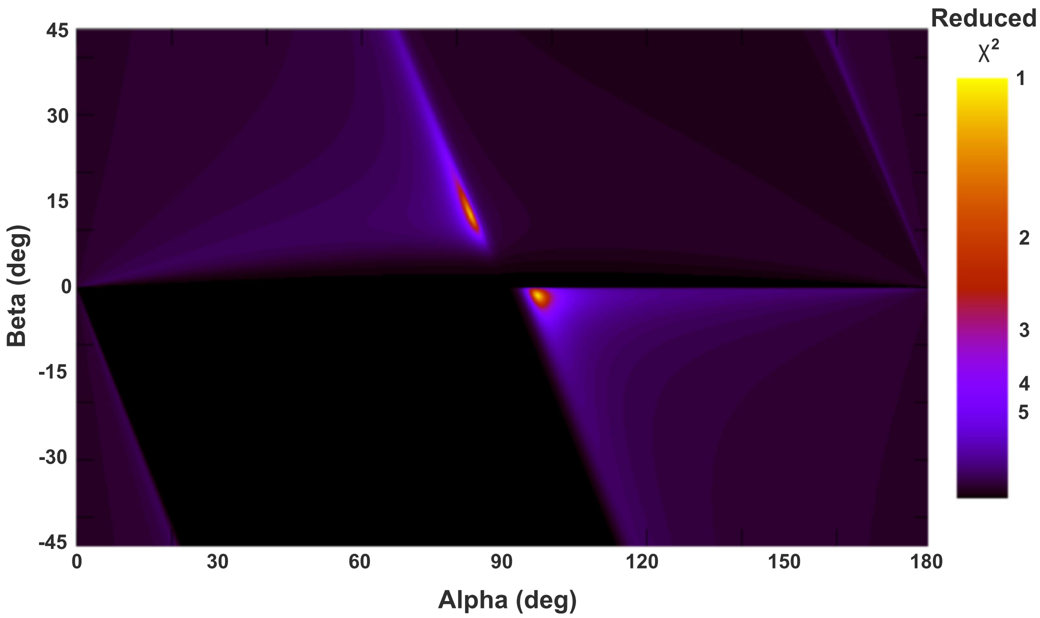

Figure 5.7: The reduced χ^2 surface for fits over a large range of α and β with the phase and position angle offsets fixed at the optimal value. Note that a pulse phase offset of 180° has been introduced for all fits with $\beta < 0$, so that both solutions are shown.

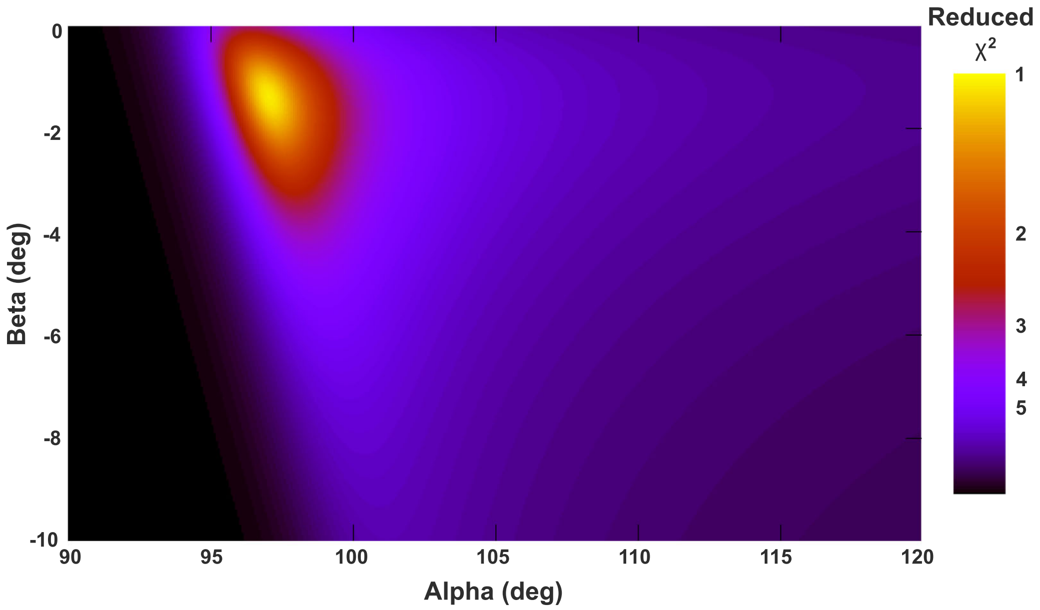

Figure 5.8: The reduced χ^2 surface for fits over a small range of α and β with the phase and position angle offsets fixed at the optimal value.

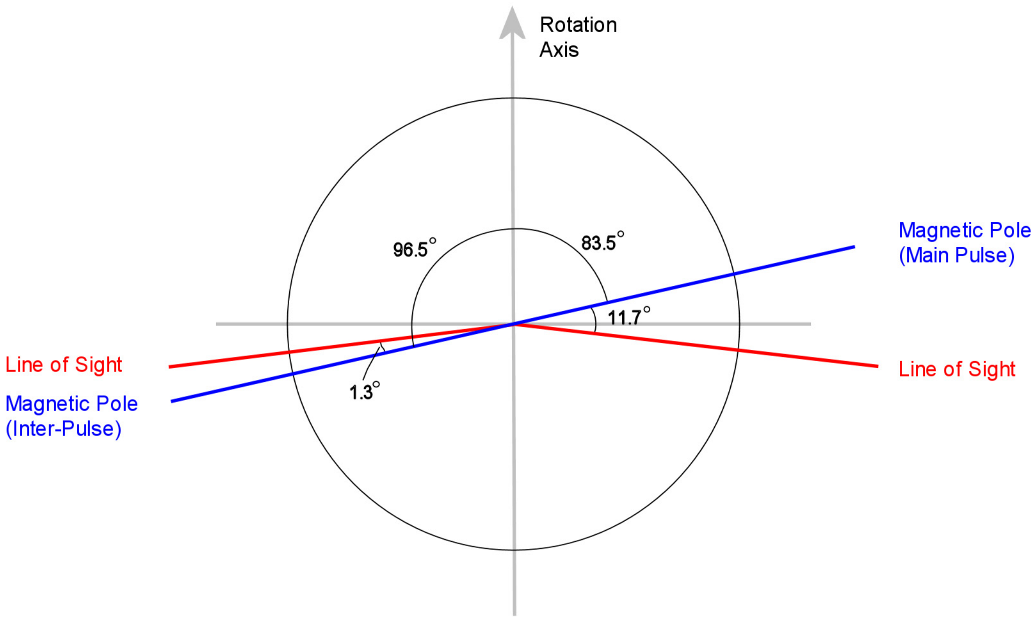

Figure 5.9: Schematic diagram showing the arrangement of the magnetic poles and the line of sight with respect to the rotation axis of the pulsar.

6

Searching for Pulsars using Grid Computing

6.1 Introduction to Grid Computing

Grid computing is an emerging field and there is no one accepted definition of a Grid. For the purposes of this work, a Grid shall be defined as a heterogeneous set of resources that can be accessed by a homogeneous interface. This means that the grid may consist of different hardware, software, or be run by different facilities, but a user of the grid as a whole can do work with the grid, without knowing which specific machine will be running the job. Some examples:

- Computer clusters in different institutions, administered locally, could be considered a grid if there is a standard user interface that allows users to run jobs on any node in the grid, without having to customise their job for a specific machine (e.g. recompile the code, or get a user account on the particular machine)

- A set of assorted computers spread throughout an institution. Each machine differs in hardware and software, however if the job scheduling software ensures that jobs that require specific hardware or software are run on appropriate machines, this can be considered a grid.

In this way, many small computers can be combined to behave as one larger computer. Grid computing does not inherently add more computing power, but it allows for it to be used more efficiently.

