Measuring the Gravitational Wave Background using Precision Pulsar Timing

by

Paul B. Demorest

B.A. (University of California, Berkeley) 2000
M.A. (University of California, Berkeley) 2002

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Physics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Donald Backer, Co-chair
Professor Steven Boggs, Co-chair
Professor William Holzapfel
Professor Carl Heiles

Spring 2007

The dissertation of Paul B. Demorest is approved:

Co-chair Date

Co-chair Date

Date

Date

University of California, Berkeley

Spring 2007

Measuring the Gravitational Wave Background using Precision Pulsar Timing

Abstract

Measuring the Gravitational Wave Background using Precision Pulsar Timing

by

Paul B. Demorest

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Donald Backer, Co-chair

Professor Steven Boggs, Co-chair

We investigate the possibility of using high precision timing measurements of radio pulsars to constrain or detect the stochastic gravitational wave background (GWB). Improved algorithms are presented for more accurately determining the pulse times of arrival at Earth and characterizing pulse profile shape variation. Next, we describe the design and construction of a new set of pulsar backends based on clusters of standard personal computers. These machines, the Astronomy Signal Processors (ASPs), coherently correct the pulse broadening caused by interstellar plasma dispersion. Since they are based in software, they are inherently more flexible than previous generations of pulsar data recorders. In addition, they provide increased bandwidth and quantization accuracy. We apply these methods to 2.5 years worth

of millisecond pulsar data recorded with the ASP systems at the Arecibo and Green Bank Telescopes, and present pulse profiles, dispersion measure variation, and timing model parameters derived from this data.

We then develop the theory of gravitational wave detection using pulsar timing, and show how data from several pulsars can be combined into a pulsar timing array for this purpose. In particular, a new method of accounting for the effect of the timing model fit on the gravitational wave signal is presented. This method incorporates the exact timing model basis functions without relying on Monte Carlo simulation. We apply this method to the 2.5 year dataset previously mentioned and derive a gravitational wave limit of $h_c(1\ \mathrm{yr}^{-1}) < 2.46 \times 10^{-14}$. Finally, we study the 20-year timing record of PSR B1937+21 and obtain information on how severely interstellar medium effects will compromise future GWB detection. We predict that these developments, in conjunction with historical data, could provide the first successful direct gravitational wave detection on a 5–10 year timescale.

Professor Donald Backer
Dissertation Committee Co-chair

Professor Steven Boggs
Dissertation Committee Co-chair

Dedicated to Sara,

for her love, support,

and especially, patience.

Contents

List of Figures

List of Tables

Acknowledgments

The work presented in this dissertation would not have been possible without the guidance, assistance and support of a large number of people. In this section, I will attempt to give credit where credit is due, and thank those who have helped me along the way. First among these is of course my thesis advisor, Don Backer. Working with Don has been a great experience for me: He is extremely knowledgeable in both scientific and technical matters, and always has valuable insights to share. I've enjoyed a large amount of freedom to pursue the research presented here on my own terms, with guidance available when needed. Most importantly, I've always felt that Don was on "my side" and had my best interests as a student and young researcher in mind – a quality that shouldn't be taken for granted in advisor-student relationships!

Many other Berkeley people have contributed in various ways to my education: Dan Werthimer deserves credit for introducing me to the field of radio astronomy. I have to thank Matt Dexter for teaching me how to do electronics the right way, and always letting me interrupt whatever else he was working on to discuss my design issues. I enjoyed collaborating with Ramachandran during his time at Berkeley – much of Chapter 6 is the result of our combined efforts and interactions. I also thank the members of my dissertation committee: Don, Steve Boggs, Carl Heiles and Bill Holzapfel for their helpful comments about this work.

The construction of the ASP systems discussed in Chapter 3 was a result of the efforts of many people. The SERENDIP5 board was primarily due to the work of

Aaron Parsons and Jeff Mock. For their roles in writing the ARTS set of software, I have to share credit with Ramach and Rob Ferdman. The help of our collaborators around the world was instrumental in making the machines successful, especially Ingrid Stairs, David Nice and Ismaël Cognard. Ismaël deserves extra credit for giving me a reason to take two trips to France, definitely highlights of my graduate career! I also thank him and Gilles Theureau for their hospitality during my visits. The staff of the Arecibo, Green Bank and Nançay observatories were of great assistance in the installation and maintenance of our systems, as well as making the observations possible.

Finally, completing graduate school would not be possible without the support of family and friends. I thank my parents, David and Nancy Demorest for their support of my undergraduate education, for encouraging me to pursue my own interests, and for always believing in my ability to make it. Last but definitely not least, my wife Sara has been my constant companion throughout it all, providing love, support, and the necessary encouragement for me to see this to completion. She has been my inspiration to move forward, and for this I dedicate this work to her.

Chapter 1

Introduction

1.1 Introduction

The subject of this work is high-precision pulsar timing, and its application to a topic currently of great interest to the field of experimental physics and astronomy – the detection of gravitational radiation, or waves propagating in the structure of space-time. The possibility that pulsars may be the instruments by which gravitational radiation is first detected is fitting, as there has been a long association between the study of the two. As neutron stars, pulsars are the most dense objects known to us that have not collapsed to a black hole. This extreme environment allows stringent testing of general relativity and has already provided the only (indirect) experimental evidence for gravitational radiation. Pulsar timing results are arguably the most precise astronomical measurements performed. The spins of millisecond pulsars can be measured over spans of up to 20 years while accounting for every single rotation of the star, and providing frequency measurements accurate to 1 part in 10^{15}. This precision requires that a large amount of care and effort be put into both experimental apparatus with which to detect the signal and data analysis and signal processing techniques with which to analyze it. As such, these topics make up a large portion of what you are about to read. Here, we will begin with a overview of both pulsars and gravitational radiation, then summarize the structure and contents of the following chapters.

1.2 Pulsar Basics

The first pulsating astronomical radio sources were discovered in 1967 at Cambridge by J .Bell and A. Hewish. These radio pulses appeared very regular in time, with a period of approximately 1 s, but were clearly of astronomical origin due to their fixed position on the sky. Despite the famous initial designation as “LGM,” it became clear that these were of natural (non-intelligent) origin. Soon thereafter, consensus was reached that these were in fact spinning – not pulsating, although the name stuck – neutron stars (Gold, 1968), a type of object which up to that point had been purely theoretical.

Our current basic physical picture is as follows (see for example the texts of Lyne and Graham-Smith (2006) or Lorimer and Kramer (2004)): A pulsar is a neutron star, a collapsed remnant of a main sequence star whose “life” ended in a supernova. The core of this star was massive enough for gravity to overcome the electron degeneracy pressure that otherwise would have supported it as a white dwarf. Yet it was not massive enough to collapse completely to a black hole. The electrons and protons merged into neutrons, whose degeneracy pressure is strong enough to balance gravity. This leaves a very unique remnant – a macroscopic object with approximately the density of nuclear matter. The canonical neutron star has a mass of 1.4 $M_\odot$ and a radius of 10 km. The collapse of the progenitor star concentrates the existing angular momentum and magnetic flux into this small region, leading to typical spin periods on the order of 1 second and surface magnetic fields of 10^{8-12} gauss. Along the magnetic field lines, a highly energized plasma flows. In a process still poorly understood, this plasma emits at radio wavelengths as the field curves away from the pole, leading to the pulses we observe at Earth. If the magnetic axis is inclined with respect to the rotation axis, and the viewing angle is appropriate, we will observe one radio pulse per rotation of the star.

The magnetic field of a pulsar can be measured by observing its spin-down rate and applying the basic formula for the energy radiated by a spinning magnetic dipole. The observational process by which this is done is known as pulsar timing. Timing offers a wealth of information in addition to the spin-down rate, such as estimates of the motion of the pulsar through space, and measurements of binary system parameters. In addition to the neutron star and its immediate environment, the radio pulses travel through, and are affected by, the interstellar medium (ISM). This gives us a unique probe of the ISM, and many of the results presented here are measurements of ISM parameters. Along the same lines, any variation in the space-time metric along the path from a pulsar to Earth will affect the pulse's travel time. It is these small fluctuations due to gravitational radiation that we are ultimately interested in detecting.

1.3 Gravitational Radiation Basics

The theoretical possibility of gravitational radiation (GW) was recognized almost immediately following the formulation of Einstein's general theory of relativity. Indeed, it is easy to show that Einstein's equation for the space-time metric permits freely propagating wave solutions (see the standard texts by Schutz (1985) and Misner et al. (1973)). Detecting these waves, on the other hand, is a ongoing challenge for experimental physicists and has yet to be accomplished. In fact, the only experimental evidence of any sort for the existence of GW comes from pulsar timing! The double neutron star binary system PSR B1913+16 was discovered by Hulse and Taylor in 1974, and its orbit was demonstrated to be decaying at exactly the rate predicted for loss of energy due to GW. A direct detection of GW would be a major milestone for physics and would further confirm that our understanding of gravity as expressed in the theory of general relativity is correct. Furthermore, access to each new electromagnetic wavelength range, from radio to gamma rays, has resulted in new discoveries in astrophysics. The detection GW will give us a entirely new method with which to view the cosmos.

Most early GW detection experiments involved the use of some form of resonant detector: A system of masses where at a certain frequency, incident GW would excite vibrations in the system that could then be detected. However, this line of thinking has for the most part reached its endpoint, and further gains in sensitivity do not appear to be possible. All new GW detectors operate on the principle of probing the gravitational metric through the use of electromagnetic radiation. Earth-based detectors employing this principle use laser interferometers. Any GW incident upon the detector will alter the path length down the interferometer arms and can potentially be detected via the changing interference pattern. The currently-operating Laser Interferometer Gravitational-wave Observatory (LIGO[1]) is the state of the art in this field. The proposed Laser Interferometer Space Antenna (LISA) follows a similar plan, but will put the detector in space, sending laser signals between a set of three spacecraft. This removes background effects due to the noisy vibrational environment on Earth, and allows the measurement to be pushed to lower frequencies (near 10^{-3} Hz) where the GW signals are stronger.

The idea of detecting gravitational radiation using pulsar timing was first put forward by Detweiler (1979). This operates on the same basic principle as the interferometers: Any GW passing between a pulsar and Earth will change the path length that the light must travel. However in this case, we only have control of one end of

[1] `http://www.ligo.caltech.edu/`

the detector. Since we need to determine the behavior of the pulsar itself from the data, as well as try to detect GW, it is impossible to make a definitive measurement of GW in this way, although we can place limits on the total GW strength. This method is sensitive to GW with periods near the timespan of the data, $\sim 10^{-9}$ Hz for experiments lasting 1–10 years. A large conceptual improvement was made by Hellings and Downs (1983), who realized that when timing a set of pulsars, GW will cause correlations in the timing records. Given a GW source pattern (for example, an isotropic stochastic background), the correlation will have a unique signature as a function of separation angle for each pair of pulsars. No intrinsic pulsar effects would be able to mimic this, so this method offers the chance of unambiguously detecting GW. This type of experiment is known as a pulsar timing array (PTA). The final necessary ingredient came with the discovery of millisecond pulsars (MSPs; Backer et al. (1982)). MSPs are pulsars which have been spun up due to accretion, and have spin periods in the 1–10 ms range. The resulting sharp pulses, as well as the intrinsic stability of these sources make them excellent clocks. We are now reaching a era in which the length of MSP timing records, the number of known MSPs, and improvements in pulsar instrumentation make a positive detection of GW over the next 5–10 years plausible.

1.4 Overview

The following chapters each deal with different aspects of the pulsar timing process. Here we present a brief description of each. To begin, in Chapter 2 we discuss what is typically the first step in pulsar timing data analysis, the generation of pulse times of arrival (TOAs) from the raw timing data. We summarize traditional methods and present a unified framework for determining both TOAs and template profiles (average flux as a function of rotational phase) in a single step. We also develop a principal components based procedure for characterizing profile shape variation in a dataset, an effect which can negatively influence timing results. Finally, we present a new polarization self-calibration method that can be applied to timing data.

In Chapter 3, we present the design of a recently built series of coherent dedispersion pulsar backends. These machines, the Astronomy Signal Processors (ASPs), record raw data from a radio telescope and coherently remove the effect of interstellar dispersion from the pulsar signal. This is a necessary step in order to achieve high-precision ($< 1\ \mu$s) timing results. The ASPs are based upon real-time signal processing software running in a "Beowulf" cluster of personal computers, as opposed to custom electronics as their predecessors were. This design provides a large amount of flexibility, as well as increased bandwidth and quantization accuracy over previous systems. We also present an analysis of several types of systematic timing error that are directly related to pulsar instrumentation.

The ASP systems were installed at several radio observatories in 2004. In Chapter 4, we present a timing analysis of 16 millisecond pulsars that we have been regularly observing with the new systems since that time. In this analysis, we apply the data processing strategies discussed in Chapter 2. The key result of this chapter is the timing residuals, which we further analyze in the following sections. We also present the measured timing model parameters, pulsar dispersion measure versus time, and high-signal to noise ratio pulse profile shapes for all the sources.

In Chapter 5, we develop the theory behind the detection of gravitational radiation. We also present a new method for taking the effect of the pulsar timing model effect on the GW signal into account. The problem of the effect of the timing model and the related question of how to optimally detect the GW signal have been topics of much debate over the years. Our method differs from most previous approaches in that it does not rely at all on Monte Carlo simulation. Given an expected GW power spectrum, it also provides a optimal detection method. We apply this new methodology to the timing residuals of Chapter 4, and derive limits on the strength of the stochastic gravitational wave background near frequencies of 0.4 yr^{-1}. In doing so, we present both a single-pulsar analysis as well as combining all our data into a pulsar timing array. These limits are less restrictive than previous work due to the short time span of our data. However, this presentation lays the groundwork for future analysis of longer datasets. In particular, this is the first published GW limit that applies the timing array method to actual data in over 20 years.

Finally, Chapter 6 presents a detailed study of the original millisecond pulsar, B1937+21, using 20 years of data from several radio telescopes. In particular we study the effect of the interstellar medium on the pulsar signal, and draw conclusions about both the ISM itself as well as how it can negatively affect efforts to measure GW.

Chapter 2

Methods for Determining Pulse Times of Arrival

2.1 Introduction

The first step in data analysis for pulsar timing is the generation of pulse times of arrival (TOAs). This is essentially a timestamp stating precisely at what point during a given observation the pulsar's rotational phase was seen to be "zero." The definition of phase zero has no physical meaning, and the only requirement is that it be consistently defined within a single dataset. These TOAs will eventually be fit to a parameterized model of the pulsar's rotation and motion (timing model; see Chapter 4), and the residuals from this fit can be further analyzed for various purposes (Chapter 5).

The data we receive from a pulsar data recording system (for example, see Chapter 3) consists of a series of "folded" profiles as a function of time and radio frequency. These represent the incident radio wave's intensity as a function of pulse phase, and are recorded by averaging the signal power versus time modulo the pulsar's current apparent period, typically for several minutes. This requires a knowledge of an approximately correct timing model, which exist for all known pulsars.[1] Here we will refer to these measured folded profiles as $d(\phi)$. The procedure of measuring TOAs from these profiles involves the following two steps: Determining a "template profile" $p(\phi)$, and calculating the phase shift between the template and each measured profile. This procedure works (in the sense of producing physically useful results) because average pulse profile shapes are observed to be very consistent, and the pulse arrival times accurately represent the rotational phase of the neutron star itself.

In this chapter, we review the standard techniques for calculating TOAs from a new perspective, and show a connection between two traditional approaches. We then turn to the question of how to determine a template profile, and present an approach which integrates TOA measurement and template generation into a single process. Finally, we present a principal components-based procedure for characterizing profile shape variation in a dataset, and potentially correcting the resulting TOAs. These methods will be applied to a set of 16 millisecond pulsars in Chapter 4.

[1] Searching for new pulsars requires a entirely different set of procedures, which are not discussed here.

2.2 Basic TOA Fit Procedure

Historically, two classes of procedure have been used to determine the phase shift between the measured data and a template. One is to compute a discrete-time cross-correlation between the two functions, then interpolate this near the maximum point to get time resolution smaller than the sampling time. A typical interpolation scheme is to fit a parabola to the points near the maximum, and then to solve for the parabola's peak. This approach is known to suffer from systematic errors in which TOAs are "pulled" towards time bin edges.

The second method, put forward by Taylor (1992) is a χ^2 fit in the frequency domain. This is based on the insight that in the frequency domain, time-shifting becomes multiplication by a complex exponential, a functional form which can be easily dealt with in a standard χ^2 fit procedure. In this section, we will review the frequency domain method and show that it is in fact equivalent to the cross-correlation method with a different choice of interpolation.

The phase shift between $d(\phi)$ and $p(\phi)$ can be found by minimizing the function

$$\chi^2(a, \phi) = \sum_{k=1}^{k_{max}} \frac{|d_k - a p_k e^{-2\pi i k \phi}|^2}{\sigma_k^2} \tag{2.1}$$

Here $d_k = \sum_j d(j/N) e^{-2\pi i j k/N}$ is the discrete Fourier transform (DFT) of $d(\phi)$ (similarly for $p(\phi)$ and p_k), a and ϕ are the fit parameters (amplitude and phase shift), and σ_k^2 is the noise power in each frequency bin of the DFT. Under the assumption of additive white noise, σ_k will be constant, and its value will not affect the fitted parameters. For clarity, we will define $\Delta^2 = \sigma^2\chi^2$ and minimize this function rather than χ^2.

The absolute square can be multiplied out, and the formula written in the following way:

$$\begin{aligned}\Delta^2(a,\phi) &= \sum_k |d_k|^2 + a^2 \sum_k |p_k|^2 - 2a\Re \sum_k d_k p_k^* e^{2\pi i k\phi} \\ &= D^2 + a^2P^2 - 2aC_{dp}(\phi)\end{aligned} \tag{2.2}$$

where $D^2 = \sum|d_k|^2$ and $P^2 = \sum|p_k|^2$ are the total power in each signal. We can see that all the phase information is contained in the function $C_{dp}(\phi) = \Re\sum_k d_k p_k^* e^{2\pi i k\phi}$, so maximizing C_{dp} will minimize Δ^2 and give us the best-fit $\phi = \hat{\phi}$, for a given a. We can also see that when ϕ takes on the discrete values $\phi_j = j/N$, $C_{dp}(\phi_j)$ is the inverse DFT of $d_k p_k^*$, or equivalently the discrete cross-correlation function (CCF) of $d(\phi_j)$ and $p(\phi_j)$. So by using this method, we are essentially doing the same thing as in time-domain methods, except rather than interpolating the CCF's peak with a quadratic fit, we are interpolating by setting all Fourier components past some maximum frequency k_{max} to zero. In principle k_{max} is set by the number of time bins in the data. However, in practice it is better to truncate the sums sooner, based on the signal-to-noise ratio of the data. We will return to the question of how to do that in Section 2.4.

The maximum $C_{dp}(\hat{\phi})$ can easily be found via standard 1-D maximization algorithms (Press et al., 1992). In this work, we use a golden section search on C_{dp} directly, for simplicity. Another approach is to use Brent's method to solve $C'_{dp}(\hat{\phi}) = 0$ (Taylor, 1992). Finding the fitted amplitude $\hat{a}$ can be done trivially by solving $\frac{\partial}{\partial a}\Delta^2(a,\hat{\phi}) = 0$, with the result:

$$\hat{a} = \frac{C_{dp}(\hat{\phi})}{P^2} \tag{2.3}$$

The errors in the fitted parameters can also be determined by standard procedure:

$$\begin{aligned} \sigma_\phi^2 &= \left(\frac{\partial^2\chi^2}{\partial\phi^2}\right)^{-1} = \frac{\sigma^2}{-2\hat{a}C''_{dp}(\hat{\phi})} \\ \sigma_a^2 &= \left(\frac{\partial^2\chi^2}{\partial a^2}\right)^{-1} = \frac{\sigma^2}{2P^2} \end{aligned} \tag{2.4}$$

Here, note that $C''_{dp}(\hat{\phi})$ should always be negative, since it is evaluated at a maximum of C_{dp}. Since we don't generally have an independent estimate of the value of σ, we can estimate it from the fit quality by assuming $\bar{\chi}^2 \equiv \chi^2/(2k_{max}-2) = 1$ and solving for σ.

2.3 Effect of Profile Shape Errors on Timing

A assumption implicit in the above analysis is that the data $d(\phi)$ is actually well described by a scaled, time-shifted template $a_0 p(\phi - \phi_0)$. Setting aside for the moment the question of how one determines $p(\phi)$, it is useful to consider a difference in shape between the two functions, and see what effect that has on the fitted parameters a and ϕ. In this case, the data can be described by:

$$d_k = a_0\left(p_k + q_k\right)e^{-2\pi i k\phi_0} \tag{2.5}$$

Here, q represents the shape difference between d and p, and is taken to be small, i.e. $\sum|q_k|^2 \ll \sum|p_k|^2$.

We can perform the fit as above, maximizing the function $C_{dp}(\phi)$. The CCF C_{dp} can be expanded as

$$C_{dp}(\phi) = a_0 C_{pp}(\phi - \phi_0) + a_0 C_{qp}(\phi - \phi_0) \tag{2.6}$$

Since the shape difference q_k is assumed to be small in magnitude it is reasonable to assume that the fitted time-shift $\hat{\phi}$ will be close to the true time-shift ϕ_0, differing by a small amount $\delta = \hat{\phi} - \phi_0$. Then the CCFs in equation 2.6 can be expanded in a Taylor series about $\phi = \phi_0$ and we can solve for δ by setting $C'_{dp}(\delta) = 0$. This results in the leading order time shift

$$\delta = \frac{C'_{qp}(0)}{C''_{pp}(0)} \tag{2.7}$$

Since we don't usually have any prior knowledge of the shape of q, a logical next step would be to estimate q by subtracting off the scaled, shifted p from d. Then we could use this estimate in Equation 2.7 to get a corrected TOA. However, this process doesn't work. Consider creation of the residual profile $r(\phi - \hat{\phi}) = d(\phi) - \hat{a}p(\phi - \hat{\phi})$. Then the CCF's derivative is given by

$$C'_{rp}(\phi) = C'_{dp}(\phi + \hat{\phi}) - \hat{a}C'_{pp}(\phi) \tag{2.8}$$

But $C'_{dp}(\hat{\phi}) = 0$, since that is how we found $\hat{\phi}$ in the first place. And $C'_{pp}(0) = 0$ as well, due to its symmetry. This means that $C'_{rp}(0) = 0$ always, regardless of the values of d and p. We can't determine a timing correction δ based solely on the measured profile shape. However, there are ways to get around this: We can make a educated guess at a form for q, for example a exponential scattering tail (Ramchandran et al.,

2006) or a variable height Gaussian (Kramer et al., 1999). Another approach is to look for empirical correlations between the residual profiles and the measured timing residuals in a given dataset. We will explore this latter option in §2.5.

2.4 Template Determination

Application of the algorithms described in Section 2.2, requires an estimate of the template function $p(\phi)$. The ideal template would be an infinite-SNR version of the pulse profile. This performs as a matched filter for the data and gives a CCF with the best possible SNR. Of course, this is difficult in practice since at some level the template must be estimated from the data. Again there have been two typical methods for doing this. One is to fit a analytic function (usually a sum of Gaussians) to a high-SNR measured profile (Lommen, 2001). This has the benefit of producing a noise-free template, but it is generally not a fully accurate representation of the pulse shape. The second method is to take a high-SNR measured profile and simply use that as the template. While the latter approach gives a more accurate pulse shape, it produces a template which is contaminated with noise. Also, both methods are somewhat inelegant in that they require human intervention in the original selection of data to use, and in the Gaussian fitting process.

Here, we propose a method which uses all the available data to determine a template. We can consider the template components p_k as additional free parameters in Equation 2.1, and sum over all M measured profiles to create a global χ^2 for the whole data set:

$$\chi^2(a_1, \ldots, a_M, \phi_1, \ldots, \phi_M, p_1, \ldots, p_N) = \sum_{l,k} \frac{|d_{lk} - a_l p_k e^{-2\pi i k \phi_l}|^2}{\sigma_l^2} \tag{2.9}$$

Since a multi-year data set can potentially have $M \sim 10^4$ profiles, this is too large a problem to approach with usual χ^2 techniques such as solution of the normal equations. Instead we will use a iterative scheme which proceeds as follows:

1. Estimate a initial set of parameters a_l, ϕ_l.
2. Solve for updated p_k while holding a_l and ϕ_l fixed.
3. Normalize and filter p_k for optimal SNR.
4. Solve for updated a_l and ϕ_l while holding p_k fixed. Estimate σ_l from the fit quality.
5. Return to step 2 unless the process has converged.

The initial estimate of parameters (Step 1 above) can be done by taking the amplitude and phase of the first Fourier component of each measured profile.

Step 2 can be done by setting $\frac{\partial}{\partial(\Re p_k)}\chi^2 = \frac{\partial}{\partial(\Im p_k)}\chi^2 = 0$. This results in the following solution for $\hat{p}_k$:

$$\hat{p}_k = \left(\sum_l \frac{\hat{a}_l^2}{\sigma_l^2} \right)^{-1} \sum_l \frac{\hat{a}_l d_{lk} e^{2\pi i k \hat{\phi}_l}}{\sigma_l^2} \tag{2.10}$$

That is, $\hat{p}_k$ is simply a SNR-weighted sum of the aligned data, a result which makes good sense intuitively.

Step 4 can then be accomplished independently for each measured profile as described in §2.2. When run on actual data, this process usually converges in 3 or 4 iterations.

Pulse profile power spectra $|p_k|^2$ typically decline versus frequency at large values of k. This means that (assuming enough time resolution) $\hat{p}_k$ will be dominated by noise past some limiting k value. It is actually counterproductive to include these points in the fits for a_l and ϕ_l since they add no information and can only degrade the fit quality. So after each updated $\hat{p}_k$ is found, we apply a low-pass filter to remove the noise-dominated points. The optimal procedure would be to apply a Wiener filter (see Press et al., 1992, §13.3). Given a noise-free profile p_k, a noisy profile $\hat{p}_k$ and a estimate of the noise power in each bin σ_k, the Wiener filter

$$f_k^{(W)} = \frac{|p_k|^2}{|p_k|^2 + \sigma_k^2} \simeq \frac{|\hat{p}_k|^2 - \sigma_k^2}{|\hat{p}_k|^2} \tag{2.11}$$

gives the maximum SNR output when applied to $\hat{p}_k$. This filter has the property of being ~ 1 for high-SNR points, and ~ 0 for noise dominated points. Since for pulsar data the transition between these two is usually fairly sharp, we can apply a simple "brick wall" filter as a good approximation:

$$f_k^{(B)} = \begin{cases} 1, & k \leq k_c \\ 0, & k > k_c \end{cases} \tag{2.12}$$

The noise level is estimated by taking the mean value of the last quarter of the power spectrum. The cutoff frequency k_c can be determined by minimizing $\sum (f_k^{(W)} - f_k^{(B)})^2$

versus k_c. This approach is more robust than simply taking k_c as the smallest k where $f_k^{(W)} < \frac{1}{2}$, a common alternate filtering method.[2] The brick wall filter is easily implemented by truncating the sums in Equations 2.1 and 2.2 at $k_{max} = k_c$. A example of a resulting template profile for PSR J1713+0747 is shown in Figure 2.1. Also shown is the power spectrum of this profile, and the results of the noise filtering process. In this case, we determined $k_c = 227$.

2.5 Principal Components Analysis

In §2.3 we discussed the effect of a mismatch between the data and template functions. We found a systematic shift in the measured TOAs (Equation 2.7). This is not necessarily a huge problem for the timing process - as long as the data have a consistent shape, δ will be constant, and will be absorbed into the timing model as an arbitrary offset. However, a variable profile shape will cause systematic shifts that can affect measured timing model parameters and corrupt any post-timing analysis such as gravitational wave detection.

Several pulsars are known to show intrinsic profile shape variations; for example the millisecond pulsars PSR B1821-24 (Backer and Sallmen, 1997) and J1022+1001 (Kramer et al., 1999; Ramachandran and Kramer, 2003). Other MSPs may have similar variations of smaller amplitude. Additionally, there are a wide range of instrumental effects that can cause variation in the measured profiles (see Chapter 3).

[2] This is done for example in Taylor's *fftfit* program.

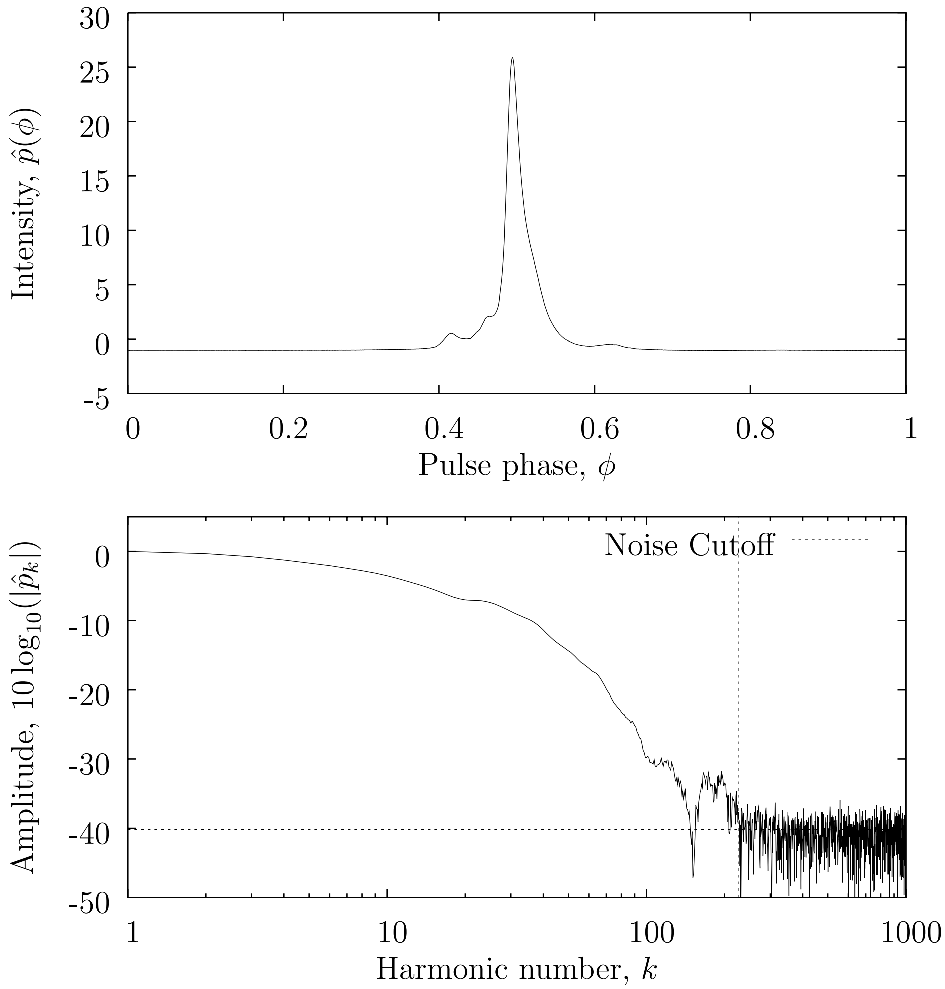

Figure 2.1: Template profile and power spectrum for PSR J1713+0747 at 1400 MHz.

Therefore, we would like to develop a method of characterizing profile shape variation in a dataset.

A useful method for detecting variations in large datasets is principal component analysis (PCA).[3] PCA characterizes variation in a dataset by determining a set of orthonormal basis vectors along which the variance is maximized. Additionally, when the data are expressed in this new basis, the covariance between different components is zero. Mathematically, this is accomplished by solving for the eigenvalues and eigenvectors of the covariance matrix of the dataset. Assume we have a set of M data points, $\mathbf{x}_j$, each of which is a N dimensional (column) vector, and has a statistical weight w_j. Compute the weighted mean of the dataset as:

$$\bar{\mathbf{x}} = \left(\sum_{j=1}^{M} w_j\right)^{-1} \sum_{j=1}^{M} w_j \mathbf{x}_j \tag{2.13}$$

Then the weighted covariance matrix is given by:

$$\mathbf{C}_{\mathbf{x}} = \left(\sum_{j=1}^{M} w_j\right)^{-1} \sum_{j=1}^{M} w_j (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \tag{2.14}$$

PCA determines a orthogonal transformation $\mathbf{M}^T = \mathbf{M}^{-1}$ which transforms the vectors $\mathbf{x}_j$ to $\mathbf{z}_j = \mathbf{M}^T \mathbf{x}_j$ whose covariance matrix $\mathbf{C}_{\mathbf{z}}$ is diagonal. The columns of $\mathbf{M}$ (eigenvectors of $\mathbf{C}_{\mathbf{x}}$), $\mathbf{m}_j$, are the principal components. The corresponding diagonal element of $\mathbf{C}_{\mathbf{z}}$ (eigenvalue), λ_j, gives the variance in the dataset along the $\mathbf{m}_j$ direction. Since any covariance matrix is positive semi-definite, $\lambda_j \geq 0$ for all j.

[3]The same procedure goes by several other names as well, including empirical orthogonal functions, the Hotelling transform, and the Karhunen-Loeve transform. It is described in many statistical texts, for example Hyvarinen et al. (2001).

They are typically ordered so that $\lambda_1 > \lambda_2 > \cdots > \lambda_N$. It is easy to see that this transformation preserves the total variance in the data:

$$\sigma_{tot}^2 = \mathrm{Tr}(\mathbf{C_x}) = \mathrm{Tr}(\mathbf{M C_z M}^T) = \mathrm{Tr}(\mathbf{M}^T \mathbf{M C_z}) = \mathrm{Tr}(\mathbf{C_z}) = \sum_{j=1}^{N} \lambda_j \tag{2.15}$$

Therefore we can say that the principal component $\mathbf{m}_j$ is responsible for a fraction $\lambda_j / \sigma_{tot}^2$ of the total variance.

The application of PCA to pulsar data has been done only a few times before. Blaskiewicz (1991) used PCA on a large set of slow pulsar observations to identify and characterize intrinsic changes in profile shape. With pulsar data, there are a few preliminary steps to perform before applying the PCA algorithm described above. For the most part we follow the same procedure as Blaskiewicz (1991), which will be briefly reviewed here.

First, it is useful to note that we can apply an arbitrary orthogonal transformation $\mathbf{F}^T = \mathbf{F}^{-1}$ to the data without changing the basic conclusions of the PCA. If $\mathbf{x}' = \mathbf{F}^T \mathbf{x}$, then

$$\mathbf{C_z} = \mathbf{M}^T \mathbf{C_x M} = \mathbf{M}^T (\mathbf{F C_{x'} F}^T) \mathbf{M} = \mathbf{M}'^T \mathbf{C_{x'} M}' \tag{2.16}$$

The eigenvalues have remained the same, and we have a new set of eigenvectors, $\mathbf{m}'_j = \mathbf{F}^T \mathbf{m}_j$ which are simply transformed versions of the original (non-primed) ones. Since the Fourier transform[4] is orthogonal, this means we can use PCA directly

[4] In this case it is best to use a "real-to-real" DFT to avoid having to solve a complex-Hermitian eigenvalue problem. The real-to-real DFT is defined by:

$$\sqrt{N} F_{jk} = \begin{cases} \cos(\pi jk/N), & \text{even } j \\ \sin(\pi(j+1)k/N), & \text{odd } j \end{cases}$$

on the frequency domain profile data.

The data is aligned and scaled using the previously determined timing parameters $\hat{a}_l$ and $\hat{\phi}_l$. The template profile is then subtracted off. This leaves us with a set of residual profiles r_{lk} which only have variation orthogonal to the template. The weights are determined from the SNR of the data.

$$r_{lk} = \frac{d_{lk} e^{2\pi i k \hat{\phi}_l}}{\hat{a}_l} - \hat{p}_k \tag{2.17}$$

$$w_l = \frac{\hat{a}_l^2}{\sigma_l^2} \tag{2.18}$$

The dimension of the problem is then reduced by only keeping the first k_c harmonics of each profile. This greatly reduces the computation involved in diagonalizing the covariance matrix, which typically scales as $O(n^3)$. It also reduces noise in the measured principal components, and keeps the analysis in the same vector space as the original timing fit.

After running the PCA algorithm on this modified dataset, we have a set of $2k_c$ eigenvectors (m_{jk}) and corresponding eigenvalues (λ_j). We must then examine the λ_j and decide the number of components which are statistically significant (N_{sig}). This is a fairly complex problem, and simple approaches typically involve a subjectively-set threshold: For example, we could estimate the noise level from the mean of the $N/2$ smallest eigenvalues, and then keep components which have eigenvalues larger than some multiple of this value. Wax and Kailath (1985) develop a information-theoretic

---

If a standard complex DFT has already been performed, it is easy to reorganize the real and imaginary parts into real-to-real format.

algorithm for finding N_{sig}, based on minimizing the following function:

$$\mathrm{MDL}(j) = -M_e(2k_c - j)\log\left(\frac{\prod_{i=j+1}^{2k_c}\lambda_i^{1/(2k_c-j)}}{\frac{1}{2k_c-j}\sum_{i=j+1}^{2k_c}\lambda_i}\right) + \frac{1}{4}j\,(4k_c - j + 1)\log M_e \quad (2.19)$$

The first term in this function is a measure of how well the first j components reproduce the data. The second term represents the number of degrees of freedom used by the first j components. The minimum occurs at $j = N_{sig}$ where these two terms balance each other. Since we used a weighted covariance estimate, this function depends on the effective number of samples, M_e (rather than M, the actual number of data points):

$$M_e = \frac{\left(\sum_{i=1}^{M} w_i\right)^2}{\sum_{i=1}^{M} w_i^2} \quad (2.20)$$

M_e ranges from 1, if all the weight is in one sample, to M if the weights are all equal.

Once we have decided on a value for N_{sig} we can project the data onto the first N_{sig} components to get a set of amplitudes, b_j for each profile:

$$b_{lj} = \Re\sum_k r_{lk}m_{lk}^*, \quad 1 < j < N_{sig} \quad (2.21)$$

We will use these amplitudes to determine timing corrections. However, since the basis vectors m_{jk} were created in the vector space spanned by r_{lk}, they suffer from the limitation described in §2.3: $C'_{mp}(0) = 0$, so we can't simply apply Equation 2.7. Instead, what we have accomplished is a further reduction in the dimension of the problem, from $2k_c$ down to N_{sig}. We can now look for a linear combination of the b_j which correlates with the measured timing residuals, then use this to correct the TOAs.

Component	Height	Width	Location
1	1.0	0.03	0.25
2	0.15−0.25	0.21	0.40

Table 2.1: Gaussian profile parameters of the simulated data.

2.6 Application to Simulated Data

In this section, we will apply the methods developed throughout this chapter to simulated profile data to explore how well they perform. We will apply the methods to actual pulsar data in Chapter 4. A trial with simulated data lets us test in a controlled way how our algorithms will respond to different situations.

We generated 1000 pulse profiles based on Gaussian parameters listed in Table 2.1. The second component's amplitude was varied randomly over the range listed. Each profile was given a phase shift that increased linearly throughout the dataset, $\Delta\phi_l = l/1000$. They were then integrated into 2048 phase bins, and Gaussian noise with a RMS value of 0.1 was added. This data was run through the iterative timing algorithm described in §2.4. The resulting template is shown in Figure 2.2. The noise cutoff value was determined to be at $k_c = 32$, but we have plotted the unfiltered template.

We then applied the PCA methods discussed in §2.5. This gave an eigenvalue spectrum (Figure 2.3) of which only the first component is significant. As expected, the last two points of the spectrum (not shown on this graph), are approximately 0 owing to the two degrees of freedom that were removed by fitting for the amplitude and phase shift of each profile. The first principal component of the residual profile, $m_1(\phi)$

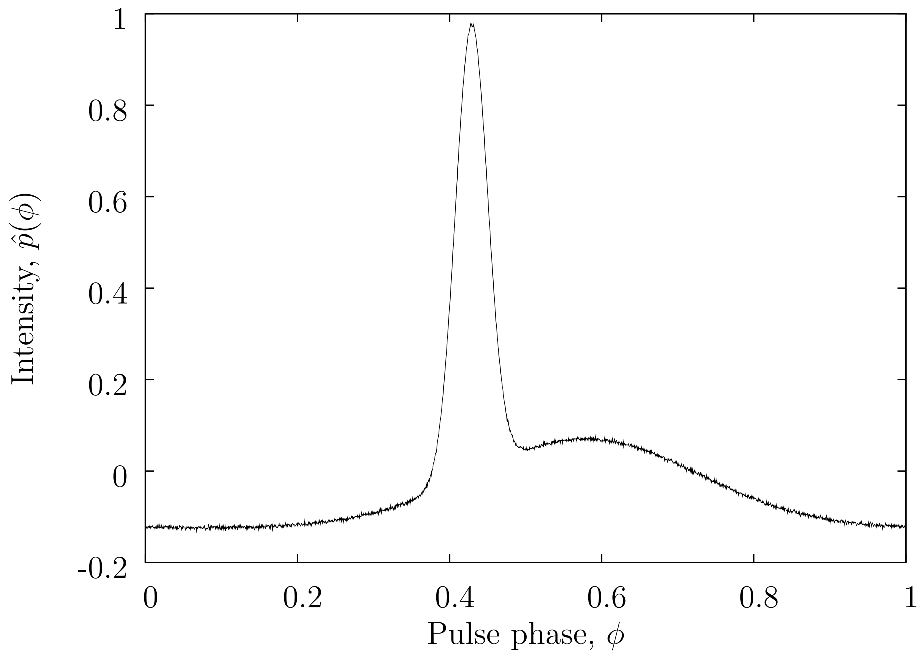

Figure 2.2: Template profile determined from the simulated dataset.

is shown in Figure 2.4. Although we only varied the second Gaussian component, this function shows features of both Gaussian components (with opposite sign). This is another consequence of the template fitting: We are only able to detect the part of the variation which is orthogonal to both $\hat{p}(\phi)$ and $\hat{p}'(\phi)$. However, even with this limitation, we are still getting a good measure of the original shape variation. This is shown in Figure 2.5, where the measured principal component amplitude (projection of $d_l(\phi)$ onto $m_1(\phi)$) is plotted versus the true Gaussian component's height for the whole dataset. The two quantities clearly have a simple linear relationship.

The process of fitting for a timing model was done by fitting the measured TOAs to a linear function of their index. This model was then subtracted off to give timing

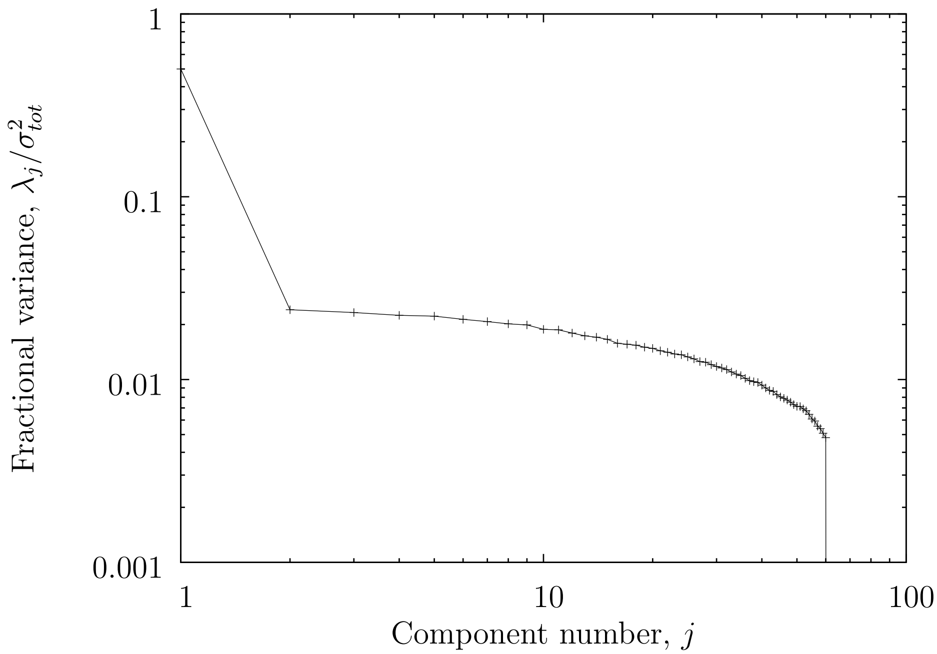

Figure 2.3: Principal component eigenvalues from the simulated data.

residuals. These residuals had a RMS value of 2.72×10^{-3}. Based on Equation 2.4, the true pulse profile and known noise level, we expect $\sigma_\phi = 2.31 \times 10^{-3}$, so we have created a systematic timing error. Figure 2.6 shows timing residual versus the measured principal component value. When we fit a line to this plot and apply the resulting correction to the TOAs, the measured RMS timing residual drops to 2.34×10^{-3}, very close to the expected value.

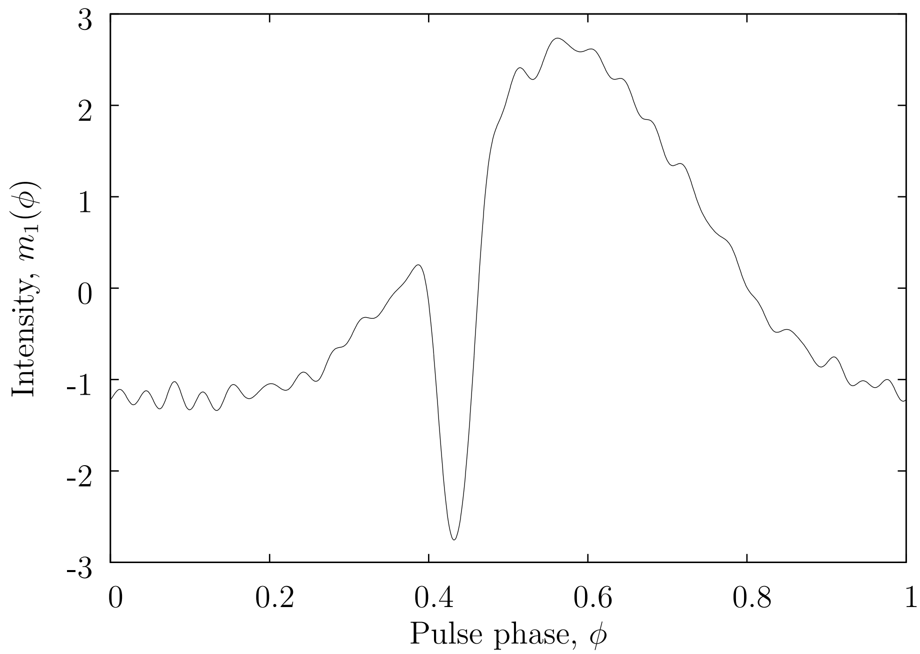

Figure 2.4: First principal component of the simulated data.

2.7 Polarization Calibration

A final effect that can change both profile shape and measured TOAs in a systematic way is incorrect polarization calibration. Pulsar emission is usually strongly linearly polarized, sometimes with a small amount of circular polarization as well. The direction and polarized fraction change systematically as a function of pulse phase. This is quantified by the 4 Stokes parameters (see Jackson, 1999, §7.2). Instead of the single function $d(\phi)$, representing intensity versus phase, we have the functions $d_j(\phi)$, where d_0 is total intensity (I in usual notation), d_1 and d_2 represent linearly polarized power (Q and U), and d_3 represents circular polarization (V). While these

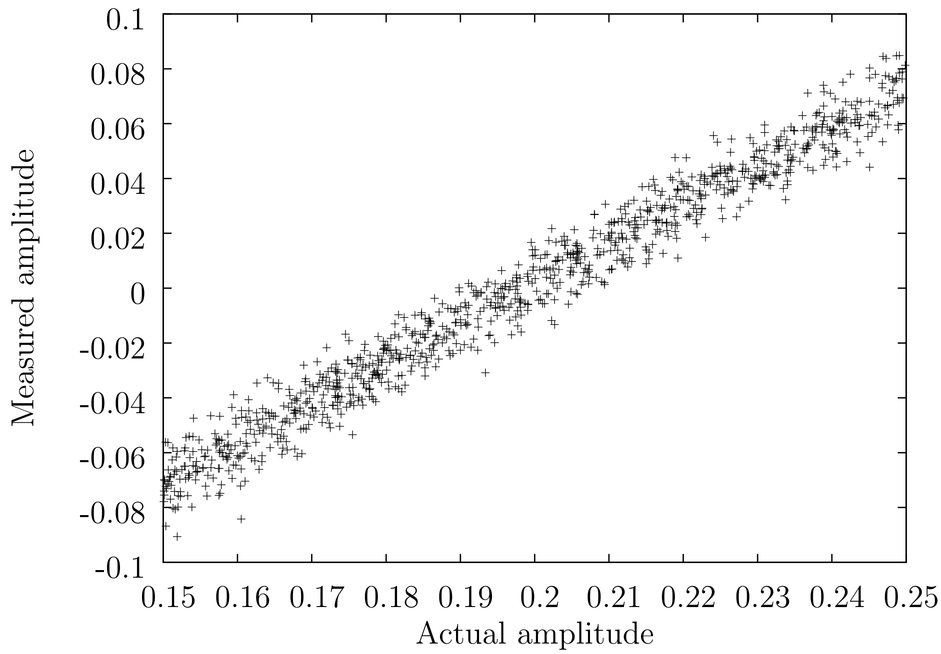

Figure 2.5: Measured principal component value versus true Gaussian amplitude.

quantities all have units of intensity (or power), only d_0 is positive definite. The other three parameters can be either positive or negative, depending on the polarization direction. Since a signal can be at most 100% polarized, the Stokes parameters also must satisfy the inequality $d_0^2 \geq d_1^2 + d_2^2 + d_3^2$.

These extra degrees of freedom also introduce extra potential for systematic error. A variety of instrumentation details can alter the polarization content of a signal, including (but not limited to) the following: Different amplifier gains for each polarization, slightly non-orthogonal receiver feeds, different cable lengths for each signal, and cross-talk between the two. The various parameters are reviewed in detail by Heiles et al. (2001). The net effect of all these parameters, as long as the system is

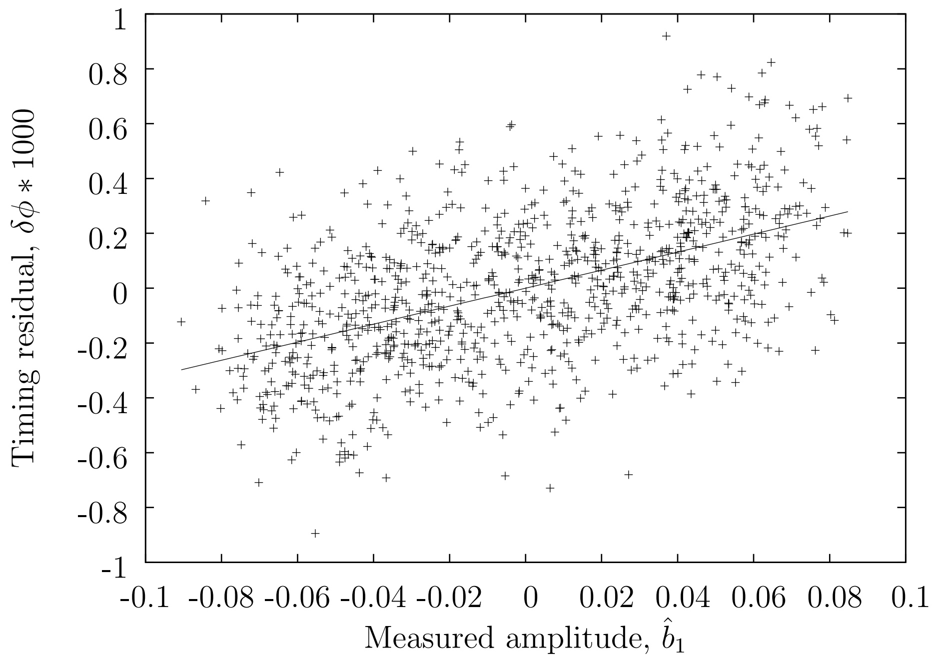

Figure 2.6: Timing residual versus measured component value.

linear, can be summed up in the following relation:

$$d_i^{(measured)} = \sum_j M_{ij} d_j^{(true)} \tag{2.22}$$

Here, M_{ij} is known as the Mueller matrix. It must be characterized and inverted in order to recover the true input signal.[5] For high-precision work, we must do this even if we only plan on working with the total intensity d_0 (as is the case here) - M_{ij} couples the polarized components into d_0, potentially altering the shape of $d_0(\phi)$. This shape change will vary as the telescope changes direction and if amplifiers or other system parameters are reset, and can thus lead to systematic TOA errors.

[5] A particularly interesting discussion by Britton (2000) reveals that Mueller matrices are in fact members of the Lorentz group.

The standard method of correcting these problems is by observing a calibrator source with well-known polarization behavior, and fitting this data for the elements of M_{ij}. Here we have developed a self-calibration method to use for pulsar timing. We can fit for the relevant polarization parameters at the same time as timing parameters, by minimizing Δ^2 as discussed in §2.2. Except we now include the additional parameters g_i as follows:

$$\Delta^2(g_i, \phi) = \sum_k \frac{|g_i d_{ik} - p_k e^{-2\pi i k \phi}|^2}{g_i g_i} \tag{2.23}$$

The g_i represent the gain that needs to be applied to each measured polarization component in order to make the profile shape more accurately match the template. These can be found in the first row of M^{-1} above. We have also absorbed the earlier parameter a into the definition of g_i, with $a = (g_i g_i)^{-1/2}$.

The effect of applying this method is shown in Figure 2.7. Here we show timing residuals calculated from observations of PSR B1937+21 taken with the NRAO Green Bank Telescope. The top two panels show residual versus radio frequency for data taken at widely different hour angles. This rotates the telescope feed with respect to the source, changing the effective Mueller matrix. The clear dependence of this effect on hour angle identifies polarization as the root source of the error. The bottom panel shows the same data after applying the polarization correction described in Equation 2.23. This has eliminated most, if not all, of the problem.

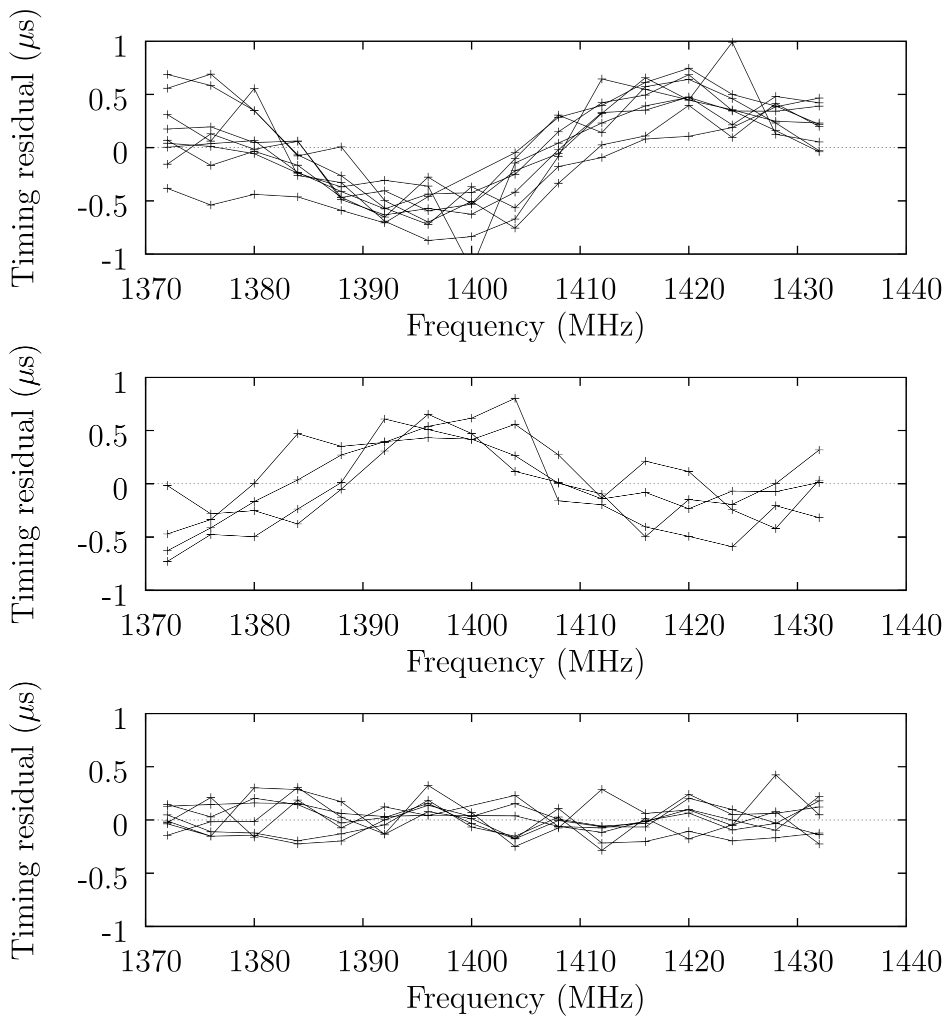

Figure 2.7: Timing residual versus frequency for PSR B1937+21. The top and middle panels show data taken at positive and negative hour angles, respectively, with no polarization calibration done. The bottom panel shows the same data after applying the method described here.

Chapter 3

New Instrumentation for Precision Pulsar Timing: The Astronomy Signal Processors

3.1 Introduction

This chapter describes a new set of coherent dedispersion pulsar backends, known as the Astronomy Signal Processors (ASPs), recently developed by our group. Three systems are currently in use, at Arecibo Observatory in Puerto Rico (ASP), the NRAO Green Bank Telescope in West Virginia (GASP), and at the Nançay Radio Telescope in France (LBP[1]). The ASP systems differ from most previous coherent systems (for

[1] LBP = Le Berkeley-Orléans-Nançay Processor

example, see §4.2) in that they employ a minimal amount of specialized hardware. Instead, most of the signal processing is done by software running in a "Beowulf" cluster of personal computers (PCs). Here we will describe the design of the ASP systems, as well as the motivations for coherent dedispersion. Where appropriate, we we also note the impact of backend design choices on precision timing results, as several types of systematic error come directly from these decisions.

3.2 Coherent Dedispersion

Unlike many other astronomical phenomena, pulsar signals show extreme variation versus time on a number of time scales, the most fundamental being the spin period of the neutron star (pulse period). This fact leads to several unique observational consequences, one of which is the *dispersion* of the observed signal at Earth. Here we define dispersion, show how it arises, and outline the preferred method of correcting for it, coherent dedispersion. This material is covered in detail in standard texts (e.g. Jackson, 1999), and the pulsar-specific case was reviewed by Hankins and Rickett (1975).

The interstellar medium (ISM) consists of a mixture of gas and dust, almost all of which is completely transparent at centimeter wavelengths. Aside from absorption due to a few atomic and molecular spectral lines (most notably the atomic hydrogen hyperfine transition at 1420 MHz), the main effect of the ISM on radio waves comes from the ionized component. The behavior of radio waves propagating through this medium is described by the tenuous (collisionless) plasma dispersion relation:

$$\omega^2 = c^2k^2 + \omega_p^2 \tag{3.1}$$

The plasma frequency ω_p is defined as

$$\omega_p^2 = \frac{4\pi e^2 n_e}{m_e} \tag{3.2}$$

and is in general a function of position and time due to its dependence on the free electron density $n_e(\mathbf{x}, t)$. Typical values for n_e in our galaxy range from 0.01 to 1 cm^{-3}, resulting in ω_p of 5 to 50 kHz. These are much smaller than the typical pulsar observation frequencies of 100 MHz to 10 GHz, so approximations based on $\omega \gg \omega_p$ can be applied.

It follows directly from Equation 3.1 that the group velocity of radio waves travelling through the ISM is given by

$$v_g = \frac{\partial\omega}{\partial k} = c\left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2} \simeq c\left(1 - \frac{\omega_p^2}{2\omega^2}\right) \tag{3.3}$$

with the final approximation being valid in the $\omega \gg \omega_p$ regime. A pulsed signal (wavepacket) travelling a distance d through the medium has a time of flight T given by

$$\begin{aligned} T = \int_0^d \frac{1}{v_g}dx &\simeq \frac{d}{c} + \frac{1}{2\omega^2 c}\int_0^d \omega_p^2(x)dx \\ &= \frac{d}{c} + \frac{1}{\omega^2}\frac{4\pi e^2}{2cm_e}\int_0^d n_e(x)dx = \frac{d}{c} + \frac{1}{\omega^2}\frac{2\pi e^2}{cm_e}\mathrm{DM} \end{aligned} \tag{3.4}$$

This expression is good as long as the variation of n_e over one wavelength is negligible. We have defined the dispersion measure (DM) as $\mathrm{DM} \equiv \int n_e(x)dx$. This has units of

l^{-2} and physically represents the column density of free electrons between Earth and the pulsar. It is usually expressed in "mixed units" of pc cm^{-3}, with typical values for sources in our galaxy of 1 to 1000 pc cm^{-3}. The second term in Equation 3.4 shows that higher frequency components of the signal will arrive at their destination earlier. The wavepacket is also dispersed, or spread in time, by an amount δT:

$$\begin{aligned}\delta T &= \left|\frac{\partial T}{\partial \omega}\right| \Delta\omega \simeq \frac{\Delta\omega}{\omega^3}\frac{\pi e^2}{c m_e}\mathrm{DM} \\ &\approx 8.3\ \mu\mathrm{s}\frac{\Delta\nu(\mathrm{MHz})}{\nu(\mathrm{GHz})^3}\mathrm{DM}(\mathrm{pc\ cm^{-3}})\end{aligned} \tag{3.5}$$

The spreading is proportional to the wavepacket's bandwidth, $\Delta\omega$. Since dispersion is a linear process, the bandwidth can be reduced at any point along the signal path without affecting the form of Equation 3.5. The appropriate $\Delta\omega$ to use is the most restrictive (or narrowest) of any such reductions. Pulsar emission is intrinsically wideband, so $\Delta\omega$ is almost always imposed by the receiving hardware.[2]

Without correcting for the dispersion somehow, observing a pulsar over a wide bandwidth rapidly becomes impossible. Once δT is comparable to the pulse period, the source is not even observed to pulse! There are two methods for correcting this problem: coherent and incoherent dedispersion. Incoherent dedispersion is the simpler of the two. In this method, the incoming signal is split via a filter bank into subbands of width $\Delta\omega$. The signal is then detected in each subband, and the DM can be measured by fitting the pulse times of arrival (TOAs) versus ω to Equation 3.4. The signals can then be aligned in time using the measured DM, and added together

[2]The diffraction of the signal by the ISM, which is discussed below and in Chapter 6 is also a factor here.

to increase the signal-to-noise ratio. This method is computationally efficient, but has an intrinsic limitation: As $\Delta\omega$ is decreased, δT also decreases proportionally. But due to the uncertainty principle, the time resolution in the subband increases as $1/\Delta\omega$. At some $\Delta\omega$, these two effects will balance, resulting in a limit on the achievable time resolution, and therefore on the TOA accuracy possible. Furthermore, since the dispersion within each subband is not corrected, any variation in the signal strength versus ω occurring on scales smaller than $\Delta\omega$ (for example, due to scintillation; see Chapter 6) will shift the measured TOAs accordingly. This systematic effect is potentially much more damaging to precision measurements than the δT smearing alone.