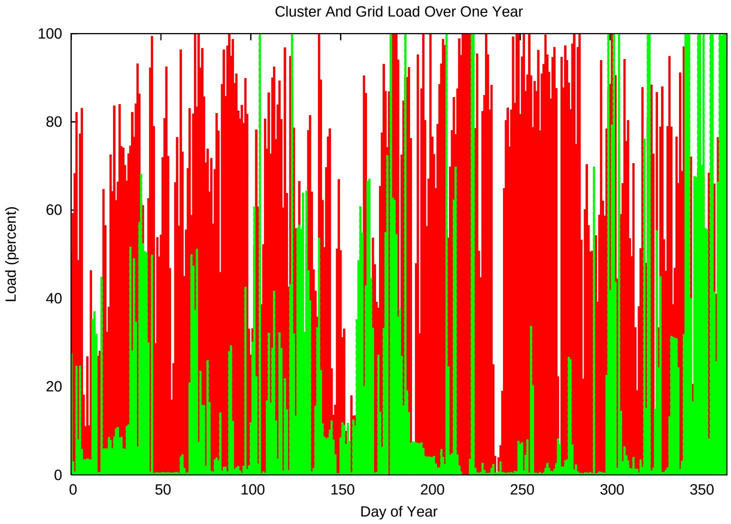

Figure 6.1: This figure shows a typical cluster load pattern for one year. The red boxes show the load for a cluster in a grid, whilst the green boxes show a cluster which is not connected to a grid.

Computer clusters are often used in bursts of processing, for example when new data becomes available. During idle periods between these bursts the processing power of the computer is wasted. By using the computer as part of a grid system, this idle time can be used by other users in exchange for the use of other machines. An example of this effect is shown in Figure 6.1. The isolated cluster, the green boxes, shows considerable variation in usage, with several quiet periods between bursts of activity. The grid cluster, the red boxes, shows much less variation in the load, and therefore the usage levels.

Unfortunately for grid systems, whilst the grid computer behaves as a single computer cluster the latency between nodes in the cluster can be increased dramatically. This means that grid computing is best suited to problems that can be run as separate processes that do not require large amounts of interprocess communications.

6.2 Pulsar searching with Grids

Pulsar searches can be considered to be highly parallelisable. That is, each task in the search can be run in parallel, without any cross communication between tasks. This makes the pulsar search problem ideal for running in a grid environment.

The pulsar search process, as described in Section 1.5, consists of a number of steps. At each step, the number of parallel jobs increases. So, for each file that is to be processed, the data must be de-dispersed at a set of trial DM values. Each de-dispersion process can be performed in parallel, and the results for each step determined separately. If acceleration searching is used, then for each DM value, a number of trial acceleration values must be used. If the data is to be de-accelerated, then each de-acceleration process can also be performed in parallel.

One cannot indefinitely parallelise the problem however, since each process requires input data. The more parallel processes are run, the more times the data must be replicated across the grid. The goal is to achieve the minimum total run time, which can only be achieved by optimising the trade-off between the time spent processing and the time spent transferring data. There is also no benefit in parallelising the processing so that there are more jobs to run than the number of nodes available. Therefore the splitting of the processing depends on the amount of data to be processed, the number of nodes available, the network speed between nodes and the processing throughput of each node.

The problem is further complicated when the connection speed between nodes is non-uniform. For example, there may be a cluster located at the institution where the data is hosted, but the majority of the processing power may be elsewhere. So, whilst the naive solution may be to send one file to each node in the grid, the optimal solution may split the work unevenly between the nodes.

The problem of arranging this processing is non-trivial, however a great reduction in total processing time may be possible with some kind of intelligent scheduling.

6.2.1 Suitable Grid Environments

Grid computing is a fairly general term, therefore there is a great variety of different types of system that can be classified as a grid. These range from very distributed systems with individual computers linked via the internet, workstation clusters distributed around an institution or dedicated computing clusters linked via the internet. The first of these is typified by the BOINC system (Anderson, 2004), the system behind the popular SETI@home experiment (Anderson et al., 2002). This type could provide a great deal of power to the pulsar search problem, however pulsar data is large and it is not feasible to split it into smaller segments. Therefore whilst this type of system has a great deal of future potential, it is only now becoming viable as the internet bandwidth available to home users is increasing.

Therefore this work has concentrated on the development of a system for processing pulsar data on a large scale grid, with each grid node being a computer cluster, although no constraints are placed on the constituents of the cluster.

6.2.2 The Pulsar Virtual Observatory

The Pulsar Virtual Observatory (PVO) project aims to develop a system that allows for researchers in all fields to access and analyse the large data sets stored in pulsar surveys, without specific knowledge about the data or the processing mechanisms. This is not the first pulsar project to be developed for grid resources, Pickles et al. (2001) and Brooke et al. (2007) describe experiments performing coherent de-dispersion in a grid environment. These experiments were limited to small scale trials with close supervision of the processing. The PVO project aims to perform autonomous processing on complete pulsar search work-flow.

This can be achieved by moving the data and processing tools to a grid resource, where the details of the processing are seen by the user as abstracted tasks. By abstracting the tasks in this way, the user only has to understand what results are required, not the details of the processing code that is used to get there. The Abstract Queueing Architecture (AQuA) system, which has been developed in parallel with the PVO core system, allows for scheduling of these abstract tasks.

This opens up large sets of radio time-series data to a wider audience, enabling greater cross field astronomy, in line with the global virtual observatory project.

6.3 AQuA - An Abstract Queueing Architecture

The AQuA system is a queueing system independent of the grid system used, or the application being run. This means that the application, for example the Pulsar Virtual Observatory, can execute in many different grid, and non-grid, environments. For example, the application may have resources available on a number of grid systems and/or some local compute resources. The initial design for AQuA was based on the core ideas behind Search Manager system described in Section 4.4.1. Adaption for use in a distributed grid system has evolved the design to a much more sophisticated level, however the core ideas of simplicity and abstraction from the application processes remains.

In the AQuA system, the jobs submitted to the underlying grid architecture are generic worker jobs that can run a set of *Task*s. For each physical ‘node’ in the underlying system, there is a *LocalManager* instance in the AQuA server. This ‘node’ may be one machine, or one cluster, or possibly a whole grid of networked machines. Once the workers start, they request a task to perform from the AQuA

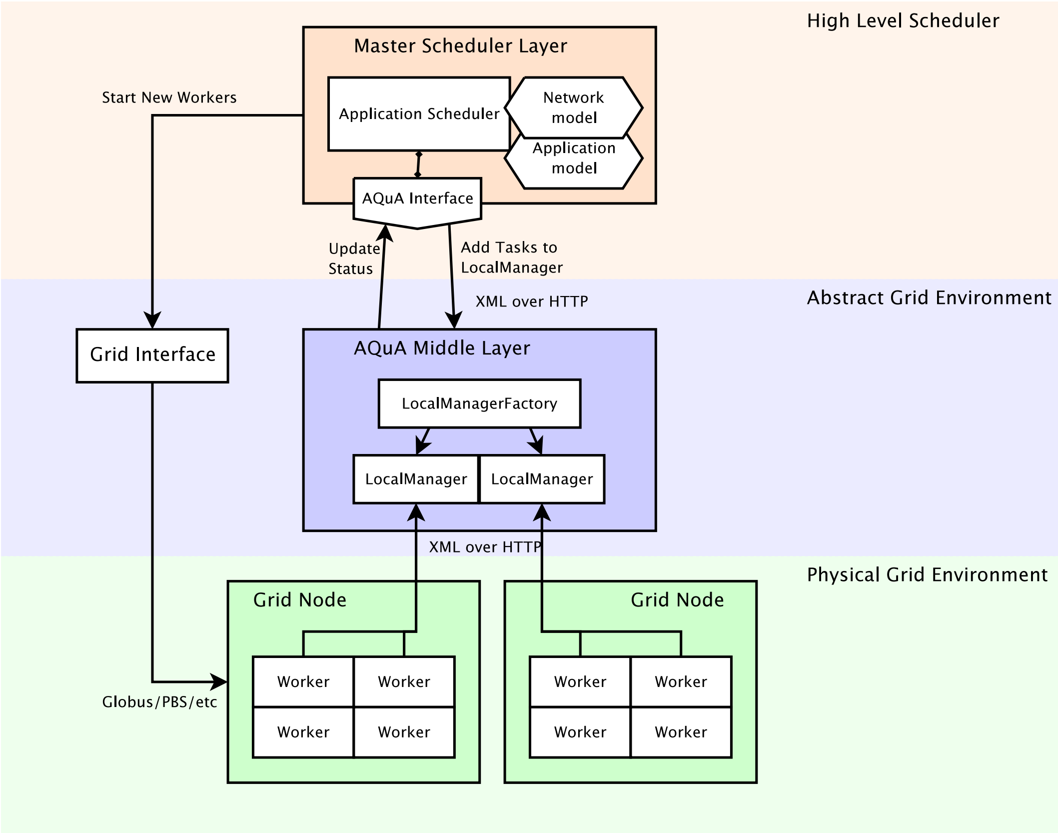

Figure 6.2: A schematic overview of the AQuA system

server, listing the tasks that the worker can run (for example, what libraries are installed), as well as the resource available. The AQuA server then returns the next suitable task from the queue for that *LocalManager*.

On top of the AQuA server is a high level scheduler that understands the processing and data requirements of tasks, as well as having a network map detailing the network speed between *LocalManager*s and the resources available. The high level scheduler is also responsible for submitting new workers when more processing power is required.

6.3.1 Description of objects used in the AQuA system

The AQuA system uses the following key components to map the actual tasks and resources. Each of these resources is identified by a unique URI[1].

LocalManagerFactory The *LocalManagerFactory* follows a factory prototype and is used to create new instances of *LocalManager* objects. This is the root component of the AQuA system, and there shall be only one *LocalManagerFactory* per AQuA instance.

LocalManager A *LocalManager* is an abstract representation of a 'node' in one of the grid systems that are served by the AQuA system. The exact physical representation of this is dependent on the grid system in question. It may represent a single machine, a cluster of machines, or an entire grid system. In the AQuA system, each *LocalManager* is a queue, so the granularity of the *LocalManager*s determines the number of queues available and limits the amount of tuning any associated scheduler can perform. A *LocalManager* has the following properties:

localmanagerfactory	A reference to the *LocalManagerFactory* that created this.
description	An arbitrary text string describing the *LocalManager*
pendingtasks	The queue of *Task*s that are waiting to run on this *LocalManager*'s resource. Stored as task URI's
activetasks	The queue of *Task*s that are currently running on this *LocalManager*'s resource. Stored as task URI's
terminatedtasks	The queue of *Task*s that have terminated from this *LocalManager*'s resource. Stored as task URI's
workers	The list of *Worker*s that are currently allocated to this *LocalManager*

[1] Universal Resource Identifier

Task A *Task* is a particular piece of work that is in the AQuA system. The application scheduler may use a more detailed model of an application task, which is broken down to a *Task* upon submission to the AQuA layer. For more details on the way application tasks are handled in the PVO, see section 6.4.2. It consists of the following elements:

localmanager	A reference to the *LocalManager* that this task is associated with.
taskdescription	The task parameters, within the AQuA system, these exist as an XML structure. See section 6.3.3 for more details.
worker	A reference to the *Worker* that the *Task* is associated with. This will not have a value if the task is not running.
activationtime	The time that the *Task* was initialised in the AQuA system.
starttime	The time that the *Task* began processing.
completiontime	The time that the *Task* ended processing.
milestone	A list of 'milestones' reached by the *Task*. A milestone has five varieties, *START* signals the task was started, *SUCCESS* signals successful completion of the task, *WARNING* signals a non-fatal error in the processing, *ERROR* signals a fatal error in the system and *MILESTONE* simply updates the current progress and gives feedback to the user or scheduler.
output	A list of output files, listed as a set of tuples of URI, checksum and file size.

Worker A *Worker* is a program that runs on a physical resource. The worker polls the *LocalManager* for *Task*s to process. It consists of the following elements:

localmanager	A reference to the *LocalManager* that this *Worker* is associated with.
numproc	The number of processors that the worker is assigned to.
allocatedtime	The time that the worker is scheduled to run for.
starttime	The time at which the worker started operating.
terminatedtime	The time at which the worker stopped operating.
lastactiontime	The last time the system received a message from the worker.
currenttask	The *Task* that the worker is currently processing

6.3.2 Interprocess communication under the AQuA system

AQuA uses XML[2] over HTTP[3] as the primary means of communication between the different layers. HTTP defines eight request methods, however we only require the use of three, GET, POST and PUT. GET is defined as a request for a resource from the server, POST is designed to give information to a resource on the server, and PUT is defined as a request to give a resource to the server. In the case of the AQuA system, the resources are XML documents that represent one of the objects described in 6.3.1 or describe some interaction with these objects. In general, performing a GET operation on a resource URI causes a XML representation of the resource to be returned. When calling a PUT or POST on the resource URI, the response depends on the XML passed to the server. For example, to create a new *LocalManager*, one must POST a '*NewLocalManagerRequest*' XML document to the *LocalManagerFactory* URI.

[2]eXtensable Markup Language. See: http://www.w3.org/XML/

[3]HyperText Transfer Protocol. See: Fielding et al. (1999)

The communication between workers and the AQuA Middle Layer is unidirectional. That is, that the workers must poll outwards to the Middle Layer to update the Middle Layer's status representation, or to fetch new jobs. This is designed in this way as the workers may be running on machines that cannot be accessed from outside the grid system. Since all the communication is powered by HTTP, if the workers cannot directly communicate with external systems, a standard HTTP proxy can be used to forward the communications.

Communication between the Master Scheduler Layer and the AQuA Middle Layer is bidirectional. If objects in the Middle Layer change status, they are passed up to the Master Scheduler, and the Master Scheduler can communicate with the Middle Layer to create new tasks. This is safe as both the Middle Layer and the Master Scheduler can be run at a location removed from the grid environment.

6.3.3 AQuA Parameters and Annotated Links

The initial design for the AQuA system required tasks to have parameters specified as an array of strings. This was designed so as to be as simple as possible, and to remove any constraints on the type of parameters specified to the workers. This became limiting however as each parameter was indistinguishable, and extra parsing work needed to be carried out by the workers. To overcome this problem, a system of parameter annotation was developed.

In the annotation system, parameters are composed of strings with associated 'annotated links'. An annotated link is composed of two elements, links and annotations. Annotations are a typed name-value pair, links are a URI with associated annotations. A single 'annotated link' consists of any number of links (with associated annotations) and any number of global annotations. An example might be an annotated link representing an input file parameter. This parameter could contain links that point to instances of the file, with annotations on each marking the permissions required to access them, and global annotations containing the file size, hash, data type and so forth.

A further enhancement that was developed was the use of multiple parameter lists. To separate logical types of parameters, a task was associated with multiple, labelled, parameter lists. For example, a list of input files.

6.3.4 REST

Representational State Transfer (REST) is an architectural style for 'hypermedia systems' (Fielding, 2000). In essence REST defines a stateless, client-server architecture wherein all operations are considered to be atomic transactions on resources. In a REST model, resources are usually described by URIs and the atomic transactions are GET, PUT, POST and DELETE (usually implemented by their HTTP counterparts).

The AQuA system as defined here can be considered as a RESTfull system (i.e. one that follows the principles of REST).

The main alternative to REST is to use a protocol such as SOAP[4] to create web services. SOAP uses a message passing architecture, usually used to wrap remote procedure calls for transport over the web via HTTP. Although REST style and web services style architectures often use HTTP as the underlying transport, in REST the HTTP operations are the operations of the architecture, whereas in web services the operations are transmitted to the server over HTTP. In web services, URIs represent objects upon which methods can be called, whereas in REST, URIs represent a resource which can be altered.

Although web services are used by a majority of grid applications, the benefits of a REST style approach for the AQuA system are clear. In a RESTfull system, clients can access and modify resources using only simple HTTP requests. Therefore the AQuA clients can run in a wide variety of locations without additional

[4]Simple Object Access Protocol. See: http://www.w3.org/TR/2000/NOTE-SOAP-20000508/

software. The simplicity of the architecture also allows for very light clients to be developed, making it easy for applications to interact with the AQuA system. Using HTTP as the core protocol makes it trivial to forward communications around the network by utilising existing HTTP proxying and caching mechanisms.