A more careful investigation of dispersion reveals that it is possible to do better. Dispersion is a coherent effect: It alters the phase of the radio waves, but in a way that is systematic and predictable (assuming the DM is known). It is possible to represent the effect of dispersion in the frequency domain by a simple phase-only transfer function, $H(\omega) = e^{i\phi(\omega)}$. Applying the inverse filter, $H^*(\omega)$ to the signal prior to detection completely removes the effect of dispersion. Hankins and Rickett (1975) give the following form for ϕ:

$$\phi(\nu;\nu_0) = \frac{2\pi D\nu^2}{\nu_0^2(\nu_0+\nu)} \tag{3.6}$$

This expression is meant to be applied to a band of width $\Delta\nu$ about a center frequency ν_0. ν gives the offset within the band, with $|\nu| < \Delta\nu/2 \ll \nu_0$. The dispersion coefficient D is related to the DM by $D = \frac{e^2}{4\pi c m_e}\mathrm{DM}$.[3] This filter has had the time shift described by Equation 3.4 removed for a frequency of ν_0, and so only accounts for the spreading of the wavepacket within the band. The Fourier transform of $H(\omega)$ gives $h(t)$, the impulse response of the dispersion filter. For the filter given by Equation 3.6, $h(t)$ is a chirped pulse of width δT, centered at $t = 0$. The dispersion process can equivalently be thought of as a convolution of the original signal with $h(t)$.

3.2.1 Implementation as a Digital Filter

Although creative analog approaches such as swept-frequency oscillators or dispersive filters have been used in the past, all modern coherent dedispersion systems work by applying the dedispersion filter digitally. This can be done either in the frequency domain by multiplying by $H^*(\omega)$, or in the time domain by convolving with $h(-t)$. The choice between these two methods depends mainly on the hardware implementation details. Time-domain methods are usually simpler to implement in specialized hardware, but are less computationally efficient. Frequency-domain methods can be much more efficient but are also more complex, as they use a Fourier transform.

Nyquist-sampling a complex digital signal of bandwidth $\Delta\nu$ requires a sampling rate of $1/\Delta\nu$, so the length of $h(t)$ in samples is:

$$N_d = \delta T \Delta\nu = \frac{(\Delta\nu)^2}{\nu^3} D \tag{3.7}$$

[3] Since the chirp rate D can be directly measured in pulsar experiments, measured DMs are by convention *defined* using the relation $\mathrm{DM}(\mathrm{pc\ cm^{-3}}) = 2.410000 \times 10^{-16} D(\mathrm{s\ Hz^2})$. This practice removes ambiguity regarding chosen values for the physical constants e, m_e, and c.

Time-domain dedispersion involves convolving the input signal with a finite impulse response (FIR) filter of length N_d. Performing the same operation in the frequency domain requires first Fourier transforming a block of N_f samples, then multiplying by the dedispersion filter $H^*(\nu)$, and finally inverse Fourier transforming to get back to the time domain. This method introduces a "edge effect" at the boundaries of the data block. If the filter given in Equation 3.6 is used, the first and last $N_d/2$ samples in each dedispersed block must be discarded, leaving $N_f - N_d$ valid output points per block. This is essentially because the Fourier transform method is equivalent to a circular convolution of the data with $h(-t)$. At the block edges, the filter "wraps" around and inappropriately combines data from the start and end. Clearly, in order for this method to work, N_f must be greater than N_d. To avoid gaps in the dedispersed time series, the Fourier transform blocks should be overlapped in time by N_d samples.

The computational cost of both methods can now be easily calculated. The direct time-domain convolution uses N_d complex multiplies and $N_d - 1$ complex adds per sample, so the cost scales as $o(N_d)$ The frequency-domain method gains efficiency when the transforms are computed via the fast Fourier transform algorithm (FFT; Cooley and Tukey, 1965). The FFT's computational cost scales as $o(N_f \log N_f)$, so the cost for frequency-domain dedispersion goes as $o\left(\frac{N_f}{N_f - N_d} \log N_f\right) + o\left(\frac{N_f}{N_f - N_d}\right)$ per sample. The best choice for N_f as a function of N_d depends on the details of the FFT implementation used, and potentially on hardware parameters such as CPU cache size, so optimization versus N_f is best done empirically. However, a good rule of

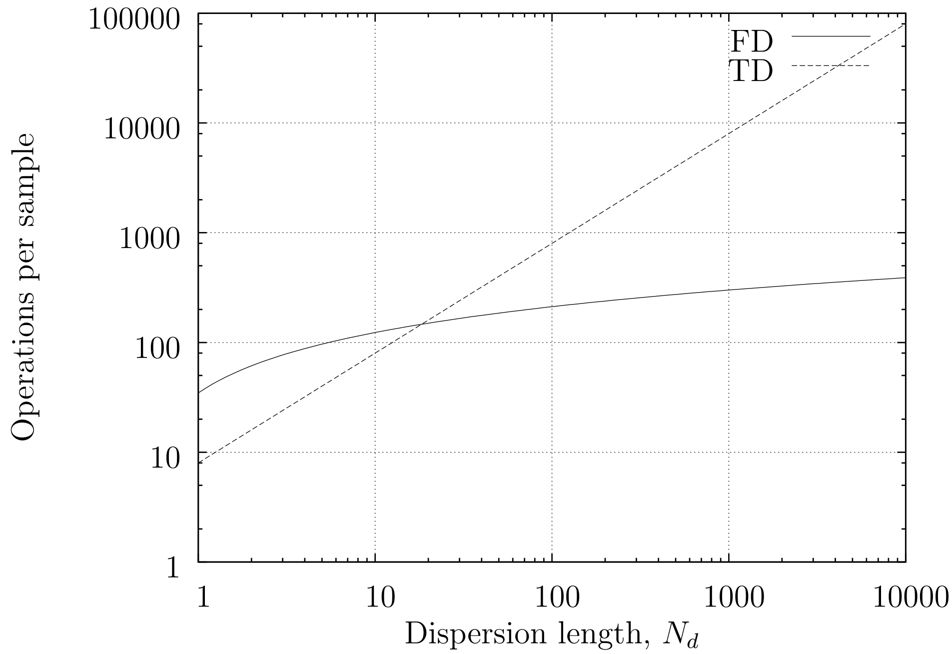

Figure 3.1: Computational cost of coherent dedispersion for time-domain (TD) and frequency-domain (FD) methods described in the text.

thumb to use for planning is $N_f \sim 4N_d$. With this approximation, the frequency-domain cost becomes $o(\log N_d) + o(1)$. Given this logarithmic scaling, the frequency-domain method is clearly more efficient for large N_d. Using realistic values for the proportionality constants, the crossover point occurs at $N_d \sim 15$ (see Figure 3.1), so the frequency-domain method is preferred in almost all cases where computation is the limiting factor.

The scaling of computational cost versus the various parameters which make up N_d can be summarized as follows:

$$\text{TD cost (op/s)} \propto N_c \left(\frac{\Delta\nu}{\nu}\right)^3 D \tag{3.8}$$

$$\text{FD cost (op/s)} \propto N_c \Delta\nu \log\left(\frac{\Delta\nu^2}{\nu^3} D\right) + C N_c \Delta\nu \tag{3.9}$$

Here, a factor of $\Delta\nu$ has been incorporated to convert from operations per sample to operations per second. N_c is the total number of channels to be dedispersed. This factor is useful in considering, for example, how to best partition a fixed total bandwidth into frequency channels.

3.3 System Architecture

A overview of the entire ASP system in shown in Figure 3.2. The basic signal path is as follows. The system receives dual-polarization analog signals from the telescope centered at a intermediate frequency (IF) of 400 MHz. These are mixed to baseband (0 MHz) in the analog front-end using a quadrature downconverter.

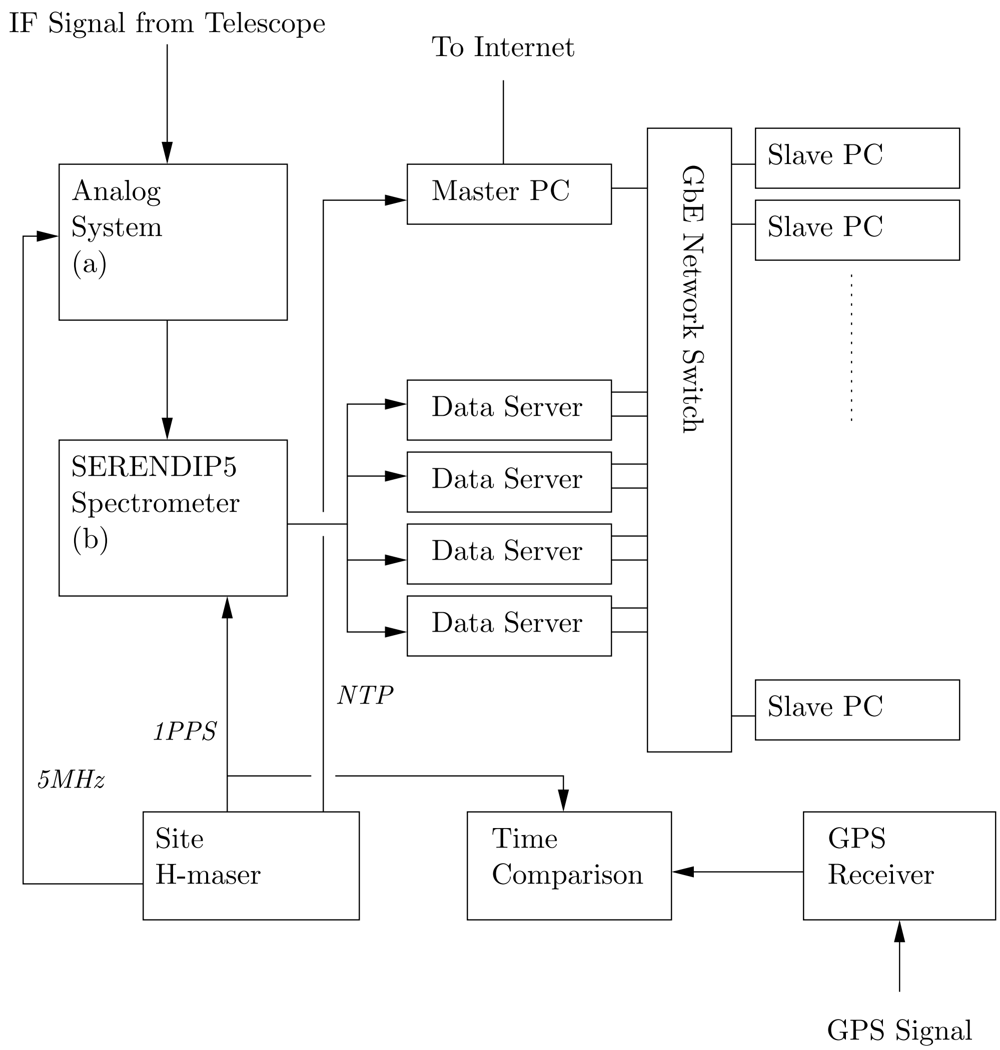

Figure 3.2: ASP System architecture overview. Major components include: (a) The analog front-end, discussed in §3.3.1, mixes the signal to baseband; (b) the SERENDIP5 spectrometer (§3.3.2) digitizes the signal and performs a polyphase filter bank; (c) the ARTS software package (§3.4) distributes data through the data servers to the slave PCs, which dedisperse, detect and fold the data. Also shown are the typical time distribution components.

The resulting signals are passed through matched pairs of 64-MHz low-pass filters, giving a total bandwidth of 128 MHz. These are digitized with 8 bits of precision in the SERENDIP5 spectrometer board. The SERENDIP5 system also divides the 128-MHz band into 32 4-MHz channels using a digital polyphase filter bank (PFB). The channelized data is then output into 4 "data server" PCs containing specialized high-speed parallel I/O cards.

The data servers reformat the raw data into ~4 MB chunks, adding timestamp and other identifying information. These chunks are passed through 2 channel-bonded gigabit Ethernet (GbE) interfaces to the "slave" PCs. The slave PCs apply the coherent dedispersion filter of Equation 3.6. They then detect and fold the data modulo the current apparent pulse period. This data is sent back to the "master" PC where it is combined into a single file and recorded to disk. These individual components are described in more detail in the following sections.

3.3.1 Analog Front End

The main purpose of the analog front-end system is to shift the incoming signal from IF centered at 400 MHz down to baseband. In quadrature downconversion, a single incoming signal $s_0(t)$ with frequency content near ω_0 is split in two, then multiplied by "local oscillator" (LO) tones of frequency ω_m whose phases are separated by 90° (i.e., sine and cosine waves). The resulting two physical signals can then be thought of as a single complex number, $s_m(t) = s_0(t)\cos(\omega_m t) - is_0(t)\sin(\omega_m t) =$

$s_0(t)e^{-i\omega_m t}$. This signal has components near frequencies $\omega_0 + \omega_m$ and $\omega_0 - \omega_m$. Low-pass filtering with cutoff $\omega_c < \omega_m$ removes the high frequency component, leaving a copy of the original signal, now shifted to zero frequency (assuming the choice $\omega_m \approx \omega_0$). The benefit of quadrature mixing is that this complex signal contains unique information over the full frequency range $-\omega_c < \omega < \omega_c$, as opposed to duplicate content in positive and negative frequencies as in a real signal. To be represented digitally, each component (real and imaginary) must be sampled at a frequency of $2\omega_c$. Obtaining equivalent bandwidth with a real (single-sideband) signal would require sampling twice as fast.

The drawbacks to quadrature downconversion are increased system complexity and potential systematic effects due to the "image rejection" problem. The former comes in the form of needing twice the amount of mixers, filters and samplers that a single-sideband system would. The systematic effects are more subtle. We can investigate the frequency content of the downconverted signal by taking the Fourier transform of $s_m(t)$:

$$s_m(\omega) = \left(s_0(\omega) \star \frac{\delta(\omega - \omega_m) + \delta(\omega + \omega_m)}{2} \right) g_I(\omega) - \left(i s_0(\omega) \star \frac{\delta(\omega - \omega_m) - \delta(\omega + \omega_m)}{2i} \right) g_Q(\omega) \quad (3.10)$$

Here, the factors inside the parentheses represent an ideal quadrature mixing operation - the Fourier transform has changed the time-domain multiplication to convolution. The $g_{I,Q}(\omega)$ factors represent the low-pass filtering process, allowing for the possibility that the two filter shapes are not identical. The commonly used subscripts I and Q refer to the real and imaginary components, respectively. Non-90° phase separation of the LO signals can be absorbed into the g factors by allowing them to be complex. Simplifying this expression results in the following:

$$s_m(\omega) = s_0(\omega + \omega_m)g_+(\omega) + s_0(\omega - \omega_m)g_-(\omega) \tag{3.11}$$

$$s_m(\omega) = \tilde{s}_m(\omega)g_+(\omega) + \tilde{s}_m^*(-\omega)g_-(\omega) \tag{3.12}$$

Here we have used the fact that since $s_0(t)$ is real, $s_0(\omega) = s_0^*(-\omega)$, and also defined $g_\pm(\omega) \equiv \frac{1}{2}\left(g_I(\omega) \pm g_Q(\omega)\right)$. The further definition of $\tilde{s}_m(\omega) \equiv s_0(\omega + \omega_m)$ as the result of a ideal downconversion highlights the main result of this analysis: Any non-ideal implementation of the mixing procedure, including gain and filter shape mismatches or non-90° LO phases, results in the final product being contaminated with a frequency-reversed “image” of itself. The magnitude of the image component depends on the magnitude of the mismatch. The effect of this error on precision timing is explored in §3.5.2.

In our implementation, the downconversion is done by a custom circuit board (quadrature downconverter; QDC) based around the Analog Devices AD8348 quadrature demodulator chip. This chip takes a input at $2\omega_m$ (800 MHz) and internally converts it into the necessary sine and cosine components, which it then uses to perform the mixing. This approach results in particularly good LO phase separation, with a typical error of $\lesssim 0.5°$. It also contains a variable-gain amplifier with a ~30 dB range that can be used to compensate for differences in the IF signal level. Each

Signal Description	I/E	Part Number	Power (dBm)		
			Min.	Typ.	Max.
IF Input (336-464 MHz)	E	AD8348	-28		-10
Frequency reference (5 MHz)	E	Valon 5003 (Si4333)	-10	0	+13
Time reference (1PPS)	E	IDT49FCT	TTL Levels		
Baseband I/Q	I	SERENDIP5		0	
LO (800 MHz)	I	AD8348	-12	-10	0
Sample Clock (128 MHz)	I	FIN1102	-16		+13

Table 3.1: Recommended input power levels for ASP systems. E: Externally applied signals; I: Internally generated signals.
[a] TTL levels.

QDC board can handle one polarization, and contains two matched 64-MHz low-pass filters. The output signals are sent to the SERENDIP5 spectrometer to be digitized.

In addition to the QDC boards, the analog system contains several other critical components. The 800-MHz signal is generated by a Valon 5003 dual synthesizer module. This synthesizer also generates 128 MHz, which is used as the sampling clock (see §3.3.2). It takes a reference frequency input at 5 MHz, typically generated by a hydrogen maser at each observatory. A custom level converter board changes the 128-MHz signal and a externally supplied 1 pulse-per-second (1PPS) timekeeping signal to low voltage differential signal (LVDS) pairs. These are sent to the SERENDIP5 via RJ-45 connectors and Category-6 cable. Finally, a DAC interface board allows the QDC gains to be set via computer control, using a standard PC parallel port. Table 3.1 lists the recommended input power levels for all analog signals in the system.

3.3.2 SERENDIP5 Spectrometer

The SERENDIP5 spectrometer is a general-purpose signal processing board developed by the UC Berkeley SETI group. It contains four 8-bit analog to digital converters (ADCs) that can be run at speeds up to 200 MHz. The ADC outputs are routed to a Xilinx Virtex-II series field-programmable gate array (FPGA) chip. The board is designed to be attached to a compact PCI (cPCI) backplane, which provides power and a low data rate I/O connection for programming the FPGA and controlling the board's operation. This control is done through a single-board host computer also attached to the cPCI backplane. An optional extension board provides four high data rate (up to 200 MB/s each) outputs designed to be used with the Engineering Design Team (EDT) PCI-CD series parallel I/O boards.

In our application, the SERENDIP5 board contains a Xilinx XC2V4000 FPGA that is programmed to perform a digital filter bank operation on the data. We use a efficient implementation known as a polyphase filter bank (PFB). A traditional digital filter bank divides a (complex) input signal into N subbands by applying frequency-shifted copies of a M-point finite impulse response (FIR) filter, h_j. This requires $M \geq N$, and the problem is simplified by setting $M = ON$, where O is a integer. The output in each channel is then downsampled by a factor of N, preserving the total data rate. For an input data stream s_j, the output in the k-th channel, $s_j^{(k)}$ is given by:

$$s_j^{(k)} = \sum_{l=0}^{M-1} s_{jN+l} h_l e^{2\pi ikl/N} \tag{3.13}$$

Since the exponential factors and filters can be precomputed, this results in a total computational cost of $o(M) = o(ON)$ operations per input time step.[4] A PFB implementation takes advantage of the filter bank's frequency symmetry to reduce this computational load. Consider rewriting the sum in Equation 3.13 as two sums, using the following transformation:

$$\begin{aligned} l &\to Nm + p, \\ \sum_{l=0}^{M-1} &\to \sum_{m=0}^{M/N-1} \sum_{p=0}^{N-1}. \end{aligned} \tag{3.14}$$

Applying this transformation lets us rewrite Equation 3.13 as follows:

$$s_j^{(k)} = \sum_{p=0}^{N-1} e^{2\pi ikp/N} \left(\sum_{m=0}^{O-1} s_{jN+mN+p} h_{mN+p} \right) \tag{3.15}$$

The filter bank expression now has taken the form of a N-point discrete Fourier transform (DFT) of the factor inside the parentheses. The DFT portion of the computation can be implemented as a FFT, reducing the computational cost to $o(O) + o(\log N)$ per input time step.

In the ASP systems, the SERENDIP5 board is clocked at 128 MHz with the clock signal supplied by a external synthesizer (see §3.3.1). The four analog inputs are interpreted as real and imaginary components of the two antenna polarizations. Each polarization is then fed through a 32-channel PFB with $O = 24$, which results

[4]Alternatively, this is $o(MN)$ operations per output time step. However, the duration of a output step depends on N. O is commonly referred to as the PFB "overlap factor."

in 4-MHz channels.[5] The PFB outputs are truncated to 8-bit precision and output through the four EDT connectors to the data server computers. Each EDT connection receives 8 of the 32 channels.

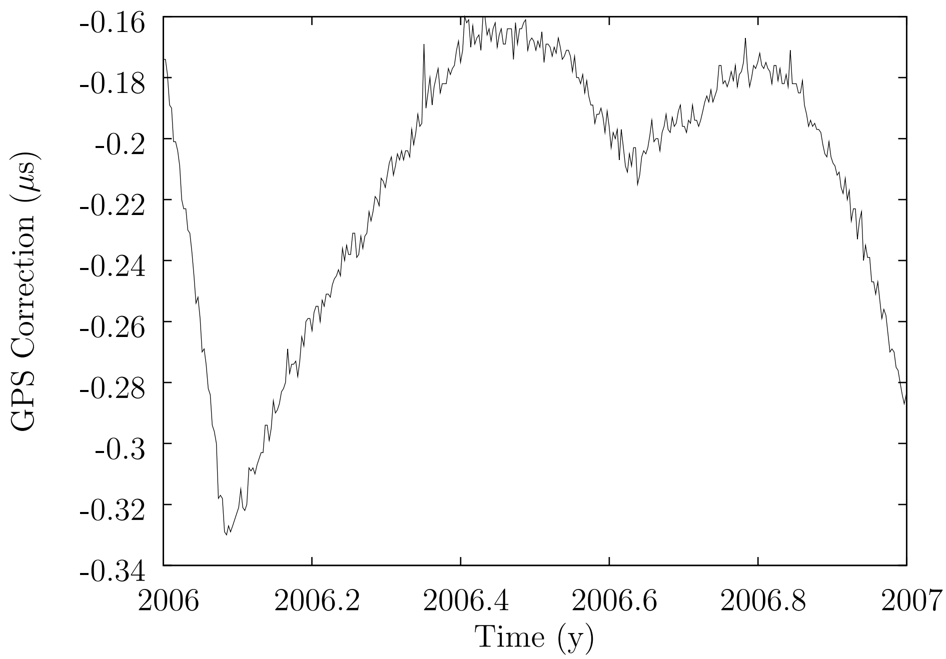

Figure 3.3: Example of the H-maser−GPS time correction from the GBT clock comparison system. The correction is typically on the order of 100 ns over timescales of years.

For precision pulsar timing, accurate timekeeping is critical. Each sample must be time-tagged with a precision of ∼10 ns to avoid systematic errors. In our system this is accomplished through the use of an externally supplied 1PPS tick. This signal is usually supplied by the observatory clock system, which is often based on a local hydrogen maser. While masers are excellent clock sources up to timescales of days,

[5]The original design, still in use by GASP, used $O = 8$.

we need a more stable source for long-term pulsar timing. The drift of the maser relative to the more accurate global positioning system (GPS) timescale is logged at the observatory and can be corrected for later. The magnitude of the drift is typically on the order of 100 ns over timescales of years. An example of this "GPS correction" from the Green Bank Telescope clock comparison system[6] is shown in Figure 3.3. Alternatively, GPS receivers can directly supply the 1PPS signal. Measurements based on the GPS timescale can be converted to a number of standard timescales via records kept by the BIPM[7]. The FPGA is programmed to start data output to the EDT cards on a 1PPS tick, after being armed mid-second via the EDT "FUNCT" output line from the data server. Since the data output is known to start on an integer second, it can be accurately time-tagged by the data server computer, whose internal clock is kept synchronized to the ms level using Network Time Protocol (NTP). Finally, the SERENDIP5 keeps track of the misalignment between the 4-MHz output clock and the 1PPS tick in order to provide sub-sample time-tagging of $\sim$1 ns accuracy. This number is provided through the cPCI interface, and recorded along with the final data.

[6] see `http://www.gb.nrao.edu/~fghigo/timer`

[7] Bureau International des Poids et Mesures

3.4 Software Components

Following the PFB operation performed by the SERENDIP5 board, all further signal processing is accomplished by software running in a “Beowulf” cluster of personal computers (PCs). The different ASP installations use slightly different PC configurations, however all are Intel-based systems running a form of Red Hat Linux. A typical PC system is in a rack-mounted enclosure containing dual Xeon CPUs clocked at 2.4 GHz, 1 GB of RAM, and between 200 GB and 500 GB of hard drive space. A full cluster consists of one master PC, four data servers and 16 or more processing (or “slave”) nodes. These are networked together through a gigabit Ethernet switch. The master system provides user account information (Network Information Services; NIS), shared disk space (Network File System; NFS) and time synchronization (Network Time Protocol; NTP) services for the rest of the cluster. This PC is also typically the only computer in the group to have an external network connection, either directly to the Internet or to the observatory’s general network.

The incoming data are handled in real time by a set of programs called the ASP Real Time Software (ARTS) package. A block diagram of the various programs and their interactions is shown in Figure 3.4. The software is designed as a set of modular components, with minimal interactions between them. All components are written in the standard ANSI C programming language. The signal flow follows a “push” architecture: When ready, each program places its output data in a shared-memory ring buffer from which the data are picked up by the next program in the chain. An

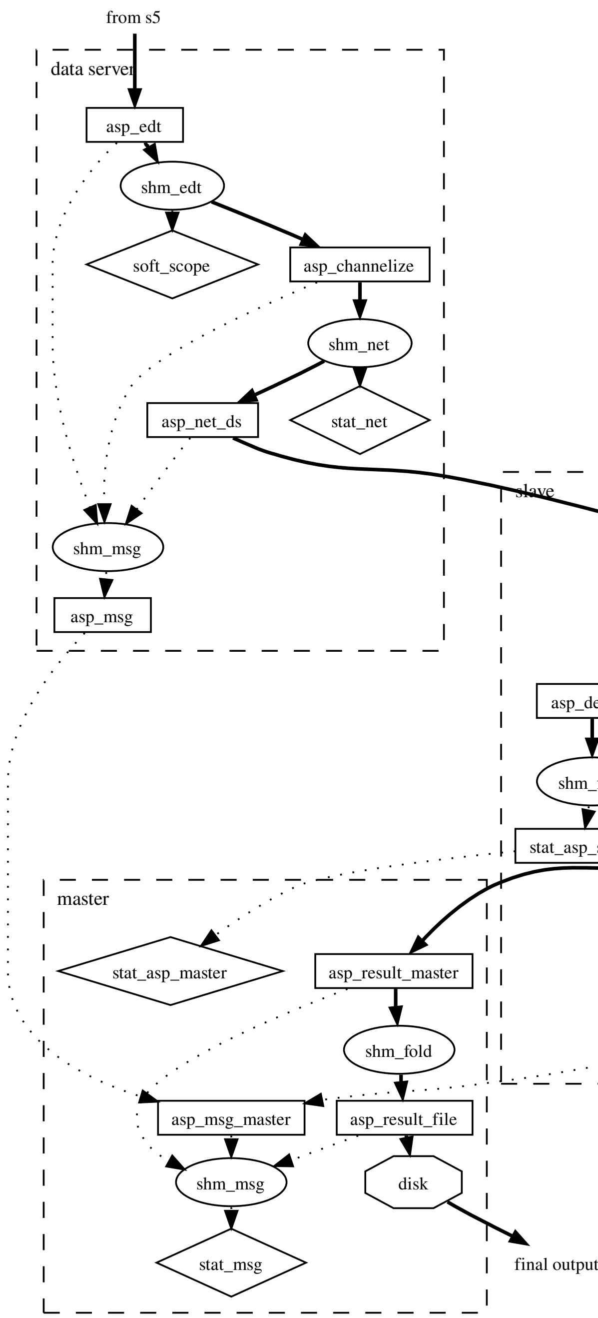

Figure 3.4: Logical block diagram of the ARTS software package data flow. The primary data path is the bold line, while monitor information follows dotted lines. Ellipses represent shared memory buffers, and diamonds are monitor programs. See text for details.

individual program does not need to know who will be receiving its output. The ring buffer (or FIFO) allows the programs to operate asynchronously, compensating, up to a point, for variations in their speed. This can be contrasted with a "pull" architecture in which later programs actively request data from preceding ones. Since all communication between programs happens through shared memory buffers, it is easy to construct monitor programs that attach to the shared memory segments and inspect the data without disrupting its flow. The modular design also allows for a large amount of flexibility: New analysis procedures can be implemented without having to rewrite the entire data flow. The following sections describe in more detail the operation of the package in the typically used mode, real-time coherent dedispersion and folding.

3.4.1 Data Servers

The data server computers have three main tasks: To read the raw data in from the SERENDIP5 spectrometer via the EDT PCI-CD I/O cards; to reorganize (or "corner-turn") the data from channel-order to large single-channel time ordered chunks; and finally, to distribute these chunks to the slave nodes for further processing. Each of these tasks is handled by a separate program: *asp_edt*, *asp_channelize* and *asp_net_ds* respectively. The data is passed between programs through two shared memory buffers: *shm_edt* and *shm_net*. The *asp_edt* program also sets the initial data start trigger, and time-stamps the first sample received.

Since the data servers' purpose is mainly reorganization and transfer of the incoming data (as opposed to mathematical computations), their operation is not very CPU-intensive. Instead, the performance-limiting factor is the speed of several internal data busses. Each data server handles a total data rate of 1 Gb/s. This data stream comes in from the EDT card over a PCI bus, is stored via DMA in the computer's main memory, takes several trips over the front-side bus (FSB) to the CPU cache memory while being reorganized, and is finally sent out to the Ethernet adapters over another PCI bus. In order to handle the full data rate, several design choices were made: The data server motherboards (Supermicro XD5GPE) have dual-PCI busses so the EDT card and network adapters can be separated. The *asp_channelize* program is optimized to minimize FSB traffic; the correct optimization is a function of CPU cache size. Finally, two GbE adapters are tied together using a process called "channel bonding" and used simultaneously to distribute the outgoing data. Due to network overhead, achieving a 1 Gb/s data rate through a single GbE adapter is not possible.

After channelization, the data is formatted as ~4 MB single-channel time blocks. Each data server has a list of slave nodes to which it sends data in a round-robin fashion, assigning one data block to each node, and looping back to the beginning of the list when necessary. There are no software restrictions on the number of nodes or channels that can be used. If the full data rate is not able to be processed by the cluster, the bandwidth can be reduced at this point by removing channels.

Several additional features that are not yet implemented would increase performance: First, each slave node currently receives a fixed data rate during operation. A dynamic load-balancing system would more efficiently make use of a set of nodes with varying capability, as well as provide more insurance against temporary computing interruptions. Second, the current design only allows each node to receive data from a single data server. Removing this restriction would allow a more even allocation of computing resources across the band. These improvements are planned for future versions.

3.4.2 Processing Nodes

The processing, or "slave" nodes are where all of the actual signal processing computation takes place. The data blocks are received over the network from the data servers by the *asp_net_slave* program and placed into another *shm_net* buffer. They are then read by the *asp_dedisp* program that performs the following tasks:

1. Converts the raw data to floating point.

2. Applies a Fourier transform of appropriate length. The FFT is implemented using the free, open-source FFTW package (Frigo and Johnson, 2005).

3. Multiplies by the dedispersion filter of Equation 3.6.

4. Inverse Fourier transforms back to the time domain.

5. Detects (squares) the data to get power versus time, and cross-multiplies polarization terms.

6. Averages (or "folds") the data modulo the current apparent pulse period. Ephemera for folding are supplied by the standard pulsar timing program TEMPO.

The folded output is placed into a *shm_fold* buffer. Once the required integration time, typically several minutes, has elapsed, the *asp_result_slave* program sends the result to the master node.

Alternatively, the raw data can be saved to disk for later (non-real time) processing. The *asp_write_raw* program reads raw data from *shm_net* and records it to disk. Later, *asp_read_raw* will read the saved data and play it back into the *shm_net* buffer. The disk writing procedure is useful if the required processing cannot be completed in real time, or if the data needs to be reprocessed several times with different parameters.

A typical slave node is able to dedisperse approximately one 4-MHz channel in real time. As discussed in §3.2.1, the computational cost of coherent dedispersion grows as the dispersion measure is increased. Figure 3.5 shows the number of channels which can be handled in real time per slave node in the GASP system as a function of dispersion length (N_d). Also shown is an estimate of how this value is expected to scale, based on the arguments presented in §3.2.1. Assuming each node achieves a constant number of operations per second, Equation 3.8 can be used to predict the scaling versus N_d. The measured value tracks the prediction fairly well for $256 <$

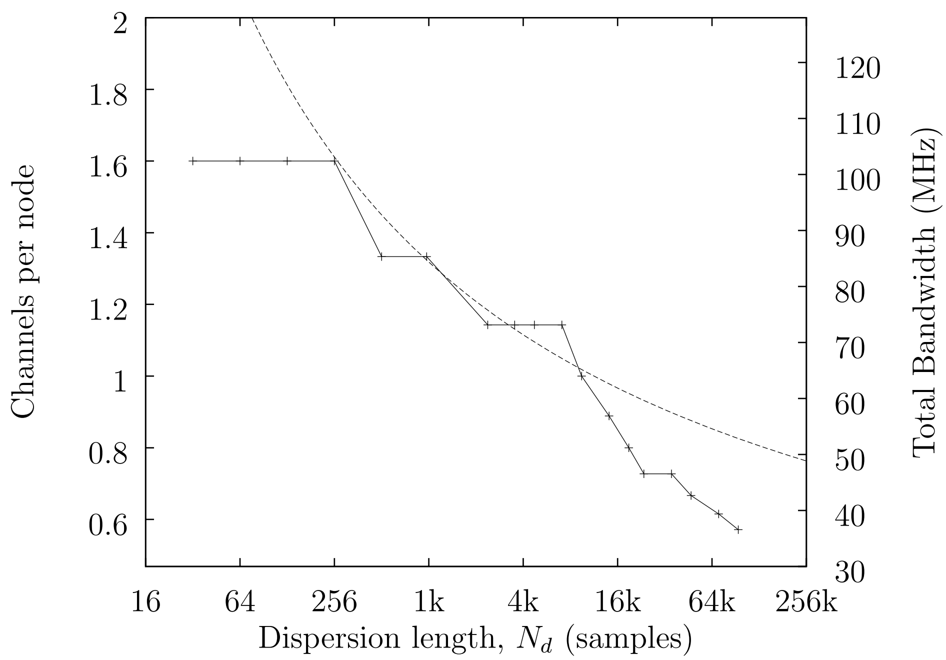

Figure 3.5: Number of 4-MHz channels that can be processed in real time per slave node in the GASP system as a function of dispersion length (N_d). The dashed line shows the prediction based on computational cost as discussed in §3.2.1. The equivalent total bandwidth for a 16-node system is shown on the right.

$N_d < 8192$. In this range, deviations are mainly due to quantization effects: The number of nodes and channels are integer values, and the FFT length used (N_f) is always an integer power of 2, whereas the prediction assumes these parameters are continuous. Outside this range, the measurements differ significantly from the expected scaling law. For small N_d, the prediction is an overestimate because it neglected to count several other tasks that must be performed simultaneously in the same system: Detection, folding, networking and memory operations become a larger fraction of the total workload as N_d decreases. For large N_d, the FFT efficiency is reduced due to hardware limitations. Once $N_f \geq 32768$, a single transform does not fit into the Intel Xeon's CPU cache memory. This incurs more overhead in the form of data transfers to and from the main system memory.

3.4.3 Master Node

The master node serves primarily as a connection to the outside world and as a provider of various network services for the rest of the cluster, as described above. Its only role in data processing is to combine the final results from all the slave nodes into a single output file. Folded profile data is read from the slaves by *asp_result_master* and placed into a *shm_fold* buffer. When all data for a single integration is collected from the slaves, the *asp_result_file* program formats it into a FITS[8] file and writes it to disk.

[8]Flexible Image Transport System (Hanisch et al., 2001).

This node also acts as the primary interface point for monitor and control of the system. Configuration information necessary for the system's operation is stored in a XML[9] file located on a networked disk. This file is read directly by all ARTS programs on startup. It contains a list of which data servers and slave nodes will be used, and the associations between the two. It also contains the list of channels to be skipped (setting the total bandwidth) and additional information to be recorded in the final data file such as source name, pointing coordinates, observatory code, and dispersion measure used for coherent dedispersion. Editing this file by hand is currently the only way to control these runtime parameters.

The incoming data can be viewed in real-time using several monitor programs: *soft_scope* provides a general-purpose software oscilloscope and spectrum analyzer that can view samples versus time, histograms of sample values, and data power spectra. *stat_asp* displays the folded pulse profiles in each channel as they integrate. Other programs (*stat_msg*, *stat_net*) display log messages and memory buffer status for the cluster.

3.5 Systematic Profile Shape Errors

In this section, we examine two common detection-related systematic effects that alter the measured pulse profile shape. If variable, pulse profile errors can negatively affect timing measurements (see Chapter 2).

[9]Extensible Markup Language

3.5.1 Effect of Quantization

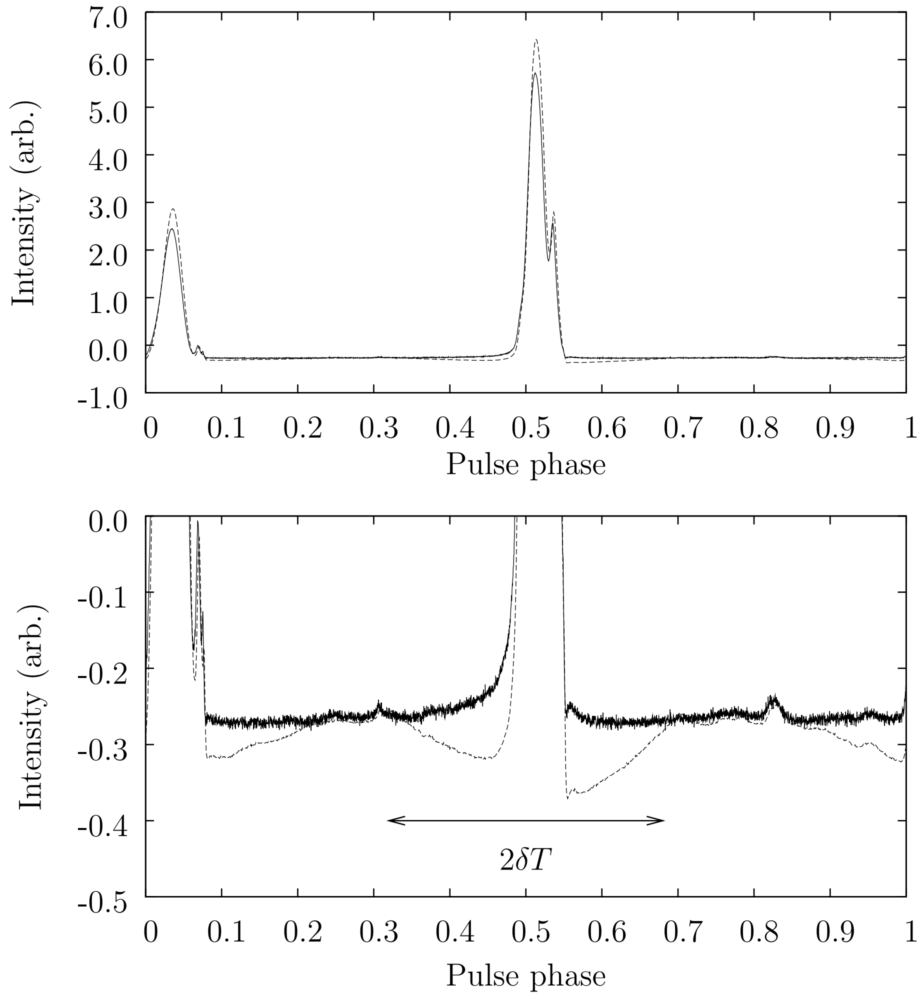

Figure 3.6: Pulse profiles for PSR B1937+21 at 1420 MHz using the ABPP (dashed) and ASP (solid) backends. The lower panel shows a blowup of the profile baselines, highlighting the dips caused by low-bit quantization. See text for details.

In order to be represented digitally, an analog signal must be both sampled at finite time intervals, and quantized with a finite amount of precision. That is, once per sampling interval (τ_s), the incoming voltage is approximated by selecting one out of a set of 2^{N_b} discrete values, where N_b is the number of bits of precision. Quantization is a nonlinear operation that distorts the signal, with larger values of N_b introducing less distortion. However, in designing digital signal processing systems, there are often limitations based on the total data rate N_b/τ_s bit/s, making it possible to trade quantization accuracy for higher total bandwidth. Radio astronomy instrumentation has traditionally gone this route, often employing 1-bit or "1.5-bit" (3-level) systems (these have the added advantage of extremely simple digital multiplication operations). The effects of quantization distortion are decreased signal-to-noise ratio, and a nonlinear relation between the input (analog) signal power and output (digital) power. The former effect is usually more than compensated for by the increased bandwidth. For stationary signals, the measured power spectrum can be fixed by applying the Van Vleck correction (Van Vleck and Middleton, 1966), or a similar method.

Pulsar signals are somewhat harder to deal with due to the rapidly changing power versus time during each pulse period. Dispersion further complicates the situation: As the dispersed pulse sweeps through the observing band, starting at the highest frequency, it increases the total input power. This acts to decrease the apparent power at all other frequencies due to the nonlinear input to output power relation. When the pulse is aligned by the dedispersion process, this apparent power decrease shows up as a "dip" around the dedispersed pulse with a width of twice the dispersion smearing time ($2\delta T$). This effect is described in detail by Jenet and Anderson (1998), who also outline methods for minimizing its effect. However, the simplest and most direct way to reduce this kind of error is to build data recording systems that use higher numbers of bits. This is the approach we have taken with the ASP systems.

Figure 3.6 shows average pulse profiles for PSR B1937+21 taken with two different backends, the Arecibo-Berkeley Pulsar Processor (ABPP; Backer et al. (1997)), a 2-bit system, and ASP which uses 8 bits. These "template" profiles (see Chapter 2) are averages over all available data from the Arecibo telescope at a radio frequency of 1420 MHz. The ABPP profile is much less noisy due to the longer span of data taken with that backend. It also clearly shows the dips caused by the uncorrected 2-bit quantization. The arrow in the plot shows the calculated dispersion smearing time $\delta T \sim 288$ μs for the 1.4-MHz channel bandwidth that was used. Any comparable effect in the ASP data is much weaker, due to the 8-bit quantization and larger channel bandwidth.

3.5.2 Quadrature Downconversion

The quadrature downconversion process was described in §3.3.1, where it was noted that any non-ideal implementation of the procedure results in a frequency-reversed image component appearing in the signal (Equation 3.11). When combined with dispersion, this leads to a systematic profile shape error. Given an LO frequency of ν_m, data taken at frequency $\nu_m + \nu$ will also contain a low-amplitude copy of data from frequency $\nu_m - \nu$. Due to dispersion, this image component will be delayed in time from the main signal according to Equation 3.4. Furthermore, due to the complex conjugate in Equation 3.11, applying the coherent dedispersion filter $H^*(\nu_m + \nu)$ does not dedisperse the image. Instead, it has now effectively been multiplied by a dispersion filter $H^*(\nu_m - \nu)H^*(\nu_m + \nu) \simeq [H^*(\nu_m)]^2$, the first term coming from the ISM and the second from the coherent dedispersion process. The result is that the image component is spread in time by twice the usual dispersion amount.

We can investigate the effect of this error on timing by applying the concepts developed in Chapter 2. Equation 2.7 described the first-order effect of a small profile variation on the measured TOA. In this case, the variation is a delayed, dispersed copy of the true profile shape, and the corresponding TOA shift can be calculated:

$$\delta(\nu) = R(\nu)A(\nu)\frac{C'_{pp}(\phi_d(\nu))}{C''_{pp}(0)} = R(\nu)A(\nu)\frac{\sum_k k|p_k|^2 S_k \sin(2\pi k \phi_d(\nu))}{2\pi \sum_k k^2 |p_k|^2} \tag{3.16}$$

$R(\nu)$ is the image rejection ratio, ϕ_d is the dispersion delay in units of pulse phase between frequencies $\nu_m + \nu$ and $\nu_m - \nu$, and $A(\nu)$ is the intrinsic signal strength ratio between the two frequencies, as follows:

$$R(\nu) = \left|\frac{g_-(\nu)}{g_+(\nu)}\right|^2 \tag{3.17}$$

$$A(\nu) = \left|\frac{s_0(\nu_m - \nu)}{s_0(\nu_m + \nu)}\right|^2 \tag{3.18}$$

$$\phi_d(\nu) = \frac{T(\nu_m + \nu) - T(\nu_m - \nu)}{P} = \frac{D}{P}\left(\frac{1}{(\nu_m + \nu)^2} - \frac{1}{(\nu_m - \nu)^2}\right) \tag{3.19}$$

S_k is a low-pass filter that describes the dispersive spreading of the image component:

$$S_k = \frac{\sin(\pi k \xi)}{\pi k \xi} \tag{3.20}$$

$$\xi = \frac{2\delta T}{P} = \frac{2\Delta\nu D}{P\nu_m^3} \tag{3.21}$$

It can be seen from these equations that the magnitude of the effect declines as dispersion increases (relative to the pulse period). This time shift will be seen most easily in low-DM pulsars with sharp profile features.

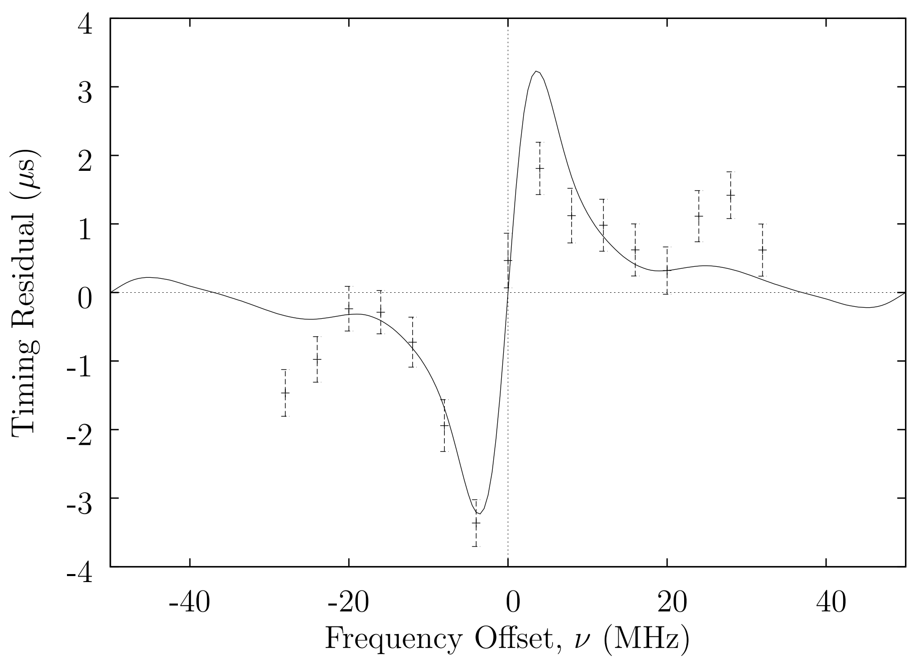

Figure 3.7: Average timing residual versus frequency for PSR J2145−0750. The solid line shows the predicted effect of imperfect quadrature downconversion (Equation 3.16).

Figure 3.7 shows an example of this effect using data from PSR J2145−0750 recorded with the GASP system at the NRAO Green Bank Telescope. J2145−0750 has a pulse period $P = 16$ ms and a dispersion measure DM $= 9.00$ pc cm^{-3}. These data were taken at a center frequency $\nu_m = 1400$ MHz, with a channel bandwidth $\Delta\nu = 4$ MHz. The predicted residual versus frequency is computed using Equation 3.16, assuming that $R(\nu)$ is constant. The inferred image rejection ratio is

-15 dB. Deviation from the model at $|\nu| > 20$ MHz is likely due to $R(\nu)$ increasing as a function of frequency. The slight asymmetry between positive and negative ν may be caused by the $A(\nu)$ term, or could be due to intrinsic profile shape variation versus frequency in this pulsar. In July 2005, the GASP analog filters were upgraded to a more closely matched set that gives image rejection of better than -30 dB.

Chapter 4

Timing Analysis of 16 Millisecond Pulsars

4.1 Introduction

In this chapter, we present timing results for 16 millisecond pulsars. This analysis applies the signal processing and data analysis ideas developed in Chapter 2 to real data, which was recorded using the hardware systems presented in Chapter 3. For our purposes, the primary results of this analysis are the timing residuals, which are analyzed for the presence of gravitational radiation in Chapter 5. Secondary results, which are interesting in their own right but not elaborated upon here, include the fitted timing model parameters, the integrated pulse profiles, and dispersion measure versus time.

4.2 Description of Observations

The data presented here are the result of ongoing long-term pulsar monitoring programs using two of the world's largest telescopes: The 305-m NAIC Arecibo telescope[1] in Puerto Rico, and the 100-m NRAO Green Bank Telescope[2] in West Virginia. The detection of low frequency gravitational radiation (GW) is a primary scientific motivation for these programs. The other primary goal is the precise measurement of the orbital parameters of pulsars in binary systems, as these can provide stringent tests of general relativity (e.g., Kramer et al., 2006). The dataset is also a rich source of information about the interstellar medium, discussed further in Chapter 6.

Both programs follow the same basic outline: A ∼12 hour block of time is allocated once per month for these observations. During this time, each of ∼10 pulsars is observed for about one hour. The time spent on each pulsar is divided between two different observing bands[3], widely separated in frequency, in order to accurately measure the current dispersion measure (DM; see §3.2) towards each pulsar. The pulsar signals are processed in real time by a coherent dedispersion backend (see Chapter 3), and are recorded as a set of "folded" pulse profiles. Each profile represents a average over several minutes of time, and several MHz of bandwidth. The total bandwidth recorded is a function of receiver bandwidth, observing frequency and

[1]The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with the National Science Foundation.

[2]The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

[3]Current radio telescopes can typically only deliver one band at a time, with a fractional bandwidth of order 0.1. This situation will likely improve in the near future.

Source	Observatory	ν (MHz)	Epochs	Span (y)	Total Time (h)
J0030+0451	AO	430,1410	23	2.12	19.7
J1640+2224	AO	430,1410	27	2.28	20.5
J1713+0747	AO	1410,2350	47	2.28	32.2
B1855+09	AO	430,1410	23	2.24	21.1
B1937+21	AO,GBT	820,1410,2350	73	2.65	37.7
J2019+2425	AO	327,430	24	2.24	12.8
J2317+1439	AO	327,430	25	2.24	29.8
J0218+4232	GBT	820,1400	31	2.29	12.2
J0613−0200	GBT	820,1400	38	2.29	23.5
J1012+5307	GBT	820,1400	51	2.65	30.3
J1455−3330	GBT	820,1400	38	2.65	18.4
J1643−1224	GBT	820,1400	55	2.65	28.3
J1744−1134	GBT	820,1400	48	2.65	24.5
J1909−3744	GBT	820,1400	35	2.65	19.8
J1918−0642	GBT	820,1400	45	2.65	21.0
J2145−0750	GBT	820,1400	37	2.64	17.4

Table 4.1: Observation parameters for the total dataset. Listed parameters are: Observatories used (AO = Arecibo Observatory; GBT = NRAO Green Bank Telescope), radio frequencies observed (ν), number of separate days (epochs) of observation, total time span of observations, and total amount of observing time per source.

DM, and is typically of order 50 MHz. A list of the pulsars that will be used for this analysis and basic dataset parameters is shown in Table 4.1. These are divided into two groups, based upon which telescope is typically used to view each one.

The Astronomy Signal Processor backend (ASP; described in detail in Chapter 3) was installed at Arecibo in late 2004. Here we will analyze a set of 6 consistently observed sources. At Arecibo, most sources are observed at 430 and 1420 MHz, with some bright pulsars observed at 2350 MHz, and some fainter ones at 327 MHz. The ability of the system to rapidly switch observing bands lets us record data at all bands during the same day. However, the restrictive elevation range of the telescope[4] confines this program to sources with declinations between 0° and 30°. Furthermore, each source is only visible for ~2 hours per day, reducing the total number which can be observed in any one session.

The NRAO Green Bank Telescope (GBT) began operation in 2001. We started a long-term pulsar timing program soon after, in mid-2004, using the Green Bank Astronomy Signal Processor backend (GASP, a clone of ASP). The GBT is a fully steerable, 100-m single dish telescope. Our GBT timing program concentrates on pulsars not visible from Arecibo, the number of which increases towards the Galactic Center (near declination −30°). The wide angular range of the telescope is important for GW detection, as the expected GW signal is a correlation of timing residual versus pulsar angular separation (see §5.4.2). Also, the longer source tracking time

[4]The extremely large primary reflector is fixed to the ground.

allows us to view more sources during each session. We are currently monitoring a set of 12 pulsars, including 1 overlap with Arecibo (PSR B1937+21) and one which is not part of this study (PSR B1821–24). All of the GBT sources are observed at 820 and 1400 MHz. Since the GBT is inefficient at switching between these two bands, our observations are scheduled in two 8-hour blocks on separate days, one for each band, typically within a week of each other. This has sometimes resulted in a reduction of our ability to measure DM(t) (if the two sessions are scheduled far apart), and in general makes the measurement dependent on telescope scheduling. Short-period binary systems (see Table 4.2) will have advanced through several orbits over the course of a week, coupling the DM measurement to the binary parameter measurements. Future development at the GBT may include dual-band feeds, which would alleviate this problem.

A table of the basic physical parameters of this group of pulsars is given in Table 4.2. The sources in this set were chosen to maximize the possible timing precision, giving the best chance of detecting a very weak GW timing component. The pulsars listed here either have demonstrated records of highly stable timing, or have parameters which suggest such stability: First, all the sources are millisecond pulsars (MSPs). The high spin frequencies imply a narrow pulse profile, typically on the order of several hundred μs, which allows accurate time tagging. Furthermore, MSPs are old: They are thought to be slow pulsars which had “died” (became radio-quiet), then later were spun up and reactivated due to accreting matter off of a companion

Source	P (ms)	DM (pc cm^{-3})	P_b (d)	$S_{1.4}$ (mJy)
J0030+0451	4.87	4.3	–	0.6
J0218+4232	2.32	61.3	2.0	0.9
J0613−0200	3.06	38.8	1.2	1.4
J1012+5307	5.25	9.0	0.6	3.0
J1455−3330	7.98	13.6	76.2	1.2
J1640+2224	3.16	18.4	175.4	2.0
J1643−1224	4.62	62.4	147.0	4.8
J1713+0747	4.57	16.0	67.8	10.4
J1744−1134	4.07	3.1	–	5.3
B1855+09	5.36	13.3	12.3	3.1
J1909−3744	2.95	10.4	1.5	2.1
J1918−0642	7.64	26.6	10.9	
B1937+21	1.56	71.0	–	12.5
J2145−0750	16.05	9.0	6.8	9.4
J2317+1439	3.45	21.9	2.45	4.0

Table 4.2: Basic physical parameters of the observed set of pulsars. Listed parameters are: Spin period (P), dispersion measure (DM), binary period (P_b) and nominal flux at 1.4 GHz ($S_{1.4}$). Sources with no P_b given are not in binary systems. Observed fluxes can vary dramatically due to interstellar scintillation.

object. This scenario explains why almost all the MSPs on the list are in binary systems. Their large age implies that the MSPs will be more rotationally stable than young pulsars, which display a wide variety of rotational instabilities such as drifts and glitches (sudden changes in spin period). Only one MSP glitch has ever been observed (Cognard and Backer, 2004), and the magnitude of MSP period drift is typically orders of magnitude smaller than in young pulsars. These intrinsic rotational instabilities, collectively known as "timing noise," will mask a GW signal, and will eventually present a limit on how good a measurement can be made. Finally, none of the MSPs presented here are members of globular clusters. Globular clusters are rich in pulsars, but the acceleration due to the cluster potential can also mask the GW signal.

4.3 Timing Analysis Procedure

Although the basic data analysis flow used for pulsar timing was described previously, in Chapters 1 and 2, we will give a brief recap here, then discuss some specific implementation details used in this analysis. The raw data output from a coherent dedispersion backend such as ASP are folded pulse profiles. The dedispersed pulsar signal is split into several (typically $\sim$16) frequency channels, and in each is integrated modulo the current apparent pulse period (folded) for several minutes. This is done in real-time, and requires a initial, approximately correct timing model. The folded profiles are recorded to disk, and all further analysis takes place off-line. The time shift between each folded profile and a standard template profile is determined (see Chapter 2), representing the difference in rotational phase between the actual pulsar and the timing model. This phase difference and the average model phase are converted to time values (using the current apparent period) and subtracted from the integration's midscan timestamp. This final quantity is called a pulse time of arrival (TOA), and represents the time nearest the middle of the integration where the pulsar rotational phase was zero. The longitude on the pulsar corresponding to phase zero is arbitrary (but constant), and is set by the template profile definition. The TOAs are then used to fit for a improved timing model using the standard pulsar timing software package TEMPO.[5]

Referring the pulse arrival time to the middle of the integration reduces systematic effects that can be caused by folding with an inaccurate timing model. A difference in spin frequency between the model and the pulsar will smear the integrated pulse profile by an amount

$$\delta t = \frac{\Delta\Omega}{\Omega} T_{int}, \tag{4.1}$$

where T_{int} is the integration time, Ω is the spin frequency, and $\Delta\Omega$ is the difference between the model and true frequencies. Since to leading order, the smearing is symmetric about the middle of the integration, the TOA will not be "pulled" one way or the other with respect to that point. In our systems, $T_{int} \sim 100$ s, and

[5]TEMPO can be found at `http://www.atnf.csiro.au/research/pulsar/tempo/` and has been developed and maintained by the Princeton and ATNF pulsar research groups. It will soon be superseded by Tempo2 (Hobbs et al., 2006), currently under development.

$\Delta\Omega/\Omega \sim 10^{-8}$, leading to a smearing of $\sim$1 μs.

Our analysis closely follows the procedure just described, with the addition of a second pass through the data to determine the time variation in dispersion measure (DM), and examine profile shape variations. We perform the following steps for each pulsar:

1. Using the methods described in Chapter 2, we simultaneously determine both TOAs for each folded profile, and template profiles. A separate template is made for each observing band.

2. Profiles with very low signal to noise ratio (less than 0.75) are cut and the rest are fit to a constant-DM timing model.

3. The residuals from this fit are used to determine DM(t) and the constant phase shift between the templates for each band.

4. The DM information from step 3 is used to integrate the folded profiles over frequency in each band.

5. The TOAs are recalculated using the frequency-integrated profiles.

6. The profiles are run through the principal components analysis (PCA) described in Chapter 2 to identify profile variation, and PCA-based timing corrections are applied, if appropriate.

7. These frequency-integrated TOAs are used for a final timing model fit, and the timing residuals and fit design matrix are saved for future analysis.

Integrating over frequency serves three main purposes: First, it reduces the volume of data which must be passed to TEMPO, the PCA analysis, and the GW analysis of Chapter 5. Second, it allows us to recover data from times when, due to scintillation, a pulsar was not bright enough for independent TOAs to be measured accurately in each frequency channel. Finally, it provides reduced sensitivity to frequency-dependent systematic errors which may be present (for example, see §3.5).

The measurement of DM(t) is a key feature of this analysis. DM was defined in §3.2 as the column density of free electrons between Earth and a pulsar. Radio waves travelling through this tenuous plasma are affected in various ways, most notably by a time delay versus radio frequency (RF) described by Equation 3.4. The DM towards each pulsar can vary versus time, as the line of sight moves relative to the intervening plasma (see Chapter 6). If this variation is not measured and corrected, it will appear as a drift in the pulse TOAs, corrupting timing model parameters and obscuring GW. Measurements of DM are most effective when data are taken at two (or more) widely spaced RFs, giving a large time-delay "lever arm" with which to pin down the DM. In our analysis, after a initial timing fit to determine the average DM and approximate timing model, we fit the timing residuals to Equation 3.4 for each epoch in which multi-band data exists. Since the time delay is linear in DM, this two-step fitting is a valid approach. The fit also determines the constant offset required to line up the two bands, as they have different average profile shapes.

The result of this procedure is a set of DM measurements and uncertainties at discrete times (see plots in §4.4). We take this time series and apply a linear prediction model (Rybicki and Press, 1992; Press et al., 1992) to both filter and interpolate it. Filtering is important since each DM point is affected by measurement uncertainty, therefore using the raw DM values can potentially add in more noise than it removes. For the measured data points, linear prediction reduces to Wiener filtering (see §2.4), an optimal filter for noisy data. It also provides a smoothly interpolated version of the signal, which we use to provide a DM correction for epochs where dual-frequency data does not exist. However, in some cases, unremoved DM variation has negatively affected our results. This is particularly true for the early GBT data, when dual-frequency observations were not regularly scheduled. Since early 2006, the scheduling has been much more favorable.

The data were also analyzed using the PCA approach described earlier. This resulted in several improvements. For one, the method offers a convenient way to identify corrupt data which might otherwise remain undetected. For example, a few folded profiles were found to contain low-amplitude time-shifted copies of the pulse shape, a signature of some form of clock error during the observation. Once identified, these data points can easily be cut. Aside from these errors, small profile variations were observed in several pulsars. The form of these variations and their effect on the timing measurements are discussed in the next section.

Keyword	Explanation	Units
RAJ	Right ascention of source (J2000)	hours, minutes, seconds
DECJ	Declination of source (J2000)	°, arcmin, arcsec
PMRA	Source proper motion along RA	mas yr^{-1}
PMDEC	Source proper motion along declination	mas yr^{-1}
PX	Annual parallax	mas
F0	Pulsar spin frequency	Hz
F1	Spin time derivative	Hz s^{-1}
F2	Spin second time derivative	Hz s^{-2}
PEPOCH	Reference time for variable position and spin parameters	MJD
EPHEM	Solar system ephemeris version	–
START	Start time of data used in fit	MJD
FINISH	End time of data used in fit	MJD
NTOA	Number of TOAs used in fit	–
TRES	RMS post-fit timing residual	μs
BINARY	Type of binary motion paramaterization (BT[a], DD[b] or ELL1[c])	–
PB	Orbital period	days
A1	Projected semi-major axis length	lt-sec
E	Orbital eccentricity	–
T0	Epoch of periastron	MJD
OM	Orbital longitude of periastron	°
EPS1[c]	First Laplace parameter	E sin(OM)
EPS2[c]	Second Laplace parameter	E cos(OM)
TASC[c]	Epoch of ascending node	MJD
M2	Mass of binary companion	$M_\odot$
SINI	Sine of orbit inclination angle	–

Table 4.3: List of timing model parameter keywords used by TEMPO, and the physical meaning and units of each. The lower section of the table lists binary model parameters.