6.4 Pulsar Virtual Observatory

6.4.1 PVO Use Cases

The Pulsar Virtual Observatory is aimed at a number of different "use cases".

Experienced pulsar astronomers, who want to determine if a pulsar, or pulsar candidate, with given parameters is detectable in the PVO data catalogues. This is useful, for example, when a new pulsar candidate is detected that lies in a region covered by the PVO data sets. If the candidate pulsar can be re-detected in the PVO data, then the signal can be confirmed as a real pulsar without having to re-observe. Whilst timing observations are necessary, knowing a set of pulsars that are definitely confirmed aids in both planning of observations and in securing telescope time.

Astronomers and other scientists who are new to the field of pulsar astronomy, may want to search for sources in the PVO data sets. This is likely when a neutron star is detected, or suspected, to be lying in one of the PVO data sets. This might be due to the detection of a high energy source consistent with neutron star emission. By performing a deep targeted search on the data in a given region, the user may be able to detect new pulsars, either confirming their neutron star detection, or otherwise providing new research material.

6.4.2 PVO Scheduler

The Pulsar Virtual Observatory scheduler has gone through several design iterations before settling on the current model. The scheduler requires knowledge about the specific tasks and work-flows that are to be run as well as an understanding of the resources that are available at any time.

Initially the scheduler used Java objects to build a model of the grid and to design the available work-flows and tasks. The scheduler tasks were extensions of the AQuA task objects, however this caused a number of implementation issues as neither the AQuA layer, nor the scheduler could authoritatively restore these instances upon system startup. This issue arises because the AQuA layer has a highly simplified model of the tasks, and therefore cannot create the more detailed scheduler objects. The scheduler however needs to match its task instances to those from the AQuA layer. Therefore it was decided to separate the AQuA tasks from the scheduler tasks, and at the same time change the scheduler to model the grid and tasks using a database. Using a database to store the state of the scheduler means that the system state does not need to be re-instantiated on system restart as the database is the authoritative system state. This prevents the system from reaching a state where the 'on-disk' state is different from the 'live' state, as well as providing transactioning so that the database state is guaranteed to always be valid.

In order to simplify the scheduling process a restriction was placed on tasks so that they could produce at most one output file. This restriction is reduced by ensuring that all tasks are split into 'atomic' units, that is the minimum amount of work that could be performed. There are then, a large number of tasks in a given system, however the scheduler is designed to aggregate the tasks so that any tasks that share common input files can be aggregated and submitted to the AQuA system as a single AQuA task. This many-to-one relationship between scheduler tasks and AQuA tasks allows for a simpler work-flow for the scheduler. Task aggregation allows for greater efficiency in submitting work, as processing many tasks at once can often be quicker than processing them separately. The level of aggregation can be altered by the scheduler depending on the system load, therefore making this system more flexible.

In practise the aggregation works by combining all tasks of the same type that utilise the same input file set. The tasks are combined by creating an AQuA task with a parameter list for each of the scheduler task parameters. Then, each of the parameters from the scheduler task are inserted into the appropriate parameter list. The worker tasks read these parameter lists with a shared index for each of the scheduler tasks and if output files are produced, they are output in the order of the tasks in the parameter lists (filling in appropriate nulls if no output file is created).

6.4.3 PVO User Interface

In the PVO system, users interact only with the top level, i.e. the scheduler. This interaction is remote, and therefore the simplest solution is to provide a HTML/HTTP web interface. Since the scheduler state is stored in a database, there are two possibilities for interacting with the scheduler.

- The user interface makes requests on the scheduler directly, for example by the scheduler providing an interface directly using JSP[5]. This gives a certain level of abstraction from the underlying database, as the database technology can be changed without changing the user interface.
- Alternately, the user interface could interact directly with the database. In this instance, the user interface is entirely separate from the scheduler code. This leads to a neater implementation, and reduces the load on, and complexity of, the scheduler. In this instance the user interface directly interprets the database, so if the database technology is changed, the user interface needs to be updated.

For the Pulsar Virtual Observatory, the second option has been selected as the preferred solution. This was chosen as it prevents complicating the function

[5]JavaServer Pages. See: http://java.sun.com/products/jsp/

of the scheduler. The problems of abstracting the database technology can be minimised by using a generic database interface for the user interface, and indeed for the scheduler, that provides an implementation independent interface to the database.

The user interface can then be written as a totally independent web application, using PHP[6], or similar.

6.5 Predicting work-flow run times

Being able to predict the run-time of a work-flow, or part of a work-flow is of great benefit to both the software scheduler and to users. Users benefit from the information as they can make better informed decisions regarding the types of analysis that need to be carried out and are able to manage their research in a more efficient manner. The scheduler also benefits from this as it can predict when resources will become available and better compute the trade off between data transfer and process time.

The run-time of a work-flow depends on two parameters, the total run-time of individual tasks and the delay between submitting tasks and running tasks. Both of these can be difficult to determine, depending on the application in question.

6.5.1 Process run-time

Process run time depends on the type of task being run, the parameters of the task and the resources that the task is run on. For the pulsar search problem, most tasks are essentially simple, they are linearly (or logarithmically) dependent on the amount of data in use. The dependence on hardware is often more complex, depending mainly on the processor clock frequency, the processor on die cache size and the bandwidth between the processor and the main system memory.

[6]PHP Hypertext Preprocessor. See: http://php.net/

Because of these complexities, it is usually necessary to benchmark each hardware system in use before predictions can be accurately made. However, once bench-marked for an average size of data set, the time to run using different size data sets can be predicted with reasonable accuracy.

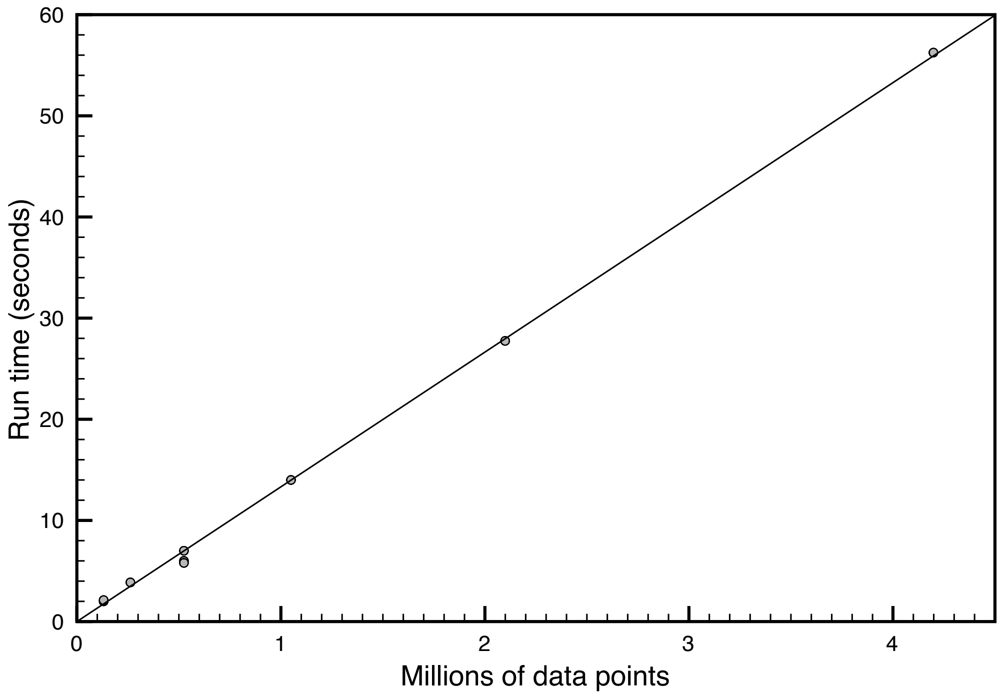

Figure 6.3: Runtime for seek (FFT search process) operating on varying length data files. Each point represents the mean of 10 trials, each processing 16 accelerations. The black line is a linear fit to these data points, passing through the origin.

Figure 6.3 gives an indication of the ability to predict run-times for the FFT part of the pulsar search process. The data were collected using a system fitted with an Intel E6300 Core 2 Duo processor and 1 GB of RAM. In all cases the data files were large enough to overflow the processor's on die cache, but small enough to fit in the system RAM. In cases where the data does not fit these limits the linearity is broken, however nearly all pulsar searches will lie within the limits for modern computers. Over the range of 2^{17} to 2^{23} data points, a linear fit through the data is good, although as it is FFT based it is likely that over a wider range an $x \log(x)$ fit would be better. Since the fit can be constrained to pass through the origin, only a single data point is required to make a good model of the run times. Clearly as more data is collected it can be added to the fit and a better model can be built.

Therefore, although the runtime on an unknown system is difficult to determine due to the number of variables, it is possible to quickly benchmark a system. In a real system it is likely that the benchmarking will be an ongoing process where the results from each completed process will go into the fit for the prediction model. Once a reasonable benchmark is achieved, predicting future process run times is simple as the behaviour is essentially linear.

6.5.2 Process start delay

The success at predicting the delay between a job being submitted to the grid and the job starting is heavily dependent on the degree that resources are guaranteed for the system. In the case of a grid where there is only a single application running, or the available resources outweigh the demand, this prediction is simple and depends only on the ability to successfully predict the run-time of currently running jobs. However in the case where multiple applications are in use this calculation becomes more difficult (Nurmi et al., 2006). Applications can share information regarding the predicted runtime of jobs in the grid queue, however this information may not be authoritative. If this grid queue is purely FIFO[7] and the queue run-time predictions are accurate, then this information is sufficient to predict the job start delay. However most grid queueing systems use some sort of fair sharing algorithm, which means that jobs may have to wait for others that have not been submitted yet. Although it is possible to estimate bounds on the queueing time (Brevik et al., 2006), this becomes complicated and is only

[7]First In-First Out, i.e. elements are removed from the queue in the order that they are input.

applicable for systems where the usage levels are stable for long time periods.

Figure 6.4 shows an analysis of the queue behaviour on the UK National Grid Service. In this analysis four parameters were measured, the number of jobs in the queue, the number of jobs running, the delay between job submission and jobs starting, and the time to transfer a file between nodes. The fourth parameter was not observed to significantly fluctuate with time, and so has not been included here. This process was run simultaneously on all four of the NGS core nodes (Manchester, Oxford, RAL and Leeds), although the data is similar for each and so only one is presented here.

It is clear from this analysis that the time between job submission and the job starting changes rapidly when large numbers of jobs are submitted. It also seems apparent that jobs are often submitted in large batches, meaning that the queue time cannot be reliably predicted ahead of time, and mean queue time is fairly meaningless.

The Pulsar Virtual Observatory project deals with this issue in a number of ways. By submitting generic workers to the grid queue, rather than the actual process tasks, one can accurately define the run-time for the submitted grid job, improving the accuracy of the grid scheduling overall. The primary benefit of separation in this way, however, is that at any given time the PVO has a pool of 'guaranteed resource', that is each worker that is running will continue to run until the pre-determined life time expires, or there are no more tasks to run. This means that even on a grid system where the start delay for jobs is not well defined, the PVO scheduler has a measure of the minimum amount of resource that will be available for each element of the grid system. Clearly if the amount of available resource needs to be expanded (by submitting new workers to the queue), the time for this expansion to take place is not well defined. When new resources become available, the distribution of tasks may need to be adjusted, however provided that the scheduler only submits enough work for the currently running workers, then this need not be a problem. By managing tasks in this

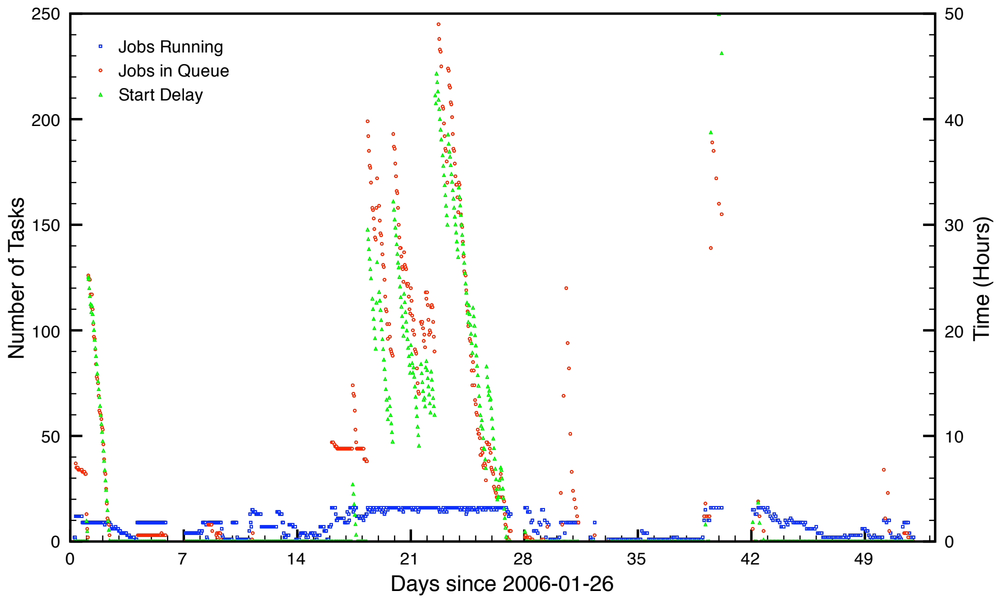

Figure 6.4: An example of the usage pattern analysis for the Manchester NGS node over a span of 52 days. The delay between job submission and the job actually starting jumps up rapidly to greater than 48 hours when large numbers of jobs are submitted.

fashion, the PVO/AQuA system does not need to understand the underlying grid environment or environments.

7

PulsarHunter

7.1 PulsarHunter Overview

As of time of writing the Sigproc package provides only basic time-domain optimisation software. To provide a complete processing package for use on the grid and elsewhere, a new processing software suite was developed, named PulsarHunter.

PulsarHunter is a Java framework for reading and analysing various types of pulsar search data. The initial focus is on filtering suspects and performing time-domain optimisation, in order to complement the features of the Sigproc package.

7.2 Design

PulsarHunter has a highly modular design. All tasks that can be performed by PulsarHunter are coded as separate modules that can be called by a PulsarHunter 'recipe'. A set of standard recipes is distributed with the software package, alternately custom recipes can be written for specific tasks. If no recipe is specified, PulsarHunter will run in interactive mode where the recipe can be input manually.

Listing 7.1: An Example PulsarHunter Recipie

```
import $ARGS1 to infile
export $ARGS2.phcf to outfile as phcf
process tune infile outfile
close infile
close outfile
```

Listed below is a selection of the most common processes, and a very brief description of their function.

- makesubints - Creates sub-integrations from a time-series. This is achieved by folding the data according to the provided period, acceleration, jerk and dispersion measure.
- tunesubints - Takes sub-integrations created by makesubints, and adjusts the phase offsets to optimise the period.
- tune - Sequentially performs the tasks of makesubints and tunesubints. This task exists for legacy support.
- phcf2candlist - Appends a PulsarHunter Candidate File to a JReaper candidate list so that it can be directly used by JReaper.
- dir2candlist - Creates a new JReaper candidate list from a directory containing any JReaper supported candidate file.
- phcfplotter - Displays a PulsarHunter Candidate File using Java Swing.
- insertfake - Adds a fake pulsar signal to the input time series.
- filter - Takes Sigproc period lists and matches the periods to form candidates. Also markes candidates that are in regions marked as interference by PulsarHunter Zap files.
- displayzapfile - Starts a interactive Java Swing GUI for editing PulsarHunter Zap files, which can be used for marking known interference signals.

Table 7.1: Summary of the data types that can be read by PulsarHunter.