[a] Newtonian orbital model (Blandford and Teukolsky, 1976).
[b] General relativistic orbital model (Damour and Deruelle, 1985). Includes all Newtonian parameters plus the Shapiro delay parameters M2 and SINI.
[c] Low-eccentricity paramaterization. The parameters EPS1, EPS2 and TASC are used in place of E, T0 and OM. This results in a more stable fit when E is very small.

The final analysis step is to fit the frequency-integrated TOAs to a timing model. The timing fit is a standard procedure in pulsar science and has been described many times in the literature (see for example Splaver et al. (2005)). It is a iterative χ^2 fit which gradually reduces the difference between the observed TOAs and a parameterized prediction of the apparent pulsar rotational phase versus time. The parameters fall into three main categories: Terms due to the motion of the Earth within the solar system, terms due to the intrinsic pulsar spin evolution, and terms due to the motion of the pulsar through its binary orbit. Since the main consumers of timing model parameter values are other pulsar astronomers, we have chosen to present these results in a format compatible with TEMPO. A guide to the timing model parameters is given in Table 4.3, which lists the TEMPO keywords and units used, as well as a brief description of each parameter.

4.4 Summary of Results

Tables 4.6 through 4.13 give the resulting timing model parameter values and associated errors. The formal fit error values given have been scaled by the fit reduced-χ value in order to represent the effect of systematic timing errors. However, the parameters themselves are presented with the full amount of precision returned by the fit, even when this is more precise than the errors suggest. This is so the exact timing model used here can be recreated in the future, but should not be taken as the true parameter uncertainty.

Source	$\bar{\chi}^2$	σ_t (μs)	N	$\bar{\sigma}_t$ (μs)	$\bar{N}$	$\bar{\sigma}_{t,prev}$ (μs)	$\bar{\chi}^2_{prev}$
J0030+0451	1.5	1.25	1181	0.43	40	–	–
J0218+4232	1.2	4.98	243	1.77	31	–	–
J0613−0200	1.4	0.76	470	0.32	38	0.62	1.5
J1012+5307	1.4	1.18	606	0.55	51	–	–
J1455−3330	1.3	3.77	368	1.57	38	–	–
J1640+2224	3.2	0.99	1232	0.85	41	–	–
J1643−1224	2.5	3.42	566	2.35	56	–	–
J1713+0747	7.2	0.25	1446	0.13	65	0.13	1.1
J1744−1134	3.9	0.92	490	0.83	48	0.90[a]	1.9
B1855+09	1.4	0.75	1265	0.29	45	3.30[a]	1.6
J1909−3744	4.3	0.38	396	0.30	35	0.15	1.2
J1918−0642	1.2	1.83	420	1.19	45	–	–
B1937+21	13.1	0.32	1250	0.24	85	0.14	2.9
J2019+2425	1.6	2.34	768	1.40	38	–	–
J2145−0750	2.7	1.34	325	1.16	34	1.30	4.1
J2317+1439	2.9	1.45	1733	0.50	46	–	–

Table 4.4: Summary of timing residuals for all pulsars. The listed values are the timing fit's reduced-χ^2 value, weighted RMS timing residual (σ_t), and daily-average RMS timing residual ($\bar{\sigma}_t$), defined in the text. The final two columns give the values presented by Hotan et al. (2006) for comparision. See the text for a discussion.

[a] RMS residual from 5 minute integration.

Table 4.4 lists the weighted RMS timing residual value observed for each pulsar, defined as:

$$\sigma_t^2 = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1} \sum_i \frac{r_i^2}{\sigma_i^2} \tag{4.2}$$

Here, r_i is the post-fit timing residual at each point, and σ_i is the associated measurement uncertainty. This number is a standard way of characterizing pulsar timing results. We present two forms, the first computed from all N data points used in the timing fit. The second is computed by first averaging all residuals from each day of observation in each band separately. This gives a reduced number of data points, $\bar{N}$. The weighted RMS residual $\bar{\sigma}_t$ is then computed from these points. The daily-average is a more appropriate value to consider when interested in long-term timing results, such as GW analysis. These numbers are comparable to the best results obtained by other current precision timing programs (for example Hotan et al., 2006), and for most pulsars we are meeting our goal of sub-μs timing.

However, we can also see from these numbers that the timing results are not limited simply by measurement noise. In that case, we expect to see $\bar{\sigma}_t \simeq \sqrt{\bar{N}/N}\sigma_t$, as the noise averages down. The fact that we are not achieving this $\sqrt{N}$ improvement after averaging suggests that there are systematic timing effects present at the $\sim$100 ns level on timescales greater than the individual observation durations (typically one hour). This is unfortunately a typical feature of pulsar datasets, and can have a variety of causes, including calibration errors, ISM effects (inaccurate DM measurement and scintillation-based shifts), and intrinsic pulsar timing noise. In Chapter 5 and Appendix A, we present a method for analyzing residuals for the presence of "red" timing noise, on timescales between the observation spacing (here, typically one month) and the length of the dataset. Our method correctly takes into account the effect of the timing fit on the statistics of the residuals, a consideration often neglected in previous analyses.

Seven of the pulsars included in this study were previously observed by Hotan et al. (2006) using the 64-m Parkes telescope in Australia. Their observing pattern was similar to ours: Each pulsar was observed for about one hour at several different bands once per month. They were also able to obtain several more intense observing sessions more suited for measuring short-period binary system parameters. Their results are summarized alongside ours in Table 4.4, where we present their RMS residual from 60-minute averages[6] (comparable to our $\bar{\sigma}_t$), and $\bar{\chi}^2$. Their numbers are generally comparable to ours, with similar RMS residual but lower $\bar{\chi}^2$. This may be expected due to the smaller telescope (more measurement noise in each data point). For J0613−0200 and B1855+09 we achieve a similar $\bar{\chi}^2$ but factors of 2 and 4 better $\bar{\sigma}_t$, indicating that these pulsars are still mostly signal-to-noise limited at this point. For J1713+0747, J1744−1134 and B1937+21, we achieve similar $\bar{\sigma}_t$, but higher $\bar{\chi}^2$. This suggests that systematic effects rather than measurement noise is the limiting factor, and more work is needed to characterize and, if possible, remove these effects. Our results for J1909−3744 are somewhat worse, most likely due to

[6]No true 60-minute RMS residual was given for B1855+09 and J1744−1134. The 5-minute averages presented here can be compared to our σ_t.

our sparse orbital coverage (see below for more on this). We would benefit greatly from the inclusion of a more intense observing "campaign" on this pulsar. Finally, J2145$-$0750 goes against the general trend. Here we achieve a slightly better $\bar{\sigma}_t$, along with a much lower $\bar{\chi}^2$. This may be due to our compensation for systematic profile variation (see §4.4.1). This pulsar's long spin period (16 ms) makes its timing much more susceptible to these sorts of errors.

Developing ways of characterizing and mitigating systematic timing effects is an active area of pulsar research. This will become more important as advances in radio telescope technology, such as larger collecting areas and higher accessible bandwidths, continue to improve achievable signal-to-noise ratios. Long-term datasets such as this one are not the best tool with which to investigate these effects, due to the large jump between the individual observation lengths and the observation spacing. Observations that focus on the "best" pulsars (J1909$-$3744, J1713+0747 and B1937+21) more intensely, but for shorter periods of time would be very complementary to our long-term data. Indeed, some of this type of observation has already been done for some pulsars, usually in the context of accurately measuring binary parameters. This is another area where short-term observations are beneficial, as short-period binaries such as J1909$-$3744 move through several orbits between our 800 MHz and 1400 MHz data points. This makes it difficult to distinguish DM variation from binary evolution. Combining existing short-term datasets and proposing complementary short-term observations to be done along with our long-term monitoring will be a subject of future work.

4.4.1 Profile Variation

We applied the PCA based method described earlier to investigate profile variation in our dataset. Aside from identifying data corruption, this uncovered little systematic profile change. The most common variation seen was a component proportional to the profile second derivative, indicating a fluctuating pulse width. This is most likely caused by inaccuracies in the online folding as described in §4.3, or in the frequency-integration step of our analysis. In most cases, this was uncorrelated with the timing residual. For J1909−3744, we were able to apply a timing correction as described in §2.5, reducing the final $\bar{\sigma}_t$ from 0.36 to 0.30 μs. One exception to this description was J2145−0750, which displays a different profile shape variation, shown in Figure 4.1. This variation is probably caused by the quadrature downconversion error discussed in §3.5.2, as it appears only in the early part of the data span. Applying a PCA-based correction reduces $\bar{\sigma}_t$ from 1.6 to 1.2 μs.

4.4.2 Dispersion Measure Variation

Dispersion measure variation and its interpretation is discussed in detail in Chapter 6, using 20 years of data on B1937+21. It is worth taking a brief look at the DMs measured here from a larger set of pulsars. The RMS variation in DM (σ_{DM}), and the variation expected due to the DM measurement uncertainty, is listed in Table 4.5

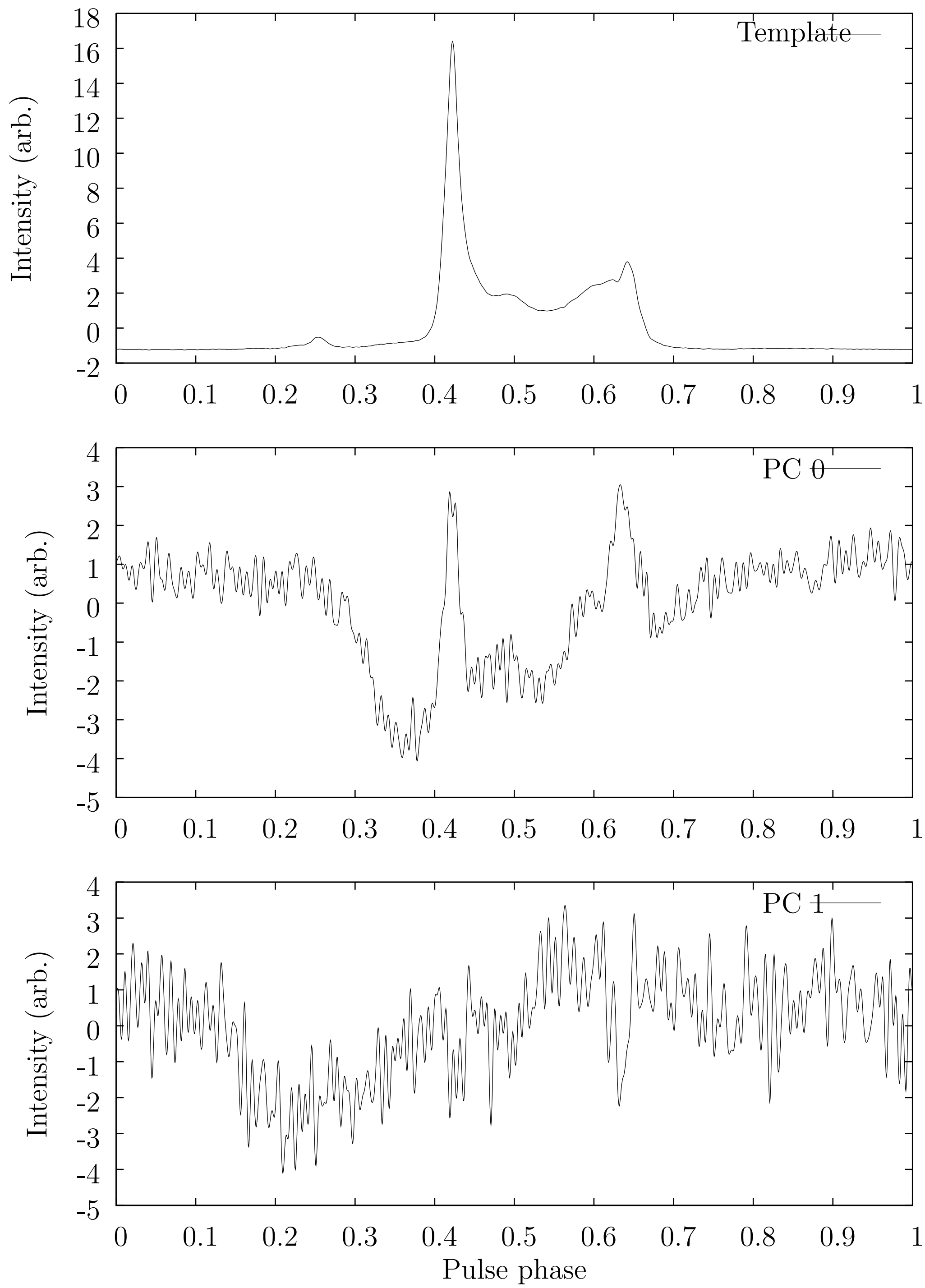

Figure 4.1: Template profile and first two principal components for J2145−0750 at 1400 MHz.

for each pulsar. σ_{DM} has been corrected in quadrature for the error contribution. This value is plotted versus DM in Figure 4.2, where we can see that the DM variation tends to increase weakly with DM, although with fairly large scatter. This behavior was previously observed by Backer et al. (1993), who proposed a simplistic model with DM variations scaling as $\sigma_{\rm DM} \propto {\rm DM}^{1/2}$. This is essentially a statistical statement that since DM is approximately proportional to distance from Earth, a high-DM pulsar will have more independent ISM components along its line of sight. These contributions will add in a "random-walk" fashion, and the RMS variation will grow as the square root. Scatter about this line is expected for several reasons. Low-DM pulsars will be affected by "low-number statistics" due to their smaller distances. Also, since DM(t) can be thought of as a stochastic process with a red power spectrum (see Chapter 6), any particular finite span of data can show different statistics. For example, over the $\sim$2.5 years presented here, B1937+21 has a relatively low σ_{DM}. However, the $\sim$20 years of data on this pulsar presented in Chapter 6 reveals much larger DM fluctuations, including other 2.5 year chunks where σ_{DM} was higher.

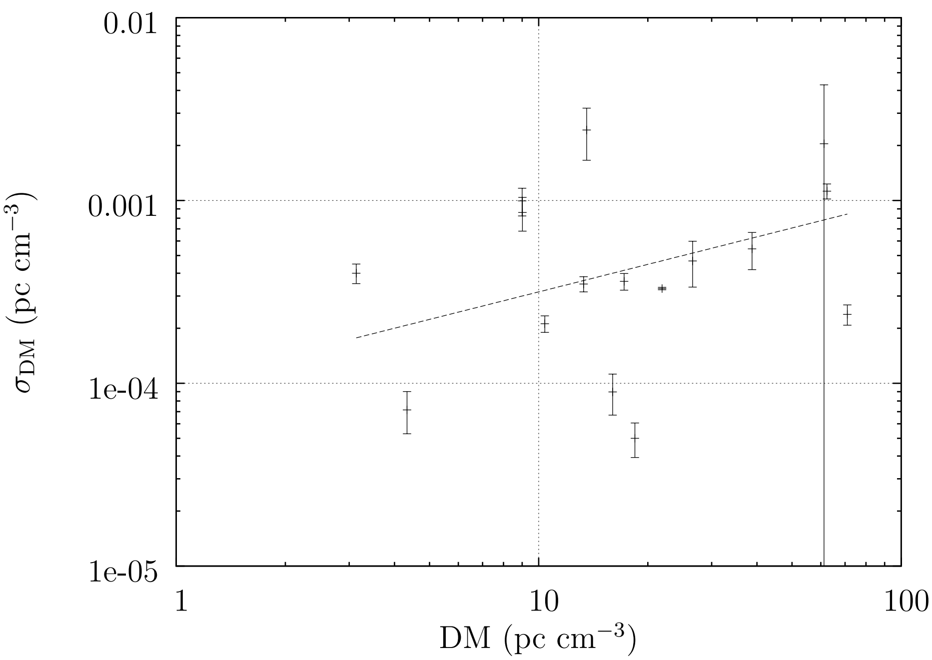

Figure 4.2: Dispersion measure variation as a function of DM. The line shows the expected $\mathrm{DM}^{1/2}$ scaling.

Source	DM (pc cm^{-3})	σ_{DM} (pc cm^{-3})	Uncertainty (pc cm^{-3})	N_{pts}
J0030+0451	4.3330	7.15×10^{-5}	1.9×10^{-5}	17
J0218+4232	61.2452	2.04×10^{-3}	2.3×10^{-3}	9
J0613−0200	38.7787	5.44×10^{-4}	1.2×10^{-4}	14
J1012+5307	9.0211	8.60×10^{-4}	1.8×10^{-4}	14
J1455−3330	13.5670	2.43×10^{-3}	7.7×10^{-4}	6
J1640+2224	18.4273	5.00×10^{-5}	1.1×10^{-5}	13
J1643−1224	62.4096	1.12×10^{-3}	1.1×10^{-4}	18
J1713+0747	15.9915	8.96×10^{-5}	2.3×10^{-5}	17
J1744−1134	3.1388	4.00×10^{-4}	4.9×10^{-5}	13
B1855+09	13.2985	3.49×10^{-4}	3.3×10^{-5}	22
J1909−3744	10.3933	2.02×10^{-4}	2.2×10^{-5}	8
J1918−0642	26.5939	4.67×10^{-4}	1.3×10^{-4}	13
B1937+21	71.0253	2.38×10^{-4}	3.0×10^{-5}	8
J2019+2425	17.1991	3.61×10^{-4}	3.8×10^{-5}	14
J2145−0750	9.0028	5.46×10^{-4}	2.2×10^{-4}	7
J2317+1439	21.9026	3.30×10^{-4}	4.9×10^{-6}	23

Table 4.5: Observed dispersion measure variation in the set of pulsars.

	J0030+0451		B1937+21	
Parameter	Value	Error	Value	Error
RAJ	00:30:27.42871853	7.0×10^{-4}	19:39:38.56125423	7.0×10^{-6}
DECJ	04:51:39.8465216	2.4×10^{-2}	21:34:59.1358485	1.3×10^{-4}
PMRA	1.2781	1.4	0.0499	5.8×10^{-3}
PMDEC	-16.3619	3.2	-0.4283	7.2×10^{-3}
PX	2.7694	1.9×10^{-1}	–	–
F0	205.53069927434390	2.7×10^{-11}	641.92826260109052	1.4×10^{-11}
F1	$-4.297025616\times10^{-16}$	1.1×10^{-19}	$-4.331108608\times10^{-14}$	2.6×10^{-20}
PEPOCH	50984.400000	–	47500.000000	–
EPHEM	DE405	–	DE405	–
START	53358.945	–	53341.797	–
FINISH	54135.866	–	54177.548	–
NTOA	1181	–	727	–
TRES	1.25	–	0.16	–

Table 4.6: Timing model parameters for J0030+0451, B1937+21.

	J1713+0747		B1855+09	
Parameter	Value	Error	Value	Error
RAJ	17:13:49.53052763	2.8×10^{-6}	18:57:36.39509190	9.0×10^{-5}
DECJ	07:47:37.5264253	9.8×10^{-5}	09:43:17.3350620	2.3×10^{-3}
PMRA	4.9359	8.6×10^{-3}	-2.8036	6.7×10^{-2}
PMDEC	-3.9090	2.0×10^{-2}	-5.3508	1.2×10^{-1}
PX	1.2032	4.0×10^{-2}	–	–
F0	218.81184391575215	1.2×10^{-12}	186.49408172864679	4.9×10^{-11}
F1	$-4.084993389\times10^{-16}$	7.7×10^{-21}	$-6.204359653\times10^{-16}$	7.7×10^{-20}
PEPOCH	52000.000000	–	46436.699400	–
EPHEM	DE405	–	DE405	–
START	53350.669	–	53358.722	–
FINISH	54177.458	–	54177.519	–
NTOA	1202	–	1265	–
TRES	0.25	–	0.77	–
BINARY	DD[a]	–	DD	–
PB	67.825129871787[a]	–	12.327171192294	1.9×10^{-9}
A1	32.342420993[a]	–	9.230779275	5.5×10^{-7}
E	0.0000749407[a]	–	0.0000217596	3.7×10^{-8}
T0	51997.578423049[a]	–	46432.777258715	3.3×10^{-3}
OM	176.191545579029[a]	–	276.361737017325	9.7×10^{-2}
M2	0.280000[a]	–	0.308421	2.7×10^{-2}
SINI	0.952352[a]	–	0.996509	1.6×10^{-3}

Table 4.7: Timing model parameters for J1713+0747, B1855+09.

[a] Binary parameters from Splaver et al. (2005).

	J1640+2224		J2019+2425	
Parameter	Value	Error	Value	Error
RAJ	16:40:16.74333189	1.8×10^{-5}	20:19:31.94315509	5.1×10^{-6}
DECJ	22:24:08.9432740	3.3×10^{-4}	24:25:15.0850505	1.1×10^{-4}
PMRA	2.6397	4.5×10^{-2}	-9.8121	1.2×10^{-1}
PMDEC	-10.9918	6.1×10^{-2}	-20.5696	1.6×10^{-1}
PX	–	–	–	–
F0	316.12398431370713	1.2×10^{-11}	254.16034544479498	1.5×10^{-12}
F1	$-2.816246993\times10^{-16}$	6.9×10^{-20}	$-4.548218076\times10^{-16}$	1.6×10^{-19}
PEPOCH	51700.000000	–	53768.196500	–
EPHEM	DE405	–	DE405	–
START	53343.622	–	53358.798	–
FINISH	54177.381	–	54177.595	–
NTOA	1232	–	768	–
TRES	0.99	–	2.34	–
BINARY	BT	–	BT	–
PB	175.460661852080	2.5×10^{-8}	76.511634718470	1.5×10^{-8}
A1	55.329722783	6.3×10^{-8}	38.767644418	1.4×10^{-7}
E	0.0007973279	2.2×10^{-9}	0.0001110910	8.0×10^{-9}
T0	51626.180616800	7.2×10^{-5}	50054.645899837	1.2×10^{-3}
OM	50.735106641882	1.5×10^{-4}	159.039383932102	5.7×10^{-3}

Table 4.8: Timing model parameters for J1640+2224, J2019+2425.

	J0218+4232		J2317+1439	
Parameter	Value	Error	Value	Error
RAJ	02:18:06.35593648	2.9×10^{-5}	23:17:09.23701054	3.2×10^{-6}
DECJ	42:32:17.3912269	7.1×10^{-4}	14:39:31.2449395	1.0×10^{-4}
PMRA	–	–	-1.3841	9.1×10^{-2}
PMDEC	–	–	4.3841	1.7×10^{-1}
PX	–	–	–	–
F0	430.46106289245108	1.7×10^{-11}	290.25460818711076	7.4×10^{-13}
F1	$-1.433643303\times10^{-14}$	9.4×10^{-19}	$-2.053981116\times10^{-16}$	1.1×10^{-19}
PEPOCH	53650.408500	–	53768.261500	–
EPHEM	DE405	–	DE405	–
START	53348.028	–	53358.854	–
FINISH	54185.150	–	54177.669	–
NTOA	243	–	1733	–
TRES	4.99	–	1.45	–
BINARY	ELL1	–	ELL1	–
PB	2.028846086573	1.1×10^{-9}	2.459331463613	1.3×10^{-10}
A1	1.984439356	8.2×10^{-7}	2.313948370	9.7×10^{-8}
EPS1	0.0000062166	6.4×10^{-7}	0.0000004888	8.1×10^{-8}
EPS2	0.0000030616	1.1×10^{-6}	-0.0000000450	6.2×10^{-8}
TASC	54001.579811266	1.7×10^{-7}	54000.254767017	1.9×10^{-8}

Table 4.9: Timing model parameters for J0218+4232, J2317+1439.

	J0613−0200		J1012+5307	
Parameter	Value	Error	Value	Error
RAJ	06:13:43.97528245	2.4×10^{-6}	10:12:33.43633434	8.0×10^{-6}
DECJ	-02:00:47.1895773	1.0×10^{-4}	53:07:02.3990763	7.8×10^{-5}
PMRA	2.0144	4.5×10^{-2}	2.6206	1.1×10^{-1}
PMDEC	-8.9753	1.4×10^{-1}	-25.9941	1.3×10^{-1}
PX	–	–	–	–
F0	326.60056720791317	1.6×10^{-12}	190.26783746758178	1.2×10^{-12}
F1	$-1.022187240\times10^{-15}$	1.1×10^{-19}	$-6.201647564\times10^{-16}$	7.2×10^{-20}
PEPOCH	53637.470000	–	53583.426500	–
EPHEM	DE405	–	DE405	–
START	53348.050	–	53214.010	–
FINISH	54185.177	–	54185.206	–
NTOA	470	–	606	–
TRES	0.77	–	1.18	–
BINARY	ELL1	–	ELL1	–
PB	1.198512556698	1.1×10^{-10}	0.604672713561	3.9×10^{-11}
A1	1.091449607	7.5×10^{-8}	0.581828078	1.2×10^{-7}
EPS1	0.0000042477	1.6×10^{-7}	0.0000021503	3.1×10^{-7}
EPS2	0.0000035199	1.3×10^{-7}	0.0000007075	4.7×10^{-7}
TASC	50315.269379722	3.2×10^{-7}	50700.081629232	2.0×10^{-7}

Table 4.10: Timing model parameters for J0613−0200, J1012+5307.

	J1455−3330		J1643−1224	
Parameter	Value	Error	Value	Error
RAJ	14:55:47.96748128	3.4×10^{-5}	16:43:38.15995291	9.6×10^{-6}
DECJ	-33:30:46.3732120	1.0×10^{-3}	-12:24:58.6890462	6.8×10^{-4}
PMRA	9.2662	7.1×10^{-1}	6.4309	1.6×10^{-1}
PMDEC	-0.0579	1.3	4.6779	8.7×10^{-1}
PX	–	–	2.5801	3.9×10^{-1}
F0	125.20024523287074	5.9×10^{-12}	216.37334060332253	2.2×10^{-12}
F1	$-3.809709019\times10^{-16}$	2.6×10^{-19}	$-8.656396699\times10^{-16}$	1.6×10^{-19}
PEPOCH	53583.477000	–	53583.503000	–
EPHEM	DE405	–	DE405	–
START	53214.079	–	53214.096	–
FINISH	54184.249	–	54184.280	–
NTOA	368	–	566	–
TRES	3.77	–	3.42	–
BINARY	ELL1	–	ELL1	–
PB	76.174567299301	7.2×10^{-8}	147.017395154750	1.3×10^{-7}
A1	32.362213498	6.6×10^{-7}	25.072592180	2.2×10^{-7}
EPS1	-0.0001166549	3.3×10^{-8}	-0.0003124988	1.9×10^{-8}
EPS2	-0.0001231841	3.8×10^{-8}	0.0003977770	1.8×10^{-8}
TASC	53884.196627028	2.4×10^{-7}	50181.618216500	3.1×10^{-6}

Table 4.11: Timing model parameters for J1455−3330, J1643−1224.

	J1909−3744		J1918−0642	
Parameter	Value	Error	Value	Error
RAJ	19:09:47.43791729	6.2×10^{-6}	19:18:48.03505002	1.0×10^{-5}
DECJ	-37:44:14.3196382	1.9×10^{-4}	-06:42:34.8648864	3.1×10^{-4}
PMRA	-9.2762	3.0×10^{-2}	-6.9782	1.5×10^{-1}
PMDEC	-35.5232	1.0×10^{-1}	-5.6279	6.1×10^{-1}
PX	1.2808	1.1×10^{-1}	2.5663	4.3×10^{-1}
F0	339.31569275865837	4.5×10^{-12}	130.78951620506928	1.6×10^{-12}
F1	$-1.614333395\times10^{-15}$	7.6×10^{-20}	$-4.395355764\times10^{-16}$	1.9×10^{-19}
PEPOCH	53000.000000	–	53583.683500	–
EPHEM	DE405	–	DE405	–
START	53214.231	–	53214.325	–
FINISH	54184.446	–	54184.415	–
NTOA	396	–	420	–
TRES	0.38	–	1.83	–
BINARY	ELL1	–	ELL1	–
PB	1.533449450812	2.9×10^{-11}	10.913177577432	1.8×10^{-9}
A1	1.897997101	1.4×10^{-7}	8.350470992	2.1×10^{-7}
EPS1	0.0000001832	8.3×10^{-8}	-0.0000125543	5.0×10^{-8}
EPS2	-0.0000000077	4.1×10^{-8}	-0.0000156171	5.8×10^{-8}
TASC	53113.950587355	1.2×10^{-8}	51569.117363352	3.8×10^{-7}
M2	0.208349	7.7×10^{-3}	–	–
SINI	0.998141	2.8×10^{-4}	–	–

Table 4.12: Timing model parameters for J1909−3744, J1918−0642.

	J2145−0750	
Parameter	Value	Error
RAJ	21:45:50.46304483	3.5×10^{-5}
DECJ	-07:50:18.4486848	1.3×10^{-3}
PMRA	-6.6344	6.7×10^{-1}
PMDEC	-18.2116	2.1
PX	2.2852	4.7×10^{-1}
F0	62.29588881744527	1.0×10^{-12}
F1	$-1.151129253\times10^{-16}$	1.2×10^{-19}
PEPOCH	53583.730000	–
EPHEM	DE405	–
START	53219.378	–
FINISH	54184.456	–
NTOA	325	–
TRES	1.34	–
BINARY	ELL1	–
PB	6.838902508281	5.4×10^{-10}
A1	10.164110251	3.0×10^{-7}
EPS1	-0.0000067802	5.9×10^{-8}
EPS2	-0.0000179971	4.7×10^{-8}
TASC	50309.897130470	2.7×10^{-7}

Table 4.13: Timing model parameters for J2145−0750.

4.4.3 Plots

The rest of the chapter is devoted to plots of the data used for this analysis. The plots are organized as follows: First, template profiles are shown for all pulsars, then DM versus time, and finally timing residual versus time. The template plots have been normalized so that the fundamental Fourier component has amplitude 1 and phase 0.5. This lines up the pulse in the center of the plot for easy viewing. The amplitude (not square) of the Fourier transform of the profile is shown below each template. The DM plots show the measured values along with the interpolation function. The timing residual plots show both the full data set in the top panel, and the residuals averaged to 1 point per band per day in the lower panel. Each band is shown with a different symbol in these plots.