Format	Read/Write	Restrictions	Description Reference
Sigproc Time Series	Both	Only 32 bit data	[1]
Presto Time Series	Read	Only 32 bit data	[2]
Sigproc .prd File	Read		[1]
EPN profiles	Both		(Lorimer et al., 1998)
Polyco	Read		[3]
JReaper Candlists	Both		[4]
PHCF Candidates	Both		[4]
PDM .ph/.aph/.sph files	Read		-

[1] See: http://sigproc.sourceforge.net/

[2] See: http://www.cv.nrao.edu/∼sransom/presto/

[3] See: http://pulsarastronomy.net/wiki/Software/PolycoFiles/

[4] See: http://pulsarastronomy.net/wiki/Software/PulsarHunter/

7.3 Performance

Traditionally Java is not thought of as a programming language optimised for performance, however development of the Java Virtual Machine means that Java processes are often as fast, or potentially faster, than equivalent programs coded in C or Fortran (Reinholtz, 2000; Bull et al., 2002). However it is important to benchmark the processing speed of PulsarHunter and compare it to that of existing software.

A common function between the Sigproc package and PulsarHunter is the folding of a time-series according to a given set of parameters. Both processes were selected to fold a 100 MB time-series file in Sigproc format with a period of 10.0 ms and export the resulting data as a small (1 kB) EPN file. Both processes were timed using the system time. PulsarHunter folded the data in an average of 1.77 ± 0.02 seconds, and Sigproc folded the data in an average of 2.20 ± 0.02 seconds.

Another comparison is between the existing time-domain optimisation software 'PDM' and PulsarHunter's recipe that simulates the process of PDM using Sigproc to de-disperse the data and PulsarHunter to fold and optimise the data. Starting from a 88 Mb SCAMP formatted time-series, both processes de-dispersed the data file into 4 sub-bands, then folded and optimised 256 sub-integrations, each 256 bins wide. PDM completed the task in 41.0 ± 0.1 seconds, PulsarHunter in 19.4 ± 0.1 seconds.

Although the tasks used different code, and produced slightly different outputs, in both cases the resultant output from either software was equivalent. Clearly the tasks performed with PulsarHunter do not take longer to run, and indeed are significantly faster in both cases. It is also important to note that this was achieved without writing heavily optimised code, decreasing the difficulty in debugging and maintaining the software.

7.4 Future Development

As well as being used as part of the Pulsar Virtual Observatory processing routines, PulsarHunter has come into use with several more traditional searches. A re-processing of the Parkes Multibeam Pulsar Survey, using time-domain coherent acceleration search routines is being carried out at Jodrell Bank. PulsarHunter is being used for the time-domain analysis and candidate confirmation part of this work. It was chosen for several reasons, in particular as it is able to operate on the partially dispersed time-series that are produced as part of the processing, therefore not requiring any extra de-dispersion steps during the time-domain optimisation.

It is also being used as part of the Methanol Multi-Beam Survey for pulsars, a pulsar survey at 6.5 GHz (the location of a prominent methanol maser line), searching the inner regions of the galaxy for pulsars that would be hidden by scattering or dispersion at lower frequencies.

New developments to the software are planned, a few examples are listed below:

- Dedispersion algorithm. The current sigproc de-dispersion algorithm is not optimally efficient for large surveys. A prototype de-dispersion algorithm has demonstrated that de-dispersion speed can be increased by working on several DM values at once.

- Working with more data formats. Handling of the Presto data format can be improved, as well as adding support to 1, 2, 4, 8 and 16 bit data types. It is intended that the PSRFITS data format also be supported.

- Development of a entirely Java pulsar processing work-flow. This is a potential long term goal, with the aim of allowing pulsar processing to be run on any machine that is capable of supporting the Java Runtime Environment. This could enable use of resources that are currently unavailable, such as computer clusters running Microsoft Windows.

8

Conclusions

8.1 Summary of Computational Aspects

8.1.1 New Software Tools for Pulsar Searching

Several new tools for processing pulsar search data have been developed as part of this work.

PulsarHunter is a new software package for time domain optimisation of pulsar candidates. It provides similar functionality to previous systems, however it is designed to operate on a wide range of input data, allowing for use in many different surveys.

JReaper is a new graphical candidate selection system, designed to aid the user distinguish between pulsars and other signals in the output of pulsar search processing. The design of this system is based on a highly successful predecessor, however there are many additional features. In particular, the development of a candidate scoring system allows for automatic ranking of candidates in an alternate way to the traditional signal-to-noise ratio. This has been shown to be effective at selecting pulsars over noise and interference signals, although it is best used in addition to, rather than replacing, traditional techniques.

8.1.2 Pulsar Searching using Grid Computing

The pulsar search process is ideal for running in a grid environment, as each task is independent of parallel tasks. The Pulsar Virtual Observatory (PVO) project was created with the aim of adapting pulsar search processing for a grid environment. This has involved development of the pulsar search tools as well as an architecture for scheduling and queueing tasks. Because of the turbulent state of grid computing systems it was decided to develop a queueing system for arbitrary grid systems, named the Abstract Queueing Architecture (AQuA). This system has been designed for maximal compatibility, therefore all client-server interaction is handled by client initiated HTTP communications, which can be easily transported to all nodes in the grid via simple HTTP proxying. The PVO system also aims to be able to estimate ahead of time the amount of time that is required to process the given tasks. This is difficult in a public grid environment, where there is no guarantee of resource. Therefore the AQuA system is designed to provide guaranteed resource even when the grid does not support it. This is achieved by submitting generic workers scripts to the grid, which fetch tasks to process from the AQuA system. Analysis of typical pulsar tasks shows that once bench marked, prediction of the process run time is simple and dependent linearly on the size of the provided data file.

8.2 Summary of Astronomical Findings

This work has analysed data from two pulsar surveys covering a 210° by 10° degree strip of the galactic plane. Candidate detections from a previous processing of the Parkes Multibeam Pulsar Survey have been analysed using the new JReaper graphical candidate selection tool. Additionally, the processing of data from the Perseus Arm Pulsar Survey has also been carried out, creating new candidates that have also been analysed in the same way.

In total these analyses have lead to the discovery of 40 previously unknown pulsars, more than 2% of the known pulsar population. Of these, there are four young pulsars (ages less than 100 kyr), a pulsar with an interesting inter-pulse, a probable double neutron star system and a pulsar with an unusual nulling phenomenon.

The binary pulsar J1753-2243 appears to be the tenth known double neutron star system. It has a highly eccentric orbit, however not as eccentric as the similar pulsar PSR J1811-1736. This apparent lack of eccentricity puts it at odds with the period-eccentricity relationship suggested by Faulkner et al. (2005).

The young pulsar PSR J1849-0020 is spatially coincident with the known supernova remnant SNR G32.45+0.1. Distance estimates to the pulsar and the remnant are both large and agree within the stated errors. If the association is valid, the transverse velocity must be of the order of 200 km/s, well within the typical distribution of pulsar velocities. Although a chance alignment is possible, future measurement of a proper motion of the pulsar would either strongly support or rule out the association.

PSR J0627+0705 has been observed with a large inter-pulse approximately 180° from the main pulse. Both peaks are sufficiently polarised that a polarisation position angle swing is determinable. By fitting the rotating vector model to the data, the geometry of the pulsar can be constrained. As expected the pulsar is an orthogonal rotator, with the angle between the rotational axis and the magnetic pole of 96.5°.

PSR J0623+0341 shows nulling on a time-scale of tens of minutes. This makes it one of the longest time-scale nulling pulsars or the shortest timescale intermittent pulsar. Detection of more pulsars filling in this phase space gap supports the theory that nulling phenomena on all time-scales are related.

Although the number of pulsars detected is insufficient to affect the general population statistics, the above individual pulsars are scientifically interesting, and the discoveries can be seen as a good demonstration of the effectiveness of the new search techniques developed here.

8.3 Future Work

The most significant obstacle in the path to the Pulsar Virtual Observatory is the auditing and accounting. Currently, no effort is made to track the amount of resource used by different users.

The main challenge in this is that the AQuA system is designed to use a range of different grid resources. If the AQuA system uses a users proxy certificate to authenticate with the grid, it has to be known that all of the grids will accept that certificate. Currently the PVO system authenticates as itself and therefore has to handle bookkeeping internally, which is not the preferred method for many grid resources.

If this is to be resolved then a global science certificate authority is required, which will be accepted by all users and grid resources. This may be possible, probably through automated certification of academics by their institutions.

Another important aspect of future work will be the ongoing monitoring and timing of the pulsars discovered in this work. In particular, continued timing of PSR J1753-2243 should allow for the measurement of the system mass and possibly the individual masses.