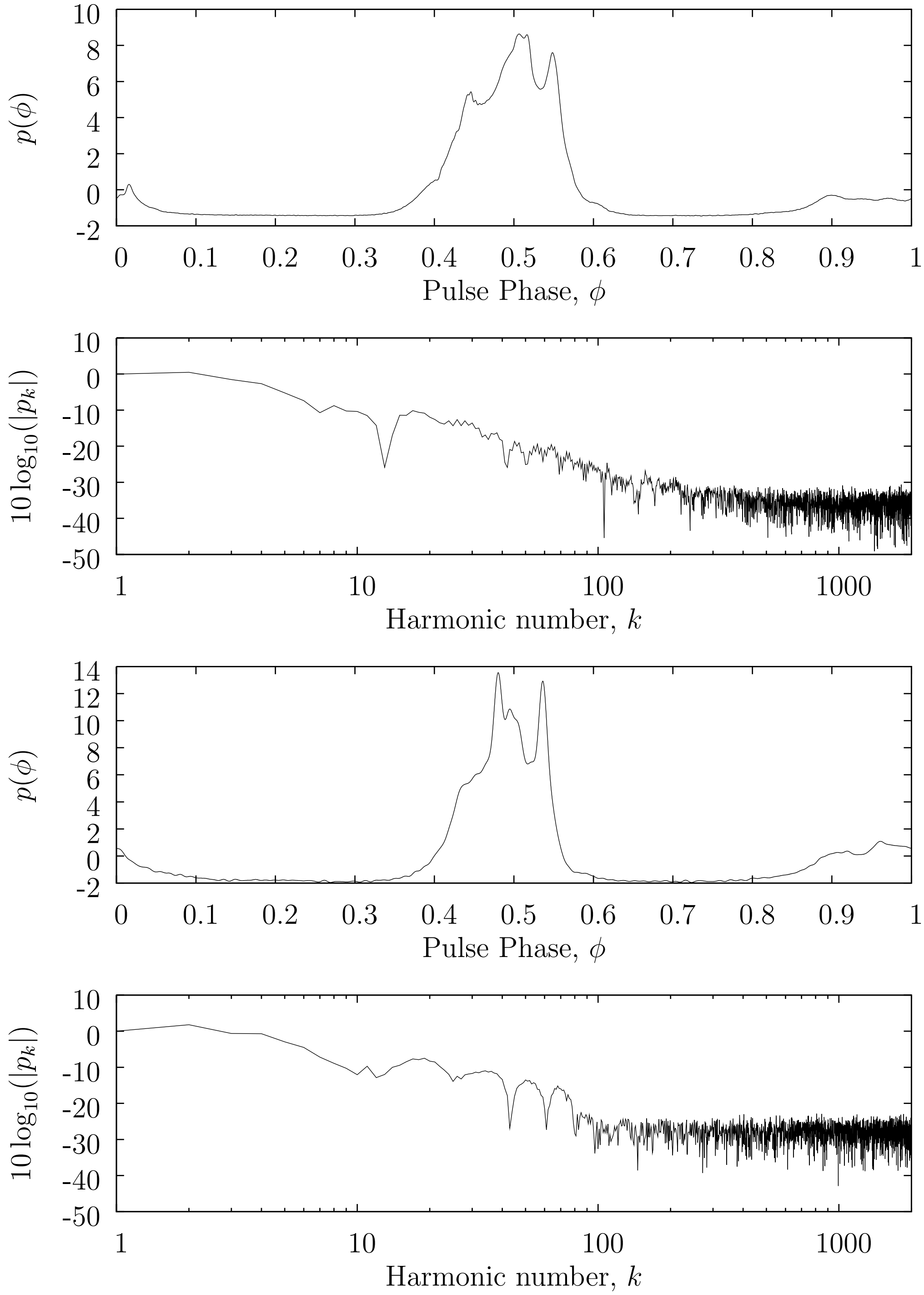

Figure 4.3: Template profiles and power spectra at 430 MHz (top) and 1410 MHz (bottom) for PSR J0030+0451.

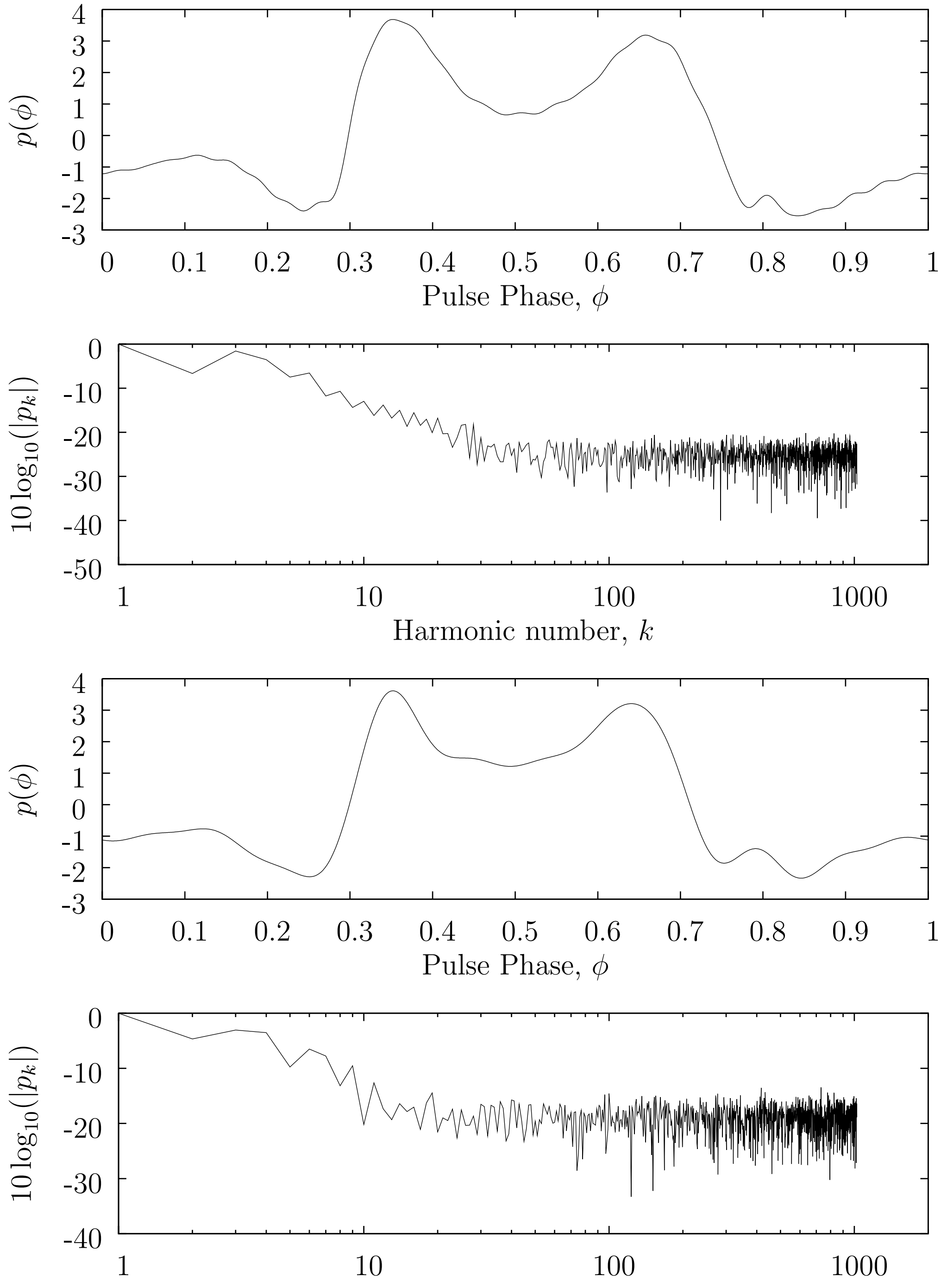

Figure 4.4: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J0218+4232.

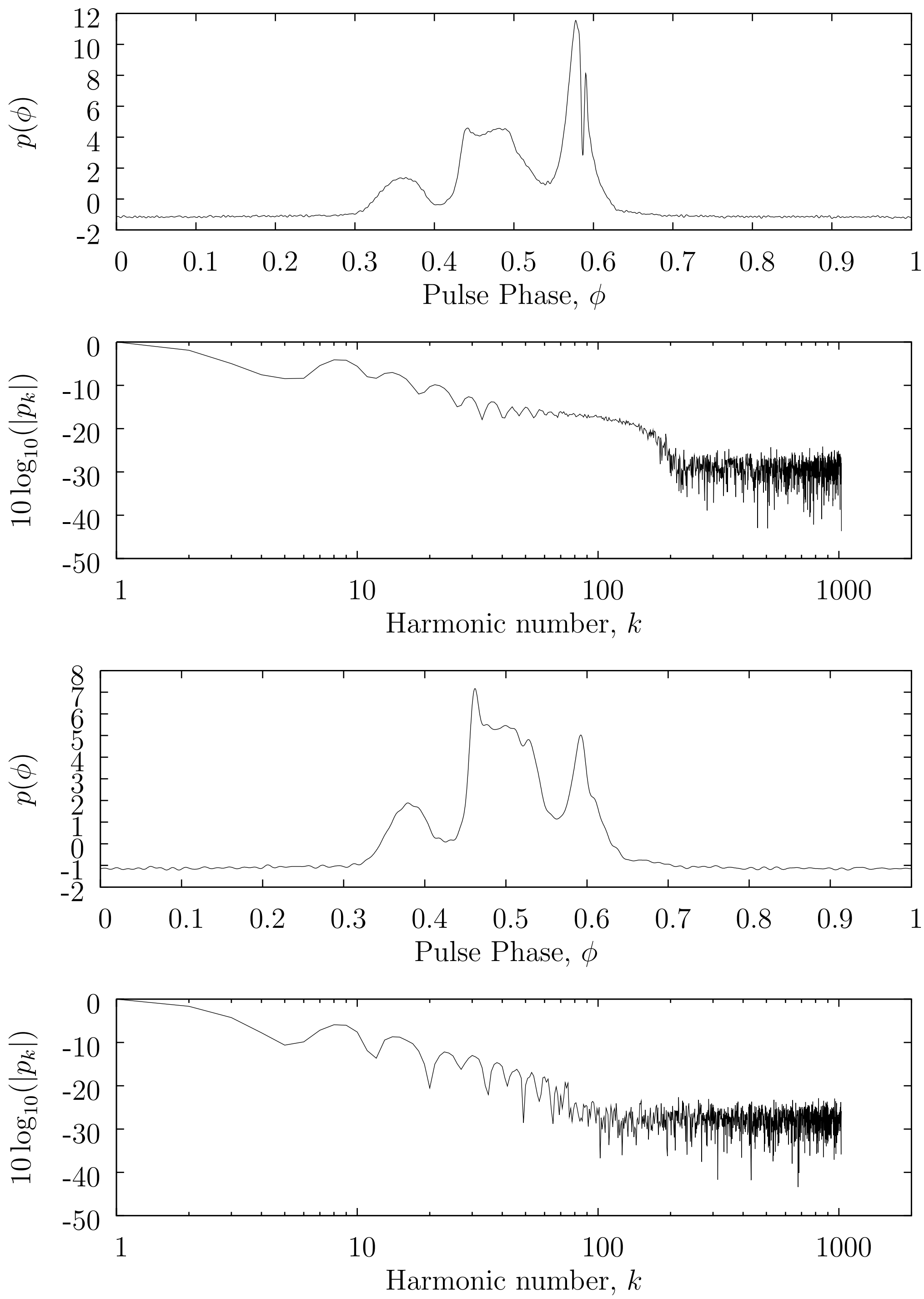

Figure 4.5: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J0613−0200.

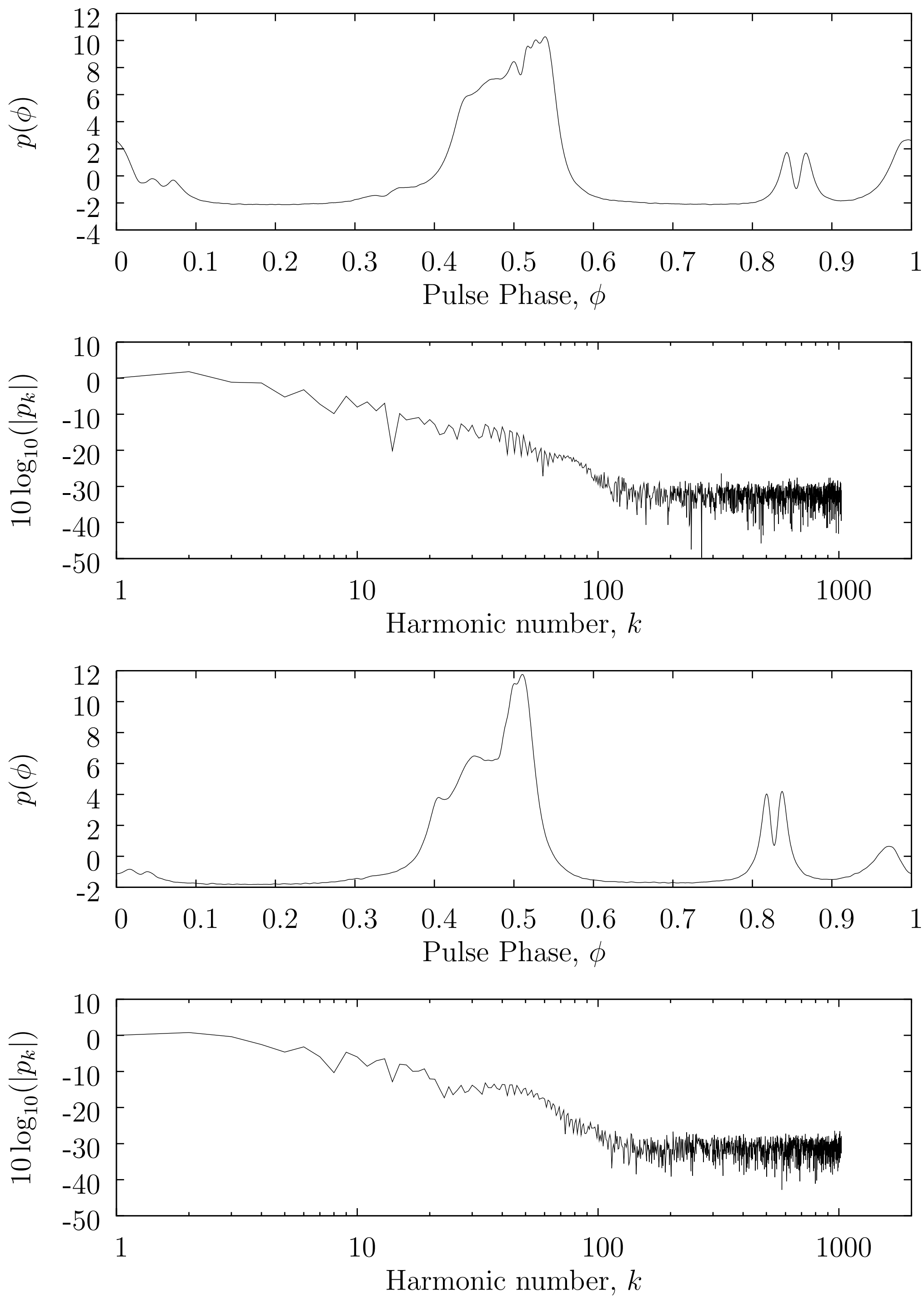

Figure 4.6: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J1012+5307.

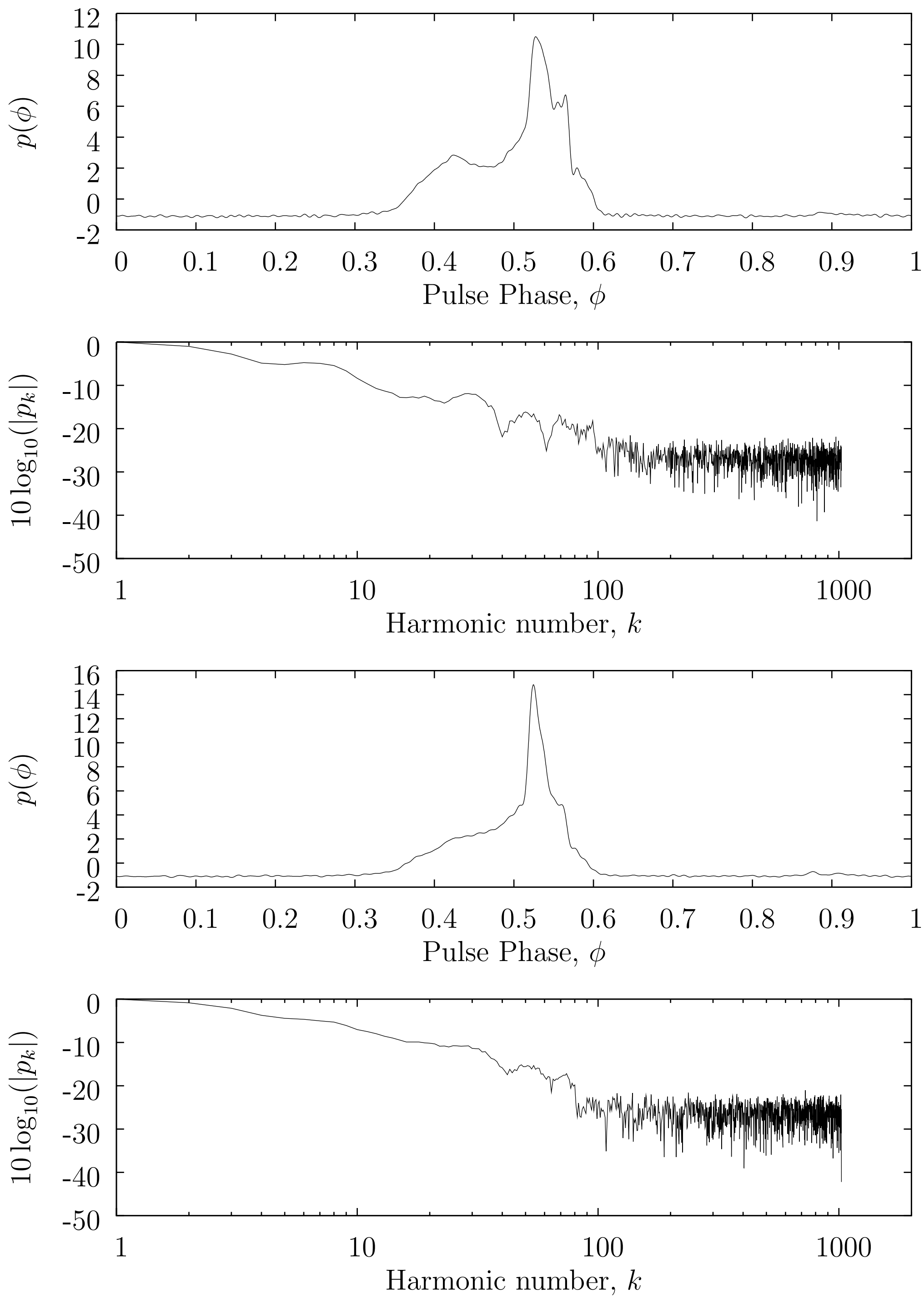

Figure 4.7: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J1455−3330.

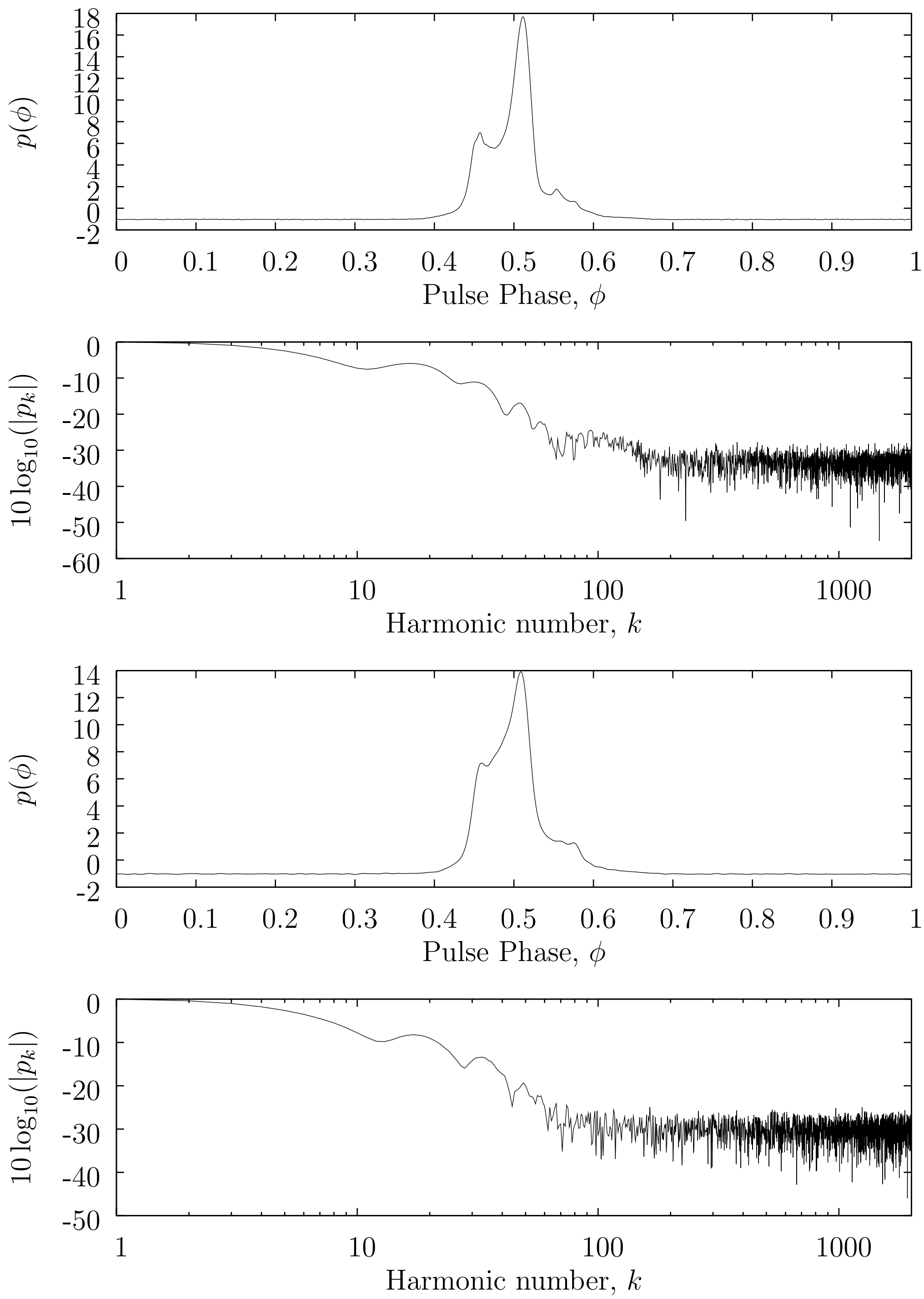

Figure 4.8: Template profiles and power spectra at 430 MHz (top) and 1410 MHz (bottom) for PSR J1640+2224.

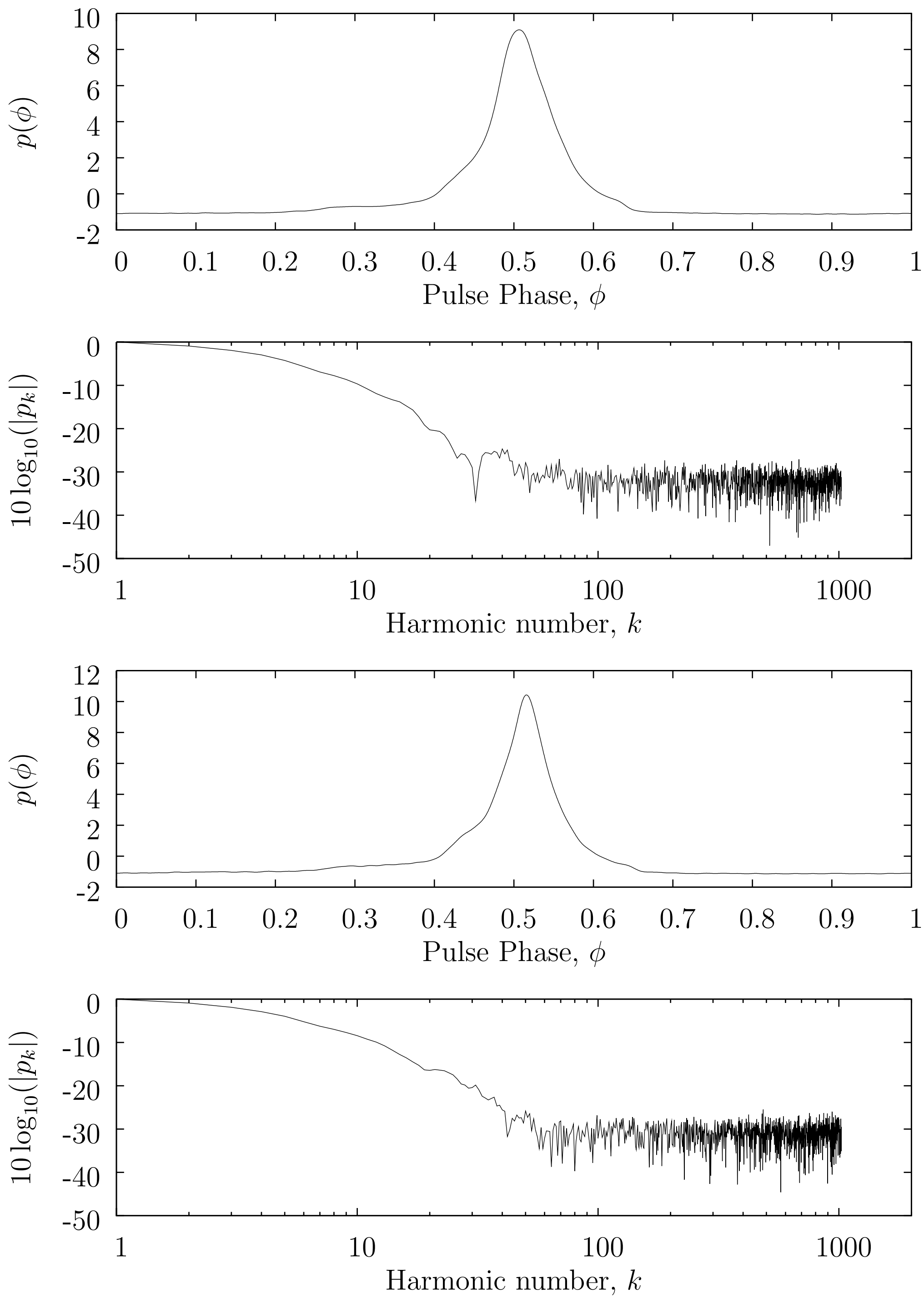

Figure 4.9: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J1643−1224.

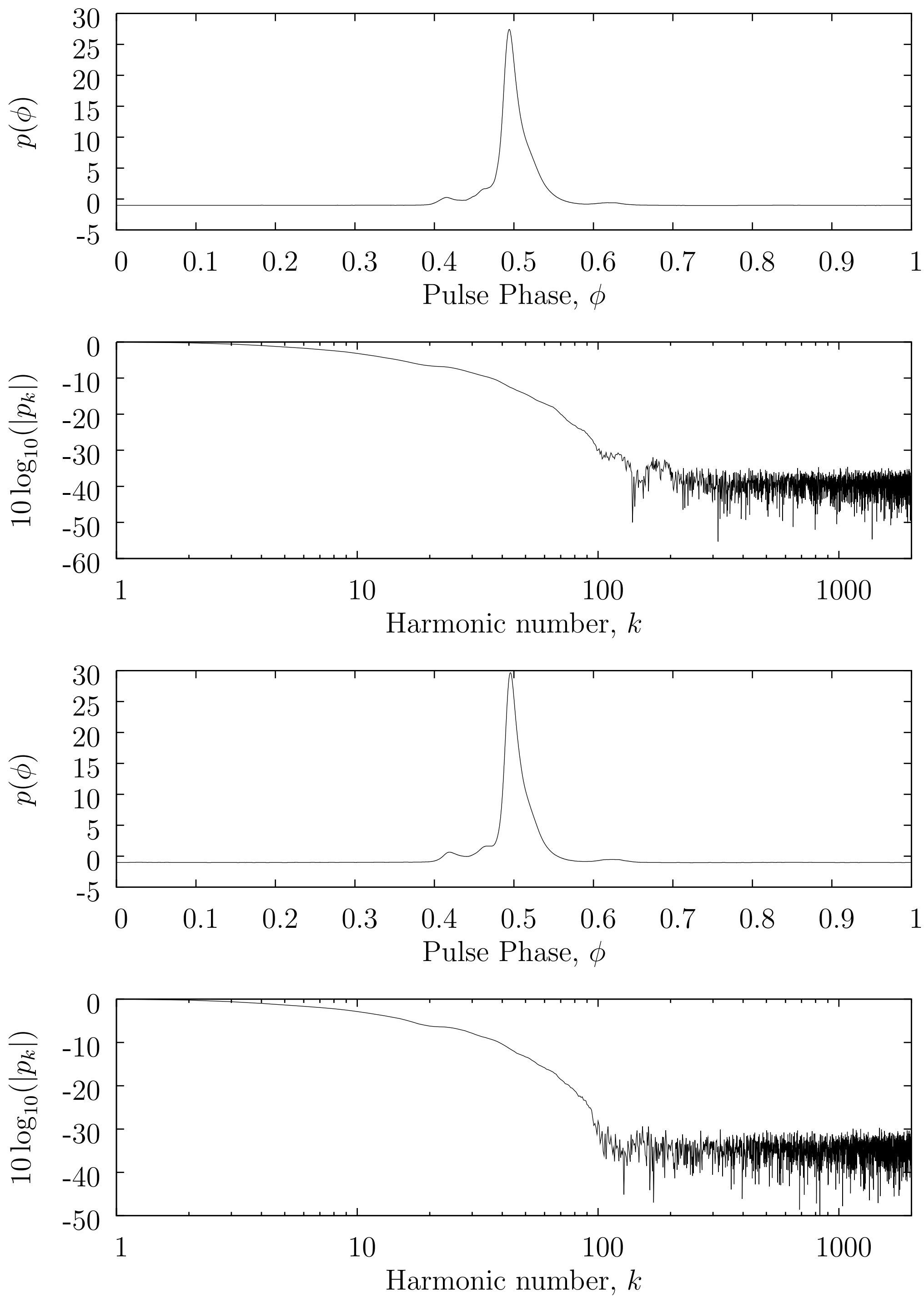

Figure 4.10: Template profiles and power spectra at 1410 MHz (top) and 2380 MHz (bottom) for PSR J1713+0747.

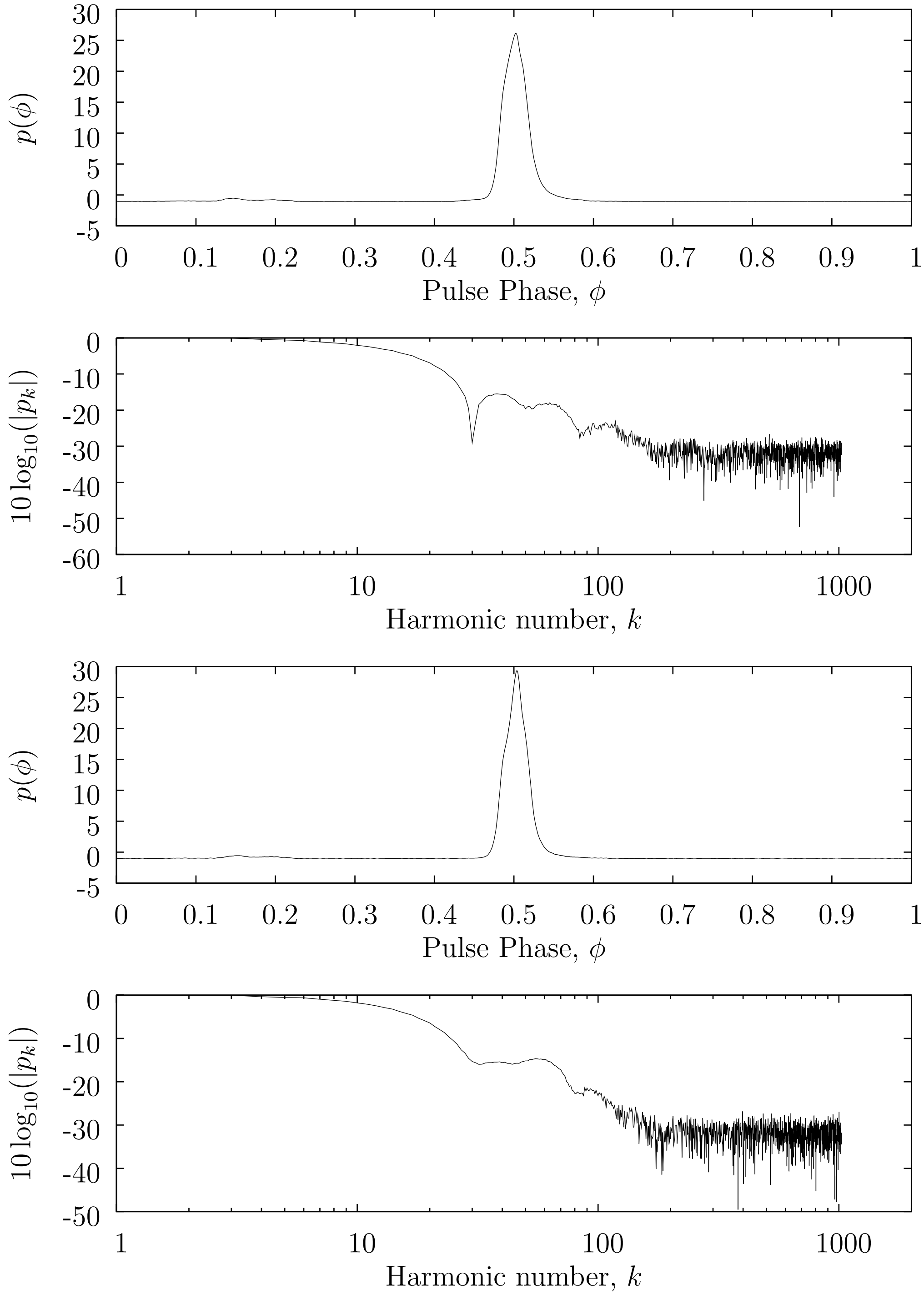

Figure 4.11: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J1744−1134.

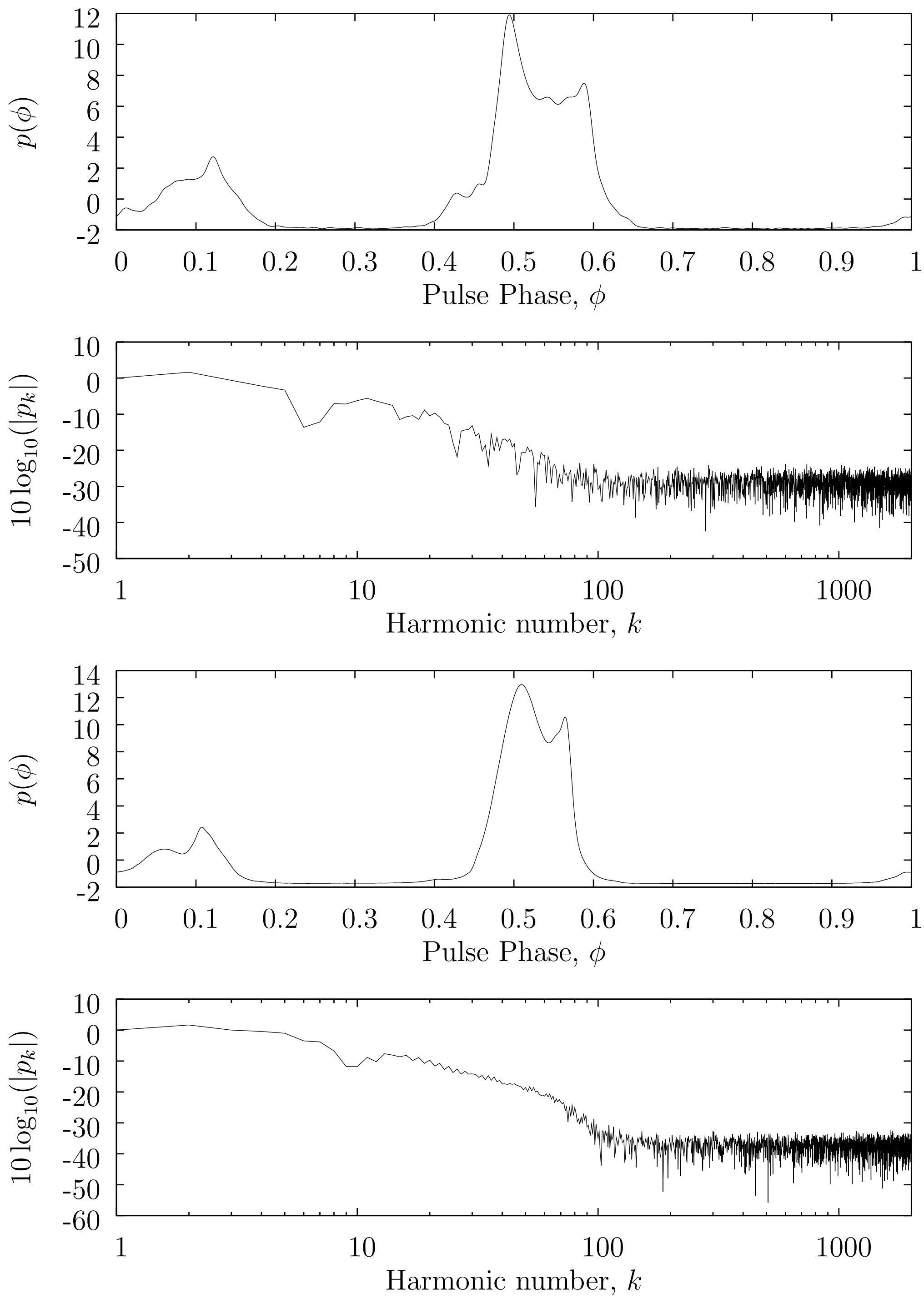

Figure 4.12: Template profiles and power spectra at 430 MHz (top) and 1410 MHz (bottom) for PSR B1855+09.

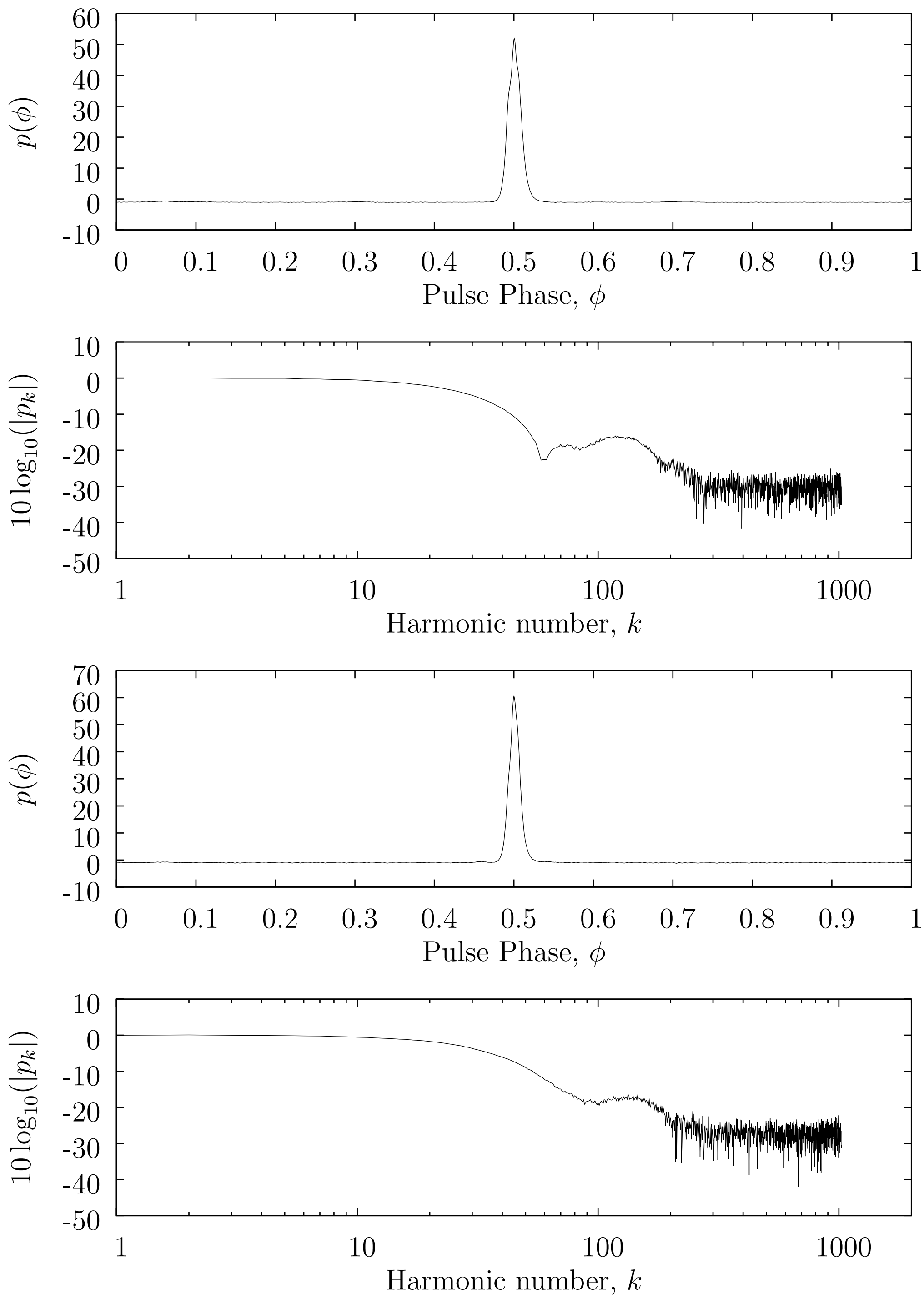

Figure 4.13: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J1909−3744.

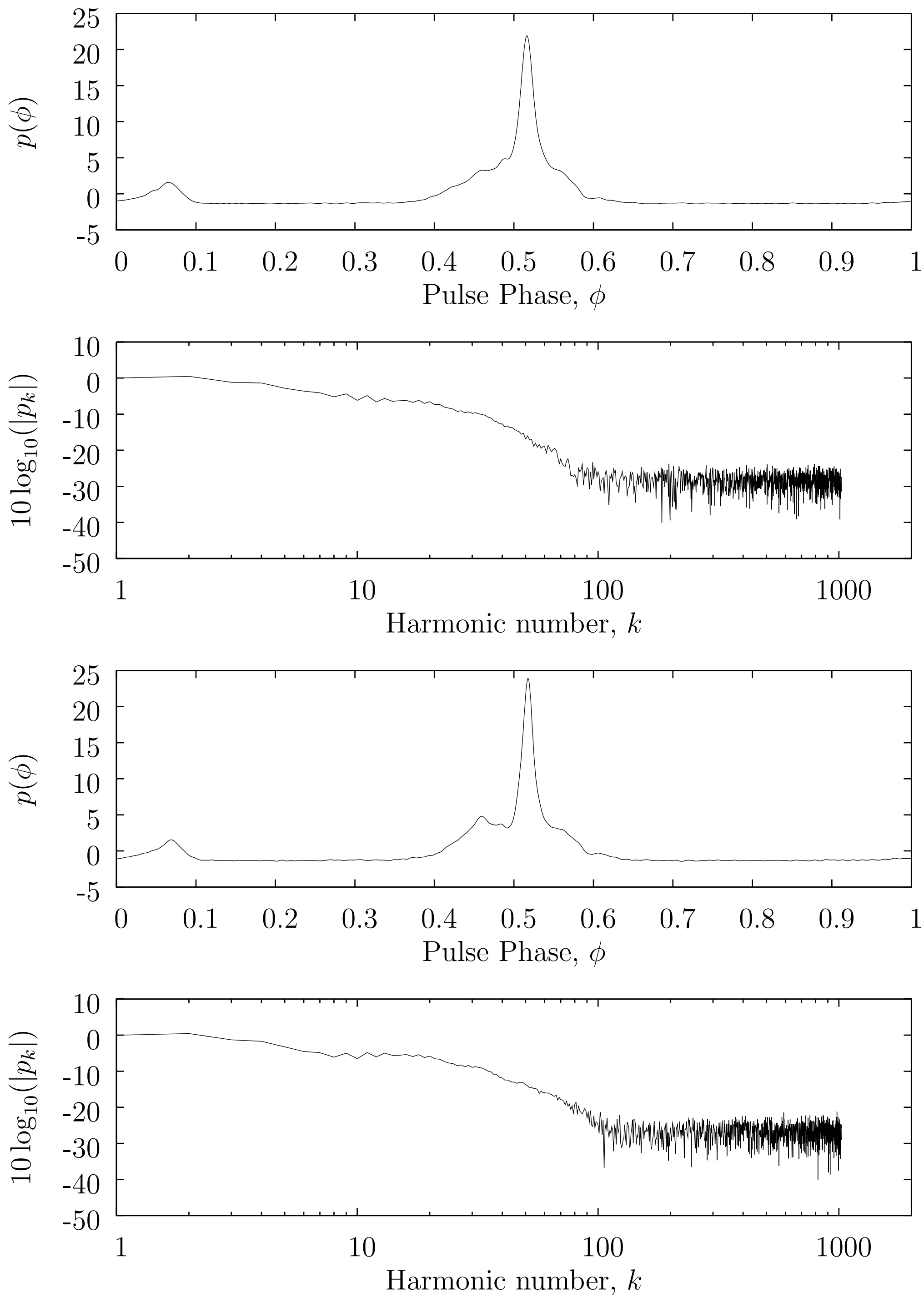

Figure 4.14: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J1918−0642.

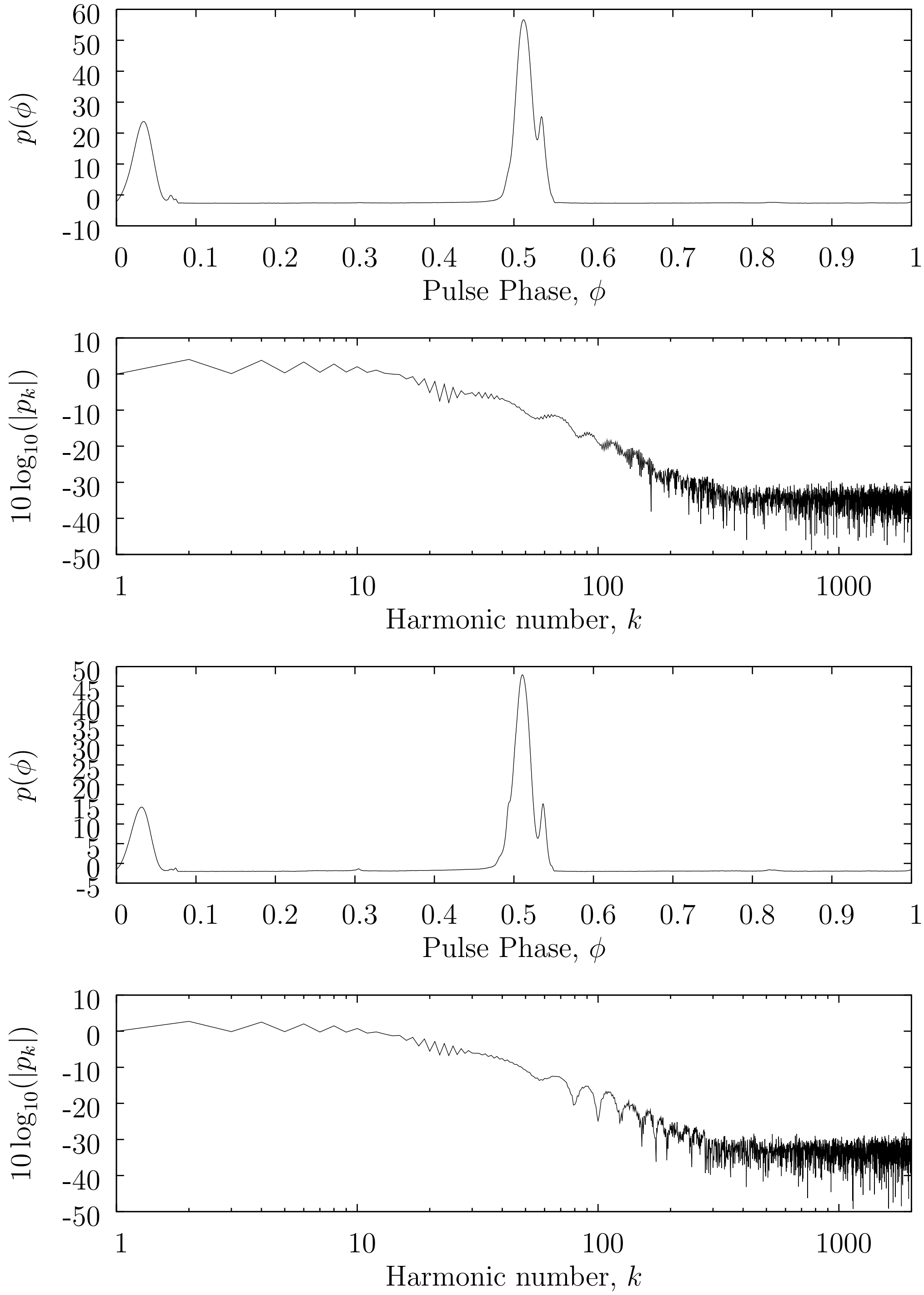

Figure 4.15: Template profiles and power spectra at 1410 MHz (top) and 2380 MHz (bottom) for PSR B1937+21.

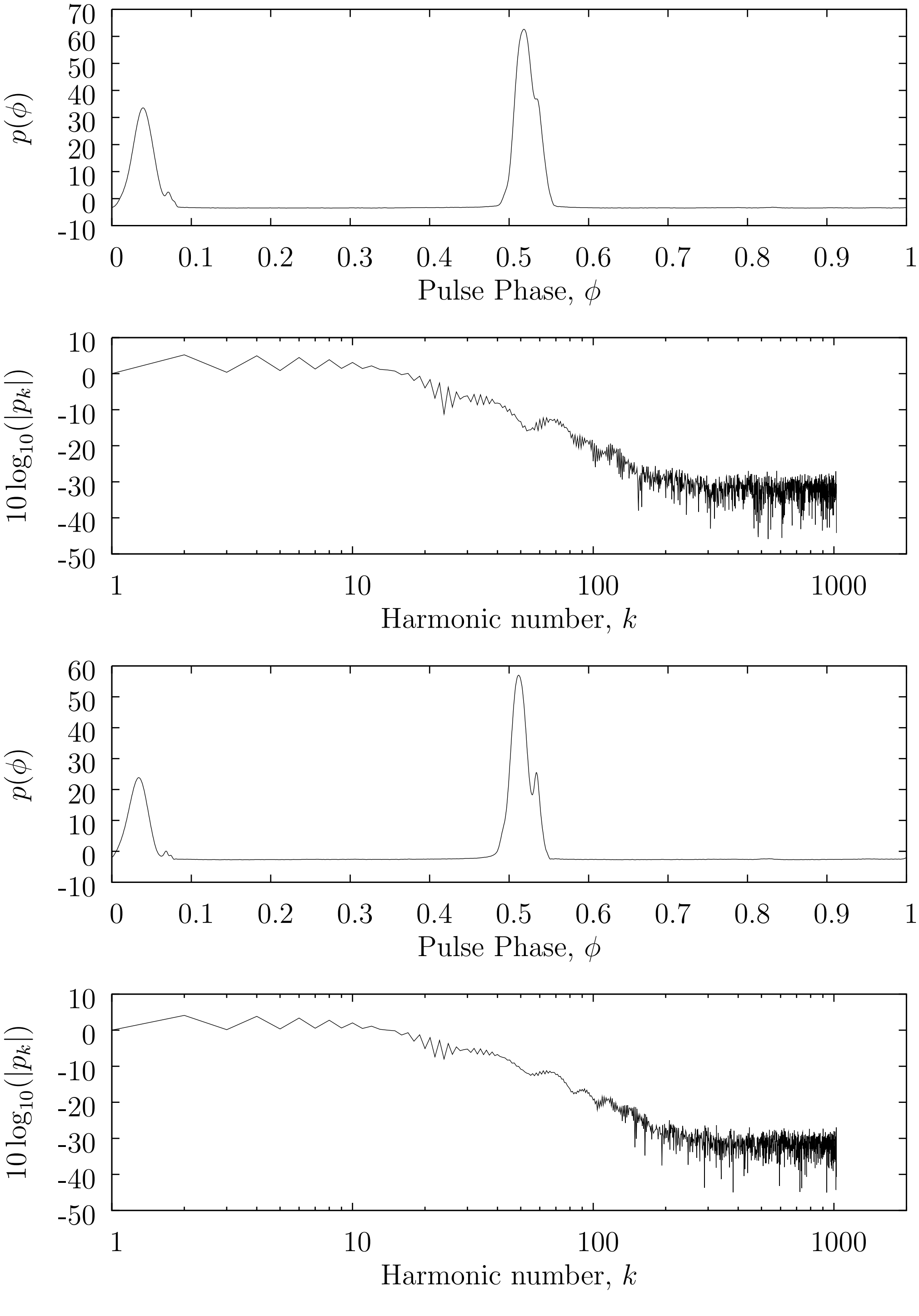

Figure 4.16: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR B1937+21.

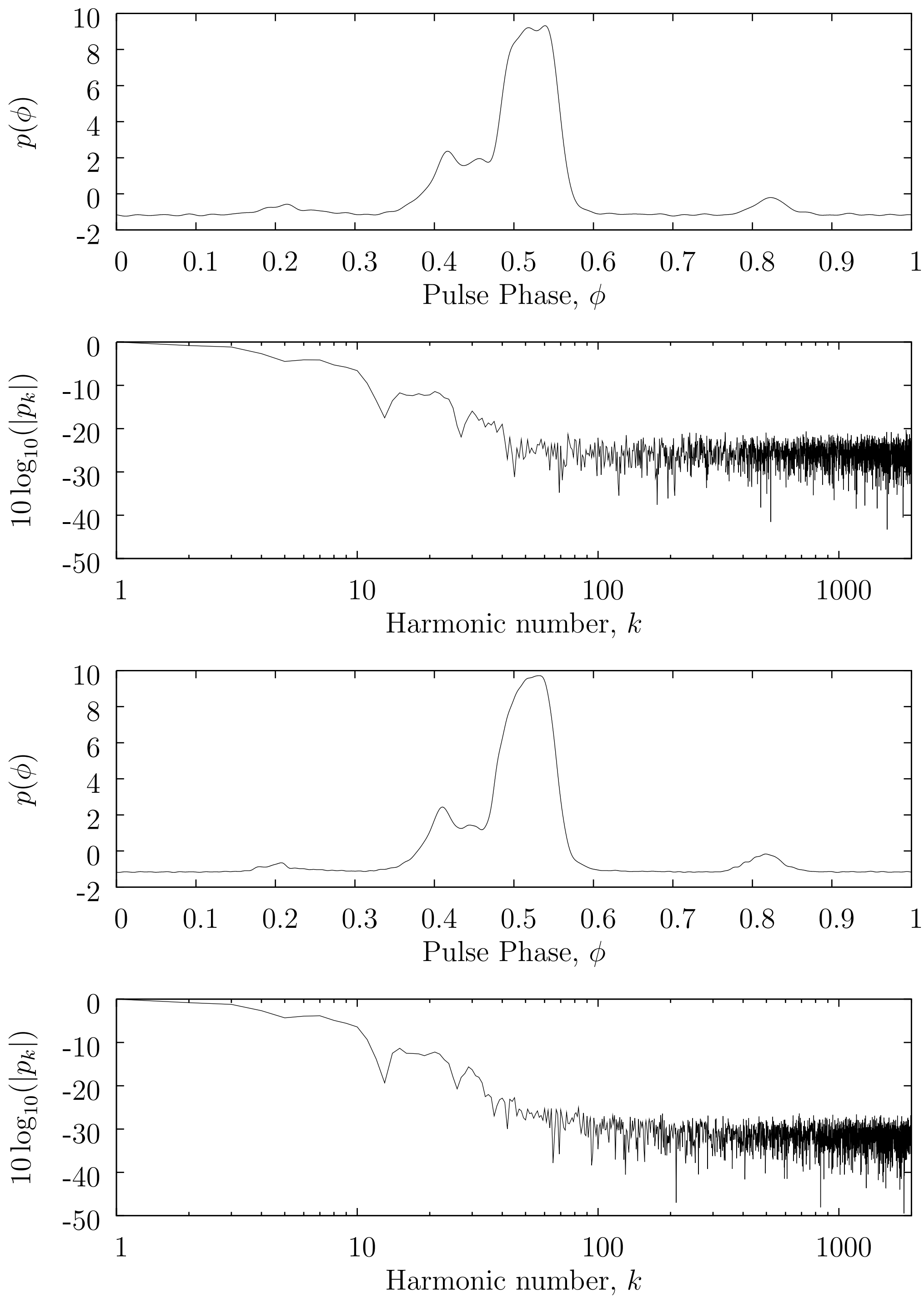

Figure 4.17: Template profiles and power spectra at 327 MHz (top) and 430 MHz (bottom) for PSR J2019+2425.

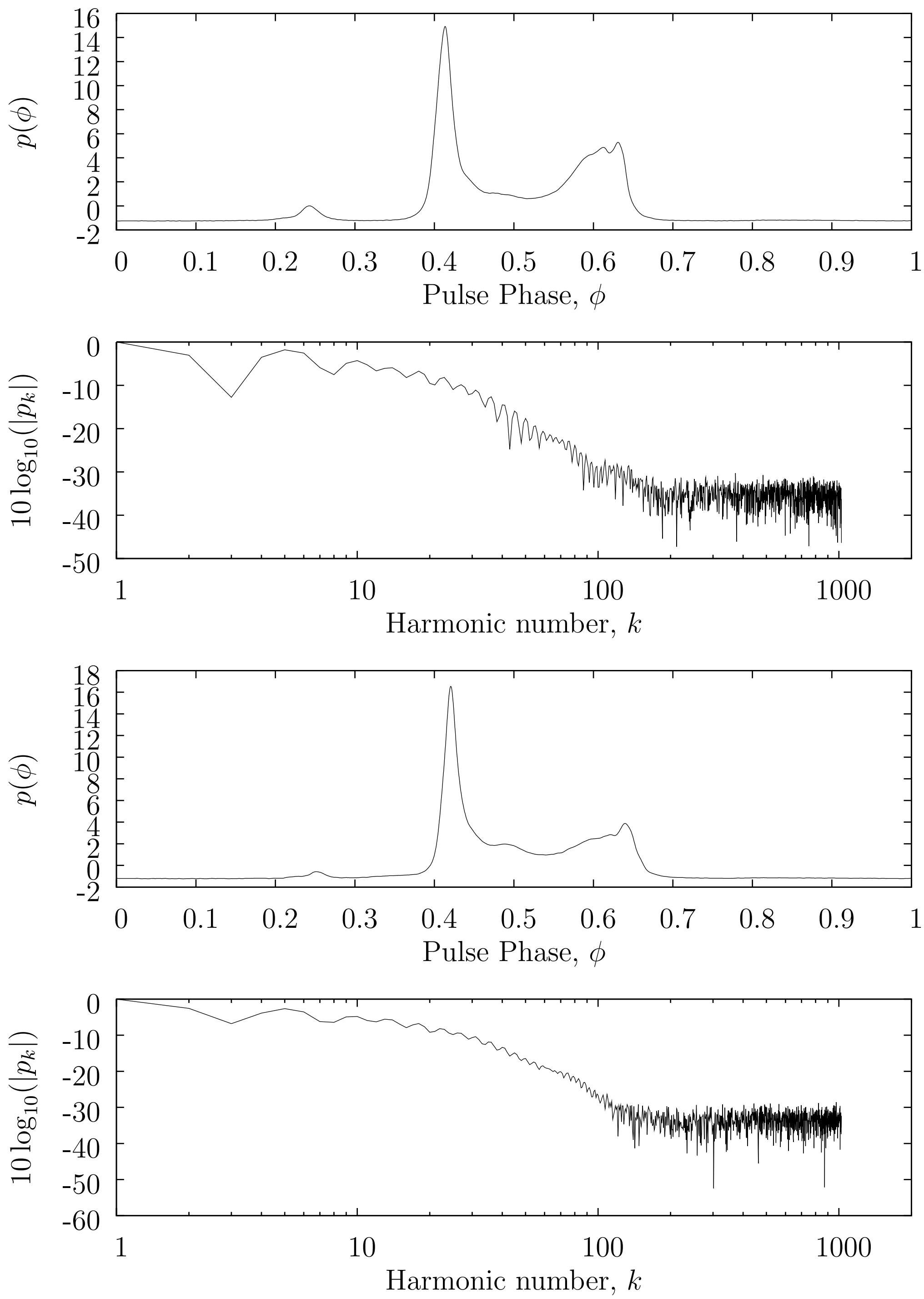

Figure 4.18: Template profiles and power spectra at 820 MHz (top) and 1400 MHz (bottom) for PSR J2145−0750.

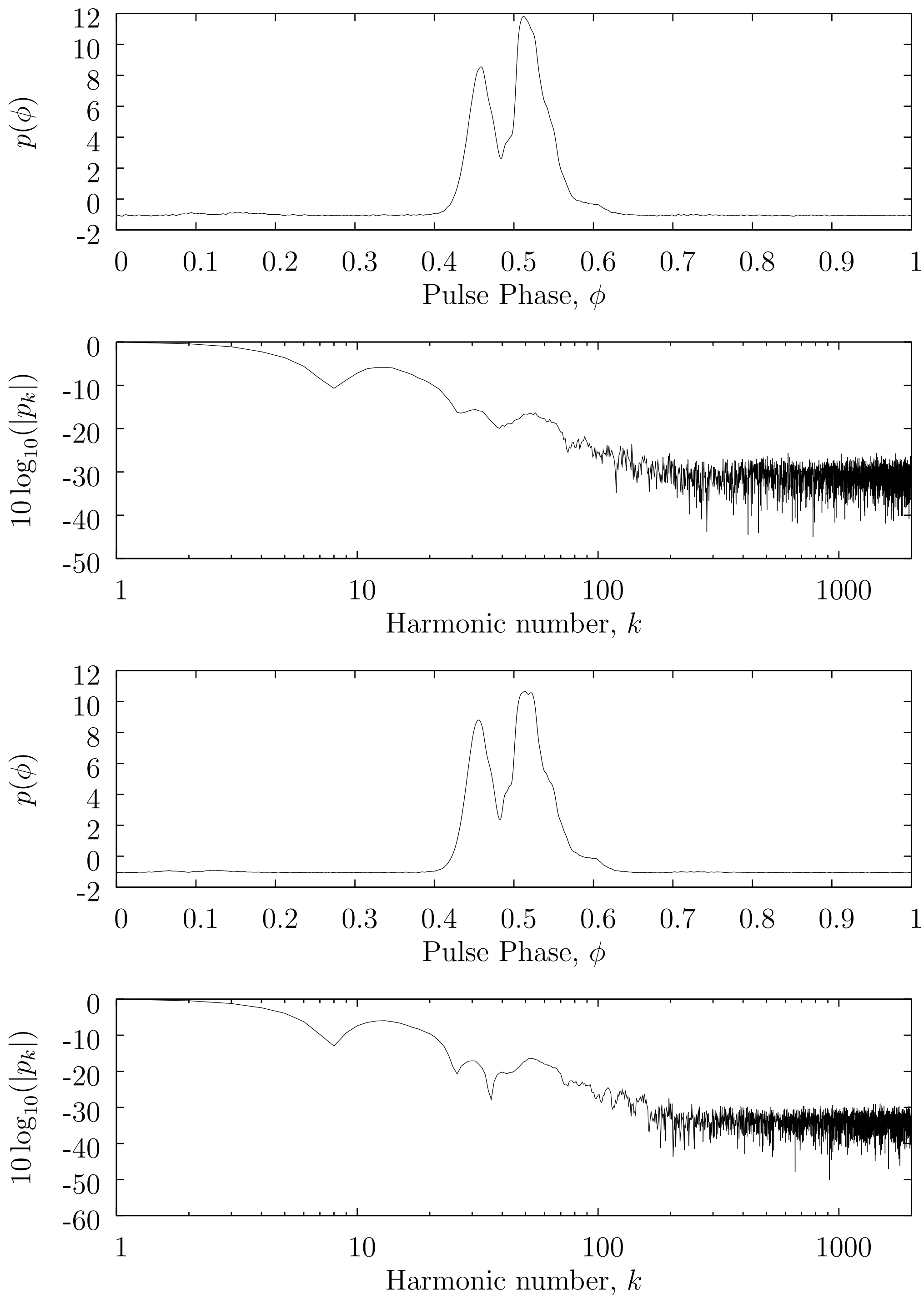

Figure 4.19: Template profiles and power spectra at 327 MHz (top) and 430 MHz (bottom) for PSR J2317+1439.