8.4 Project Outcome: Addressing the Aims

In this final section of this thesis, the project will be compared against the aims set out in section 1.6.2 will be considered. Each of the aims are addressed below:

- *Investigate current methods of searching for new pulsars.*

  Processing of the Perseus Arm Pulsar Survey using traditional pulsar search techniques has shown the strengths and weakness of these methods. In addition to this, working with the data from the Parkes Multibeam Pulsar Survey has shown that archival data is highly useful for searching and confirming pulsars.

- *Determine means for applying pulsar search work-flows to grid computer systems.*

  Application of pulsar search methods to a grid environment has been facilitated by several new software tools. The AQuA system manages job submission and tracking of the data through the pulsar work-flow.

  In order to use a wide range of data types, Sigproc was chosen as a flexible software package for pulsar searching. To complete the full pulsar search work-flow requires extra tools however, therefore PulsarHunter was developed as a complementary software package to Sigproc.

  As well as this, the scoring system that was integrated into JReaper was developed to aid unskilled users who might be processing data using the pulsar grid software.

- *Develop a prototype system for scheduling pulsar tasks in a grid environment.*

  Efficient scheduling requires an understanding of the run-time of tasks in the system. It has been shown that the runtime of individual pulsar tasks is simple to compute, however the delay between the submission of a job and the start of the job running is non-trivial. Therefore the AQuA system manages its available resources through generic workers, significantly reducing the scheduling problem in AQuA based systems.

  A scheduler for the PVO has been developed, allowing for a basic interface to the AQuA system, however it does not yet implement the features listed above.

Overall, significant progress has been made towards the realisation of a Pulsar Virtual Observatory. In doing this, improvements to pulsar search techniques have lead to the discovery of new pulsars and will also benefit the ongoing search for pulsars.

References

Anderson, D. P.: 2004, in *Fifth IEEE/ACM International Workshop on Grid Computing (GRID'04)*, pp 4–10

Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., and Werthimer, D.: 2002, *Commun. ACM* **45(11)**, 56

Anderson, P. W. and Itoh, N.: 1975, *Nature* **256**, 25

Backer, D. C.: 1970a, *Nature* **228**, 1297

Backer, D. C.: 1970b, *Nature* **228**, 42

Bisnovatyi-Kogan, G. S. and Komberg, B. V.: 1974, *Sov. Astron.* **18**, 217

Blandford, R. and Teukolsky, S. A.: 1976, *ApJ* **205**, 580

Brevik, J., Nurmi, D., and Wolski, R.: 2006, in *IISWC*, pp 213–224

Brooke, J., Pickles, S., Carr, P., and Kramer, M.: 2007, in I. J. Taylor, E. Deelman, D. B. Gannon, and M. Shields (eds.), *Workflows for e-Science: Scientific Workflows for Grids*, pp 60–79, Springer London

Bull, J. M., Smith, L. A., Ball, C., Pottage, L., and Freeman, R.: 2002, *Concurrency and Computation: Practice and Experience* **15**, 417

Burgay, M., D'Amico, N., Possenti, A., Manchester, R. N., Lyne, A. G., Joshi, B. C., McLaughlin, M. A., Kramer, M., Sarkissian, J. M., Camilo, F., Kalogera, V., Kim, C., and Lorimer, D. R.: 2003, *Nature* **426**, 531

Champion, D. J., Lorimer, D. R., McLaughlin, M. A., Xilouris, K. M., Arzoumanian, Z., Freire, P. C. C., Lommen, A. N., Cordes, J. M., and Camilo, F.: 2005, *MNRAS* **363**, 929

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J.: 1998, *AJ* **115**, 1693

Cooley, J. W. and Tukey, J. W.: 1965, *Math. Comput.* **19**, 297

Cordes, J. M. and Lazio, T. J. W.: 2002, astro-ph/0207156

Cordes, J. M. and McLaughlin, M. A.: 2003, *ApJ* **596**, 1142

Corongiu, A., Kramer, M., Stappers, B. W., Lyne, A. G., Jessner, A., Possenti, A., D'Amico, N., and Löhmer, O.: 2007, *A&A* **462**, 703

Damour, T. and Deruelle, N.: 1985, *Ann. Inst. H. Poincaré (Physique Théorique)* **43**, 107

Damour, T. and Deruelle, N.: 1986, *Ann. Inst. H. Poincaré (Physique Théorique)* **44**, 263

Deutsch, L. P.: 1996, *RFC 1952: GZIP file format specification version 4.3*

Dewi, J. D. M., Podsiadlowski, P., and Pols, O. R.: 2005, *MNRAS* **363**, L71

Duda, R. O. and Hart, P. E.: 1972, *Comm. ACM* **15**, 1

Edwards, R. T., Bailes, M., van Straten, W., and Britton, M. C.: 2001, *MNRAS* **326**, 358

Everett, J. E. and Weisberg, J. M.: 2001, *ApJ* **553**, 341

Faulkner, A. J.: 2004, *Ph.D. thesis*, University of Manchester

Faulkner, A. J., Kramer, M., Lyne, A. G., Manchester, R. N., McLaughlin, M. A., Stairs, I. H., Hobbs, G., Possenti, A., Lorimer, D. R., D'Amico, N., Camilo, F., and Burgay, M.: 2005, *ApJ* **618**, L119

Faulkner, A. J., Stairs, I. H., Kramer, M., Lyne, A. G., Hobbs, G., Possenti, A., Lorimer, D. R., Manchester, R. N., McLaughlin, M. A., D'Amico, N., Camilo, F., and Burgay, M.: 2004, *MNRAS* **355**, 147

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee., T.: 1999, *RFC 2616: Hypertext Transfer Protocol – HTTP/1.1*

Fielding, R. T.: 2000, *Ph.D. thesis*, University of California, Irvine

Foster, R. S. and Backer, D. C.: 1990, *ApJ* **361**, 300

Gaensler, B. M. and Johnston, S.: 1995a, *MNRAS* **275**, L73

Gaensler, B. M. and Johnston, S.: 1995b, *MNRAS* **277**, 1243

Gold, T.: 1968, *Nature* **218**, 731

Green, D. A.: 2006, *A Catalogue of Galactic Supernova Remnants (2006 April version)*, http://www.mrao.cam.ac.uk/surveys/snrs/

Helfand, D. J., Manchester, R. N., and Taylor, J. H.: 1975, *ApJ* **198**, 661

Hessels, J. W. T., Ransom, S. M., Stairs, I. H., Freire, P. C. C., Kaspi, V. M., and Camilo, F.: 2006, *Science* **311**, 1901

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., and Collins, R. A.: 1968, *Nature* **217**, 709

Hobbs, G., Faulkner, A., Stairs, I. H., Camilo, F., Manchester, R. N., Lyne, A. G., Kramer, M., D'Amico, N., Kaspi, V. M., Possenti, A., McLaughlin, M. A., Lorimer, D. R., Burgay, M., Joshi, B. C., and Crawford, F.: 2004, *MNRAS* **352**, 1439

Hobbs, G., Lorimer, D. R., Lyne, A. G., and Kramer, M.: 2005, *MNRAS* **360**, 974

Hobbs, G., Lyne, A. G., Kramer, M., Martin, C. E., and Jordan, C.: 2004, *MNRAS* **353**, 1311

Hough, P. V. C.: 1960, *US Patent Number 3069653: Method and means for recognising complex patterns*

Hulse, R. A. and Taylor, J. H.: 1975, *ApJ* **195**, L51

Jaffe, A. H. and Backer, D. C.: 2003, *ApJ* **583**, 616

Jenet, F. A., Hobbs, G. B., Lee, K. J., and Manchester, R. N.: 2005, *ApJ* **625**, L123

Johnston, S., Manchester, R. N., Lyne, A. G., Bailes, M., Kaspi, V. M., Qiao, G., and D'Amico, N.: 1992, *ApJ* **387**, L37

Kaspi, V. M., Lyne, A. G., Manchester, R. N., Crawford, F., Camilo, F., Bell, J. F., D'Amico, N., Stairs, I. H., McKay, N. P. F., Morris, D. J., and Possenti, A.: 2000, *ApJ* **543**, 321

Komesaroff, M. M.: 1970, *Nature* **225**, 612

Konacki, M., Wolszczan, A., and Stairs, I. H.: 2003, *ApJ* **589**, 495

Kramer, M., Bell, J. F., Manchester, R. N., Lyne, A. G., Camilo, F., Stairs, I. H., D'Amico, N., Kaspi, V. M., Hobbs, G., Morris, D. J., Crawford, F., Possenti, A., Joshi, B. C., McLaughlin, M. A., Lorimer, D. R., and Faulkner, A. J.: 2003, *MNRAS* **342**, 1299

Kramer, M., Lyne, A. G., O'Brien, J. T., Jordan, C. A., and Lorimer, D. R.: 2006, *Science* **312**, 549

Kramer, M., Stairs, I. H., Manchester, R. N., McLaughlin, M. A., Lyne, A. G., Ferdman, R. D., Burgay, M., Lorimer, D. R., Possenti, A., D'Amico, N., Sarkissian, J. M., Reynolds, J. E., Joshi, B. C., Freire, P. C. C., and Camilo, F.: 2006, *Nature*, submitted

Lorimer, D. R., Faulkner, A., Lyne, A. G., Manchester, R. N., Kramer, M., McLaughlin, M. A., Hobbs, G., Possenti, A., Stairs, I. H., Camilo, F., Burgay, M., D'Amico, N., Corongui, A., and Crawford, F.: 2006a, *MNRAS*, submitted

Lorimer, D. R., Jessner, A., Seiradakis, J. H., Lyne, A. G., D'Amico, N., Athanasopoulos, A. Xilouris, K. M., Kramer, M., and Wielebinski, R.: 1998, *A&AS* **128**, 541

Lorimer, D. R., Stairs, I. H., Freire, P. C. C., Cordes, J. M., Camilo, F., Faulkner, A. J., Lyne, A. G., Nice, D. J., Ransom, S. M., Arzoumanian, Z., Manchester, R. N., Champion, D. J., van Leeuwen, J., McLaughlin, M. A., Ramachandran, R., Hessels, J. W. T., Vlemmings, W., Deshpande, A. A., Bhat, N. D. R., Chatterjee, S., Han, J. L., Gaensler, B. M., Kasian, L., Deneva, J. S., Reid, B., Lazio, T. J. W., Kaspi, V. M., Crawford, F., Lommen, A. N., Backer, D. C., Kramer, M., Stappers, B. W., Hobbs, G. B., Possenti, A., D'Amico, N., and Burgay, M.: 2006b, *ApJ* **640**, 428

Lyne, A. G., Burgay, M., Kramer, M., Possenti, A., Manchester, R. N., Camilo, F., McLaughlin, M. A., Lorimer, D. R., D'Amico, N., Joshi, B. C., Reynolds, J., and Freire, P. C. C.: 2004, *Science* **303**, 1153

Lyne, A. G., Camilo, F., Manchester, R. N., Bell, J. F., Kaspi, V. M., D'Amico, N., McKay, N. P. F., Crawford, F., Morris, D. J., Sheppard, D. C., and Stairs, I. H.: 2000, *MNRAS* **312**, 698

Lyne, A. G. and Smith, F. G.: 2005, *Pulsar Astronomy, 3rd ed.*, Cambridge University Press, Cambridge

Manchester, R. N., Hobbs, G. B., Teoh, A., and Hobbs, M.: 2005, *AJ* **129**, 1993

Manchester, R. N., Lyne, A. G., Camilo, F., Bell, J. F., Kaspi, V. M., D'Amico, N., McKay, N. P. F., Crawford, F., Stairs, I. H., Possenti, A., Morris, D. J., and Sheppard, D. C.: 2001, *MNRAS* **328**, 17

Manchester, R. N. and Taylor, J. H.: 1972, *Astrophys. Lett.* **10**, 67

McLaughlin, M. A., Lyne, A. G., Lorimer, D. R., Kramer, M., Faulkner, A. J., Manchester, R. N., Cordes, J. M., Camilo, F., Possenti, A., Stairs, I. H., Hobbs, G., D'Amico, N., Burgay, M., and O'Brien, J. T.: 2006, *Nature* **439**, 817

Middleditch, J. and Kristian, J.: 1984, *ApJ* **279**, 157

Morris, D. J., Hobbs, G., Lyne, A. G., Stairs, I. H., Camilo, F., Manchester, R. N., Possenti, A., Bell, J. F., Kaspi, V. M., Amico, N. D., McKay, N. P. F., Crawford, F., and Kramer, M.: 2002, *MNRAS* **335**, 275

Nurmi, D., Mandal, A., Brevik, J., Koelbel, C., Wolski, R., and Kennedy, K.: 2006, in *SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, p. 119, ACM Press, New York, NY, USA

Peters, P. C.: 1964, *Phys. Rev.* **136**, 1224

Pickles, S. M., Brooke, J. M., Costen, F. C., Gabriel, E., Muller, M., Resch, M., and Ord, S. M.: 2001, *Future Generation Computer Systems* **17**, 911

Pilkington, J. D. H., Hewish, A., Bell, S. J., and Cole, T. W.: 1968, *Nature* **218**, 126

Podsiadlowski, P., Dewi, J. D. M., Lesaffre, P., Miller, J. C., Newton, W. G., and Stone, J. R.: 2005, *MNRAS* **361**, 1243

Prince, T. A., Anderson, S. B., Kulkarni, S. R., and Wolszczan, W.: 1991, *ApJ* **374**, L41

Radhakrishnan, V. and Cooke, D. J.: 1969, *Astrophys. Lett.* **3**, 225

Ransom, S. M., Eikenberry, S. S., and Middleditch, J.: 2002, *AJ* **124**, 1788

Ransom, S. M., Hessels, J. W. T., Stairs, I. H., Freire, P. C. C., Camilo, F., Kaspi, V. M., and Kaplan, D. L.: 2005, *Science* **307**, 892

Reinholtz, K.: 2000, *SIGPLAN Not.* **35(2)**, 25

Robertson, H. P.: 1938, *Ann. Math.* **38**, 101

Ruderman, M. A. and Sutherland, P. G.: 1975, *ApJ* **196**, 51

Shibazaki, N., Murakami, T., Shaham, J., and Nomoto, K.: 1989, *Nature* **342**, 656

Shull, J. M., Fesen, R. A., and Saken, J. M.: 1989, *ApJ* **346**, 860

Staelin, D. H.: 1969, *Proc. I. E. E. E.* **57**, 724

Stairs, I. H., Lyne, A. G., and Shemar, S.: 2000, *Nature* **406**, 484

Staveley-Smith, L., Wilson, W. E., Bird, T. S., Disney, M. J., Ekers, R. D., Freeman, K. C., Haynes, R. F., Sinclair, M. W., Vaile, R. A., Webster, R. L., and Wright, A. E.: 1996, *Proc. Astr. Soc. Aust.* **13**, 243

Taylor, J. H.: 1974, *A&AS* **15**, 367

Taylor, J. H. and Weisberg, J. M.: 1989, *ApJ* **345**, 434

Thorsett, S. E. and Chakrabarty, D.: 1999, *ApJ* **512**, 288

Ueno, M., Yamaguchi, H., Koyama, K., Bamba, A., Yamauchi, S., and Ebisawa, K.: 2005, in *X-Ray and Radio Connections (eds. L.O. Sjouwerman and K.K Dyer) Published electronically by NRAO, http://www.aoc.nrao.edu/events/xraydio Held 3-6 February 2004 in Santa Fe, New Mexico, USA, (E4.18) 5 pages*

Wang, N., Manchester, R. N., and Johnston, S.: 2007, *MNRAS* **377**, 1383

Weisberg, J. M. and Taylor, J. H.: 2005, in F. A. Rasio and I. H. Stairs (eds.), *Binary Radio Pulsars*, Vol. 328 of *Astronomical Society of the Pacific Conference Series*, p. 25

Yamaguchi, H., Ueno, M., Koyama, K., Bamba, A., and Yamauchi, S.: 2004, *PASJ* **56**, 1059