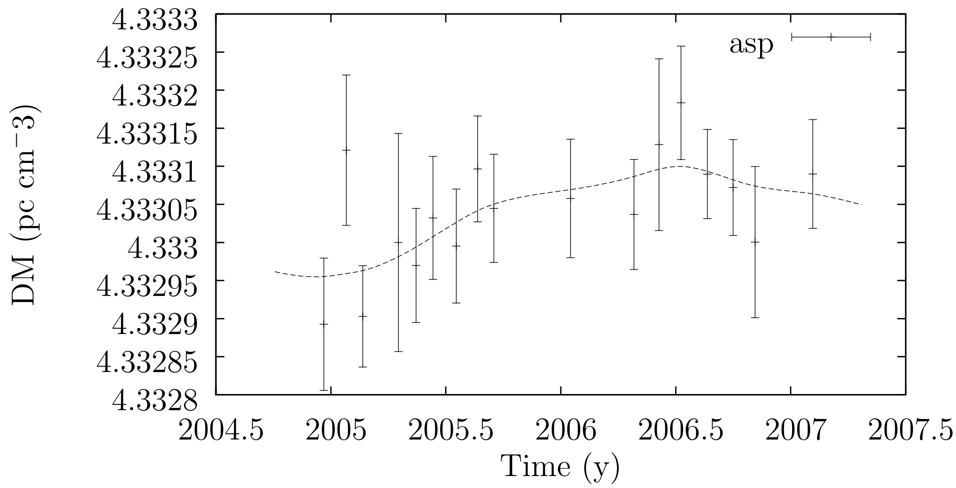

Figure 4.20: Dispersion measure versus time for PSR J0030+0451.

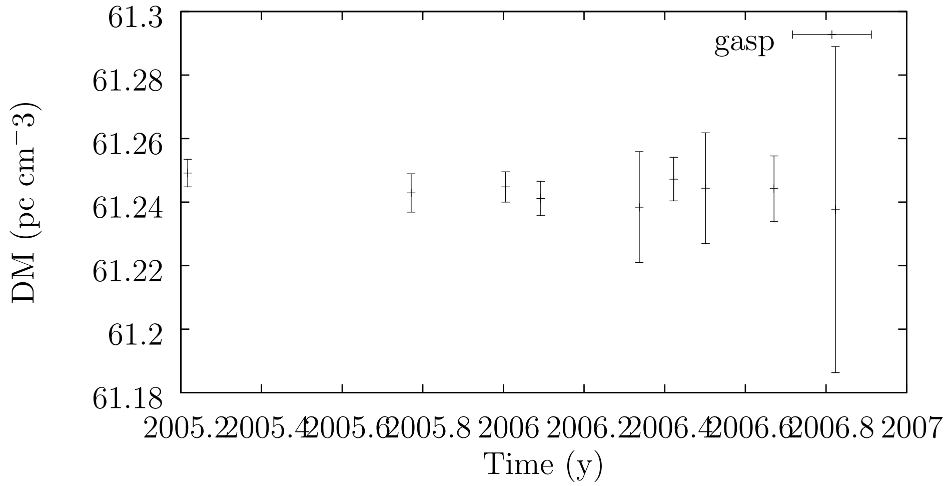

Figure 4.21: Dispersion measure versus time for PSR J0218+4232.

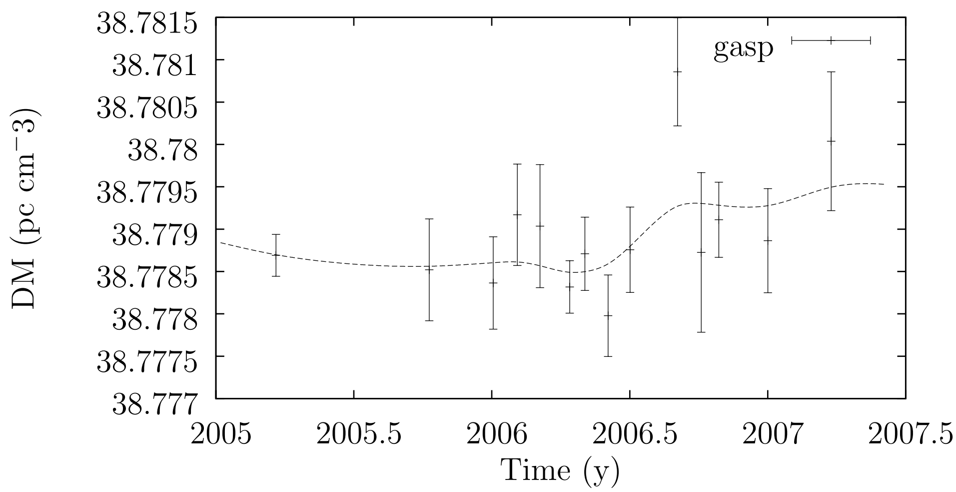

Figure 4.22: Dispersion measure versus time for PSR J0613−0200.

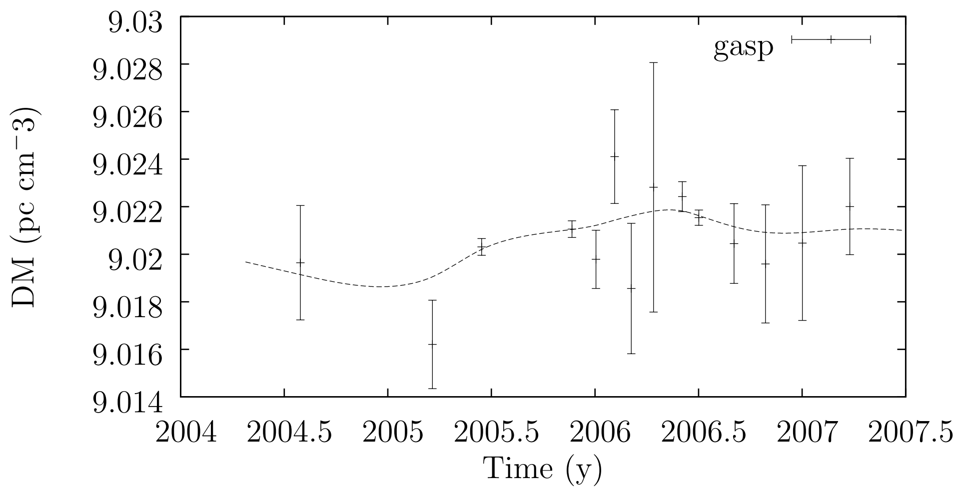

Figure 4.23: Dispersion measure versus time for PSR J1012+5307.

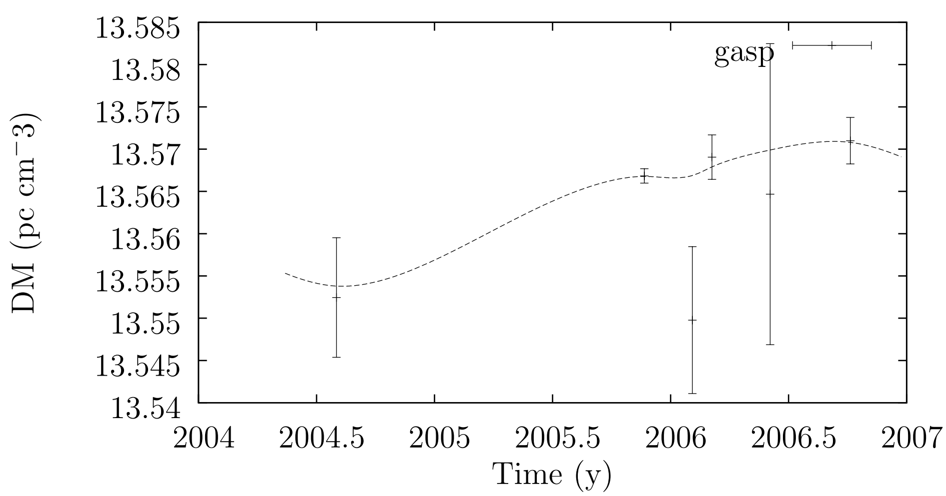

Figure 4.24: Dispersion measure versus time for PSR J1455−3330.

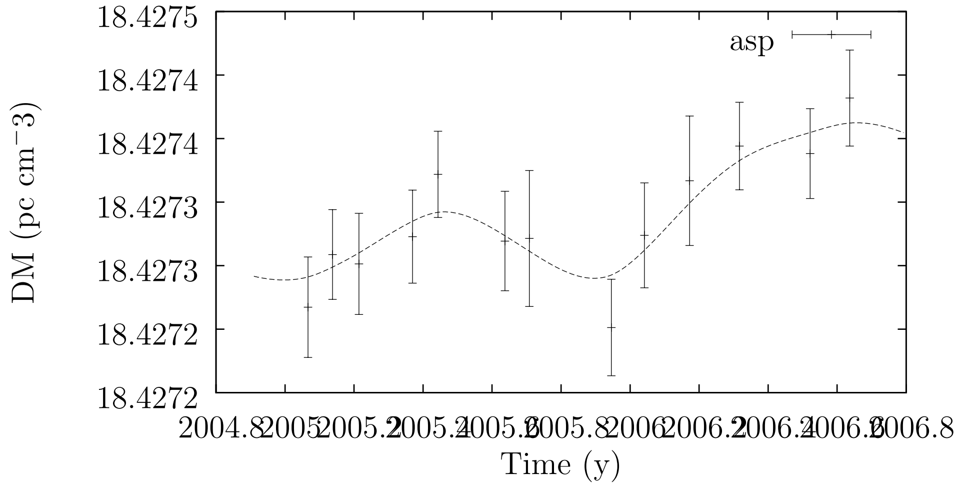

Figure 4.25: Dispersion measure versus time for PSR J1640+2224.

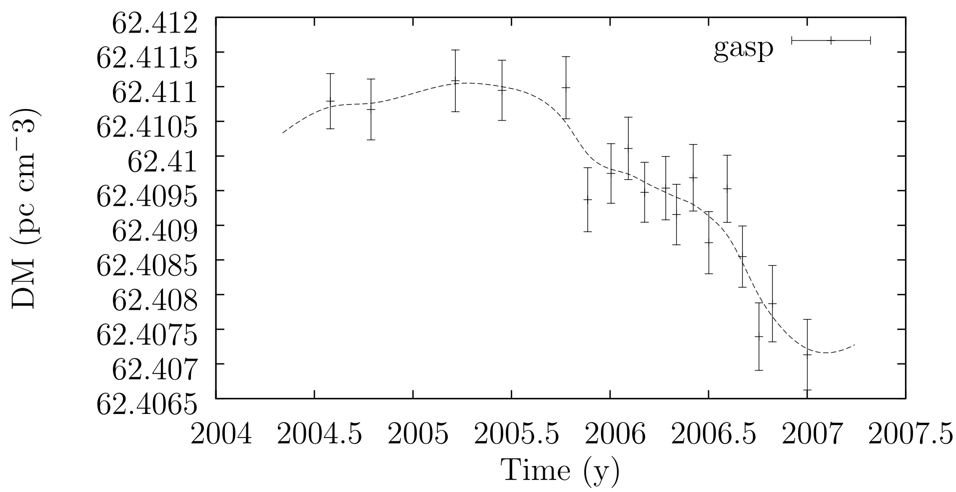

Figure 4.26: Dispersion measure versus time for PSR J1643−1224.

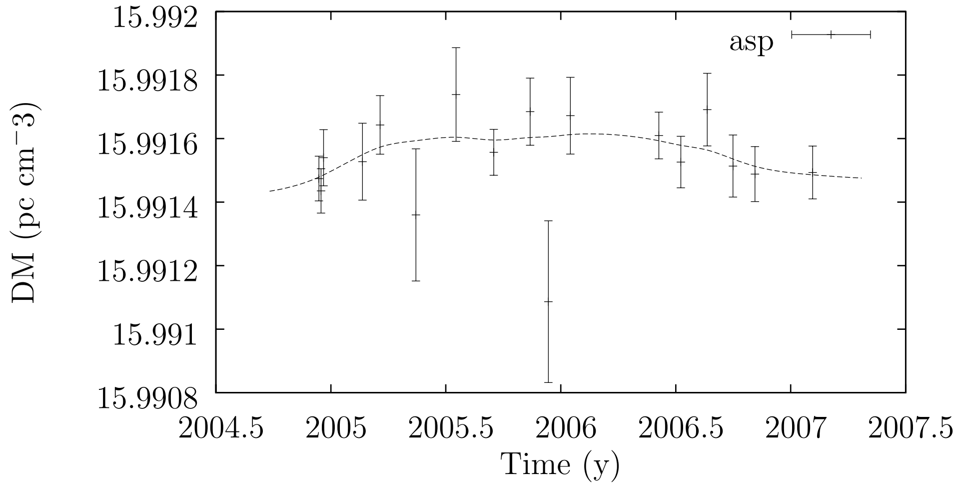

Figure 4.27: Dispersion measure versus time for PSR J1713+0747.

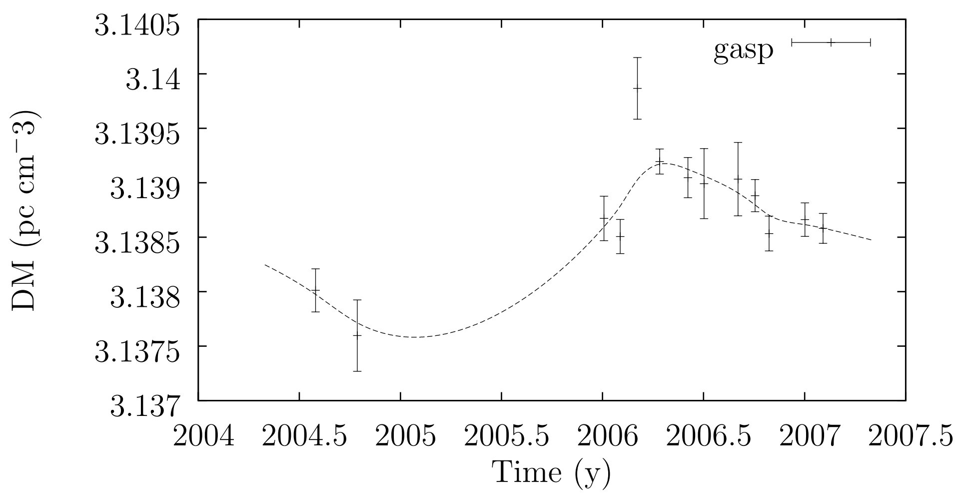

Figure 4.28: Dispersion measure versus time for PSR J1744−1134.

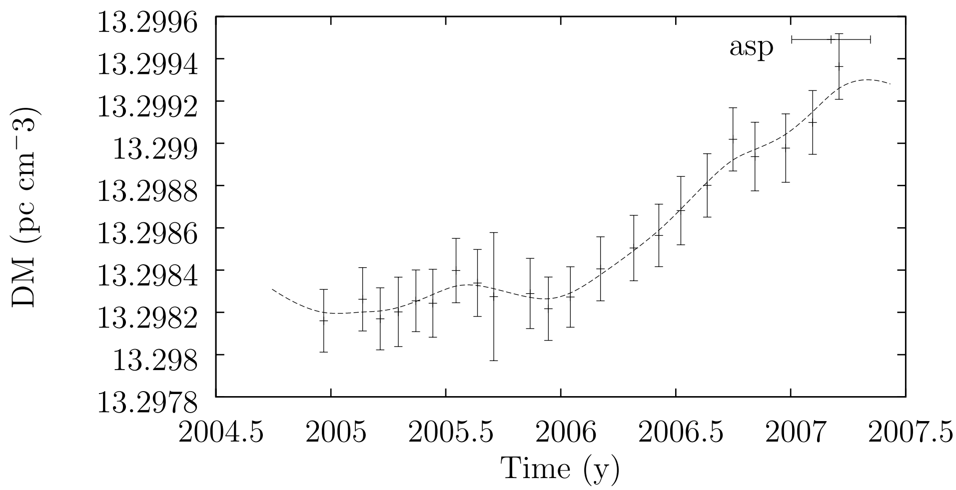

Figure 4.29: Dispersion measure versus time for PSR B1855+09.

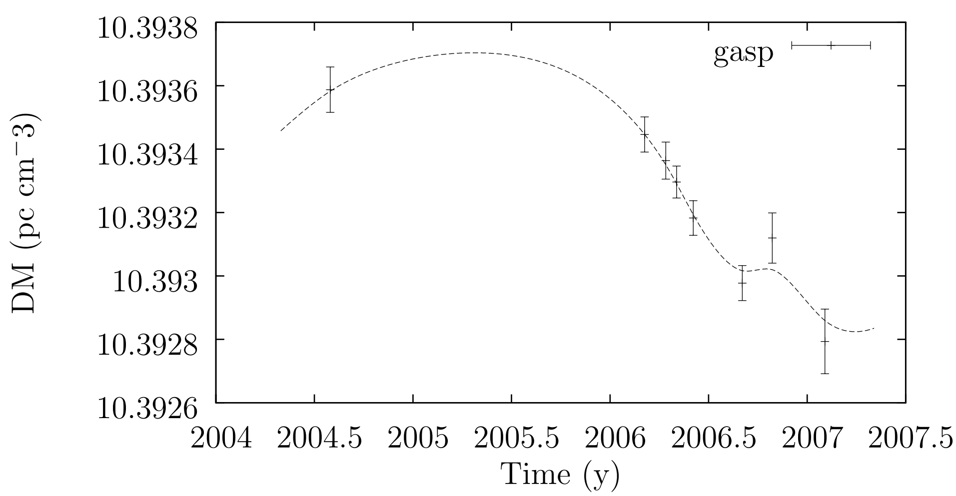

Figure 4.30: Dispersion measure versus time for PSR J1909−3744.

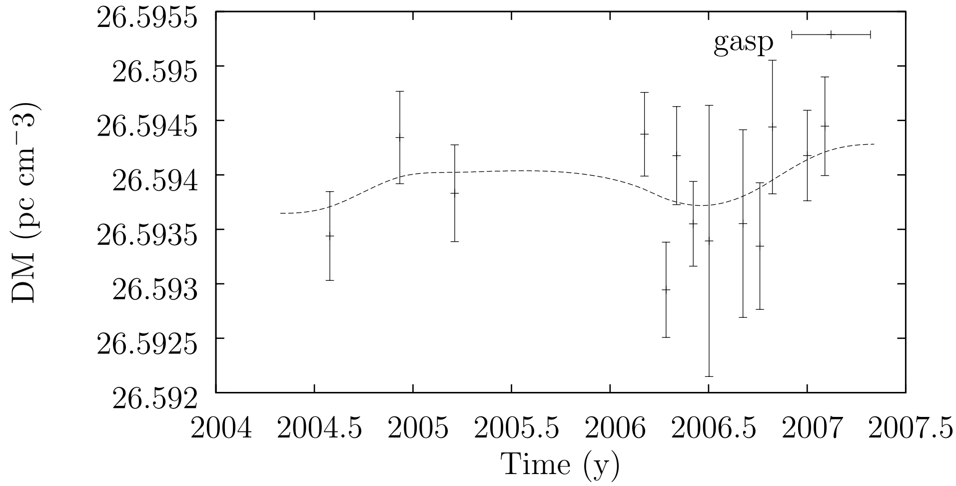

Figure 4.31: Dispersion measure versus time for PSR J1918−0642.

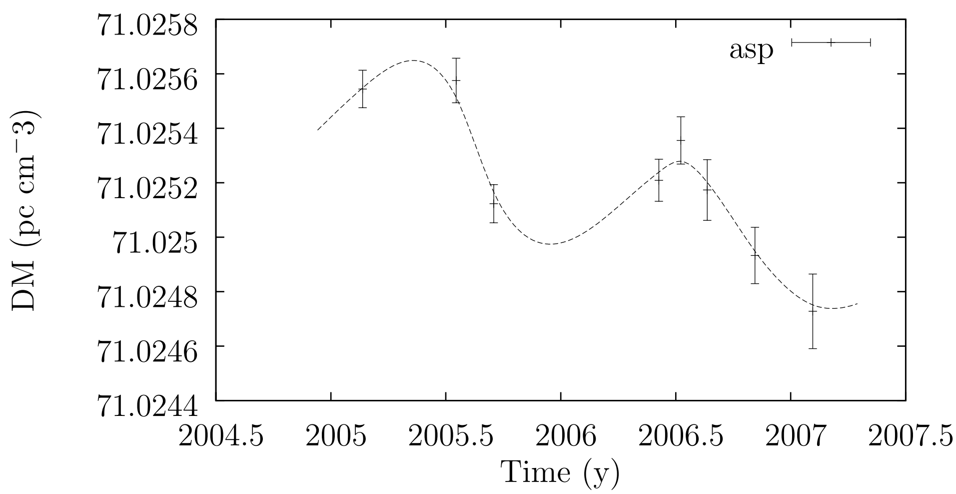

Figure 4.32: Dispersion measure versus time for PSR B1937+21.

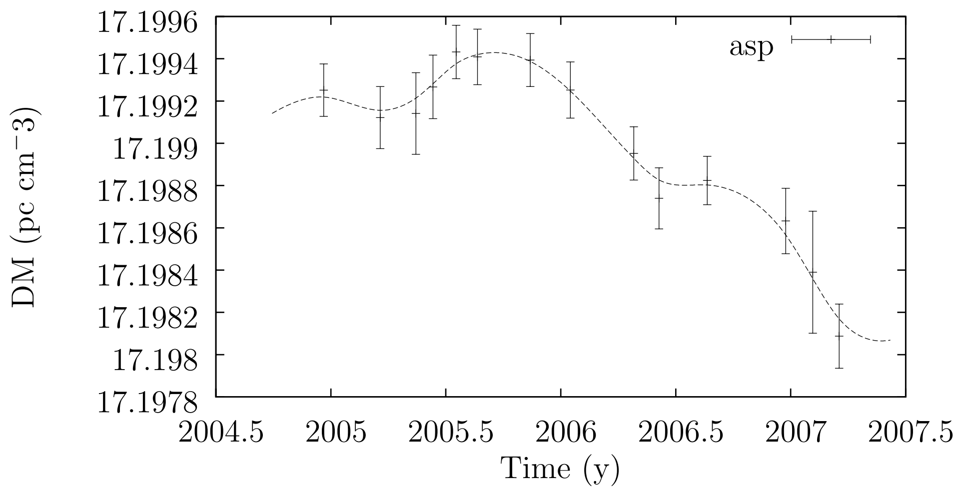

Figure 4.33: Dispersion measure versus time for PSR J2019+2425.

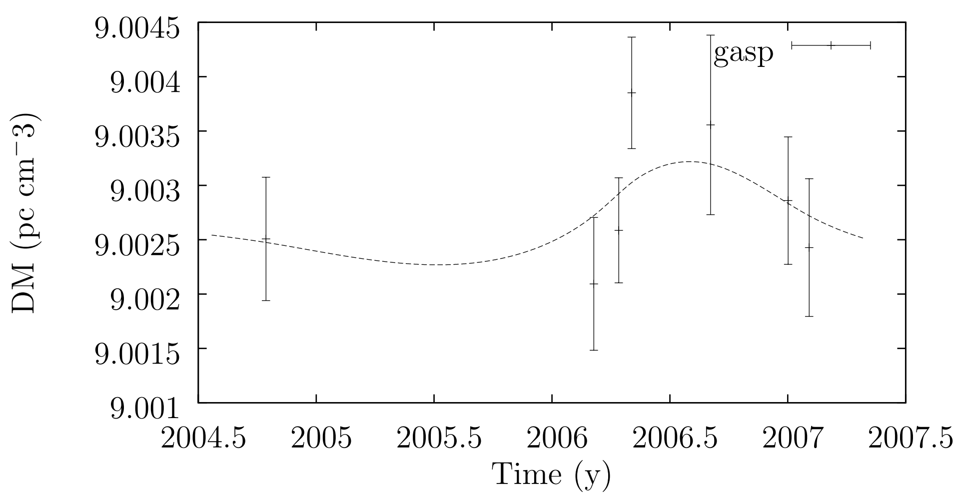

Figure 4.34: Dispersion measure versus time for PSR J2145−0750.

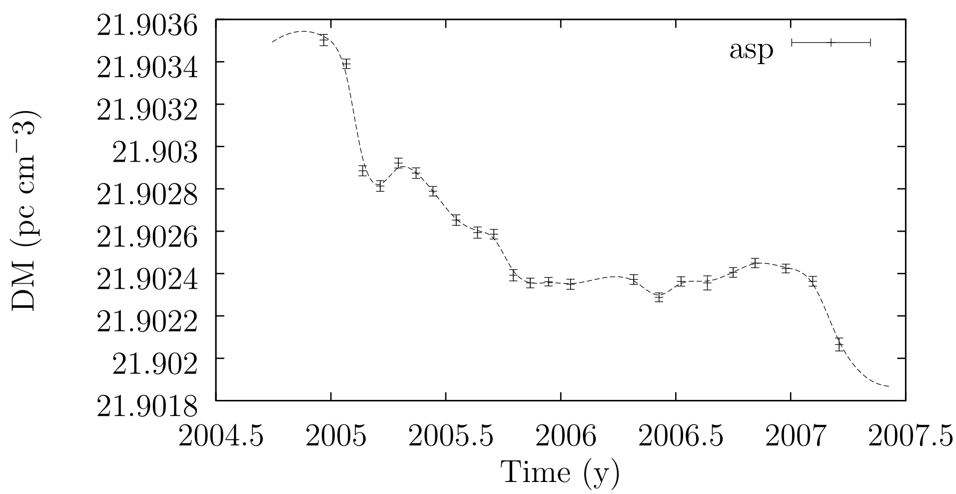

Figure 4.35: Dispersion measure versus time for PSR J2317+1439.

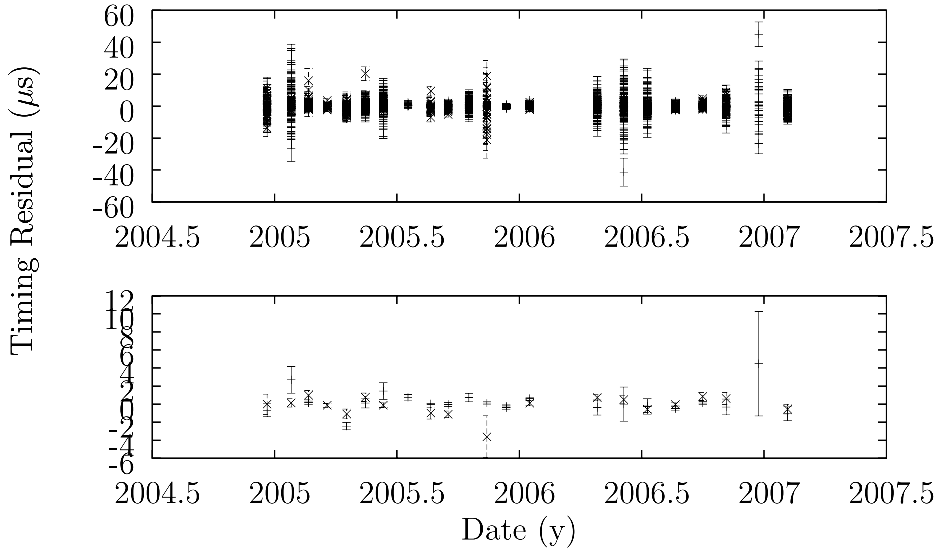

Figure 4.36: Timing residuals for PSR J0030+0451.

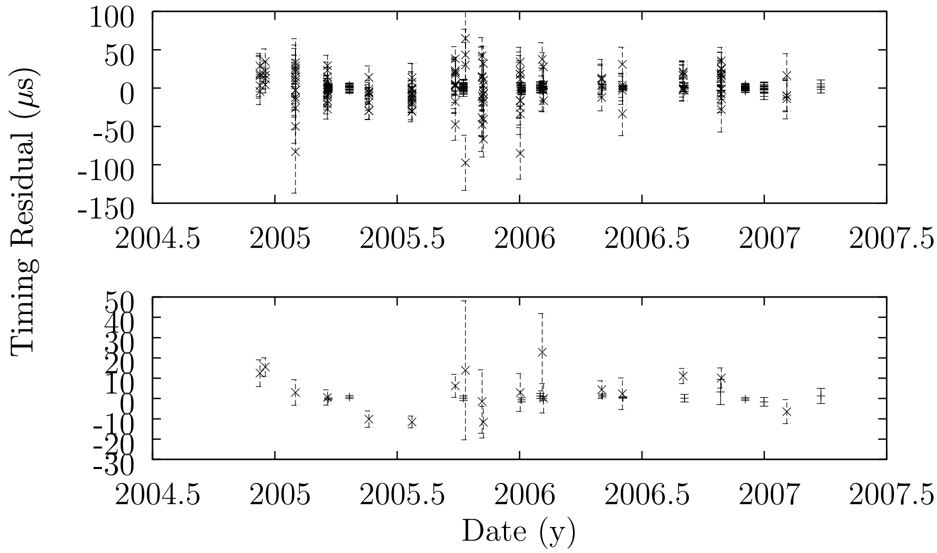

Figure 4.37: Timing residuals for PSR J0218+4232.

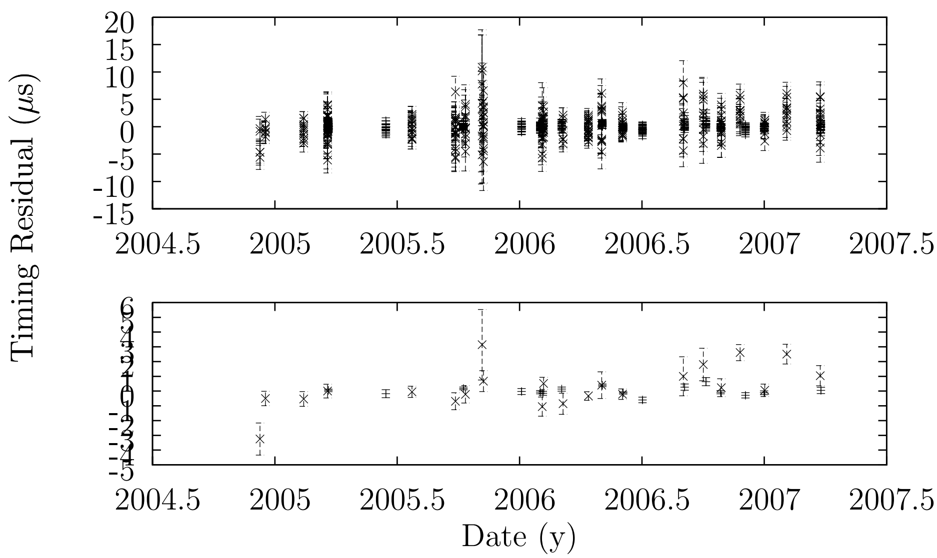

Figure 4.38: Timing residuals for PSR J0613−0200.

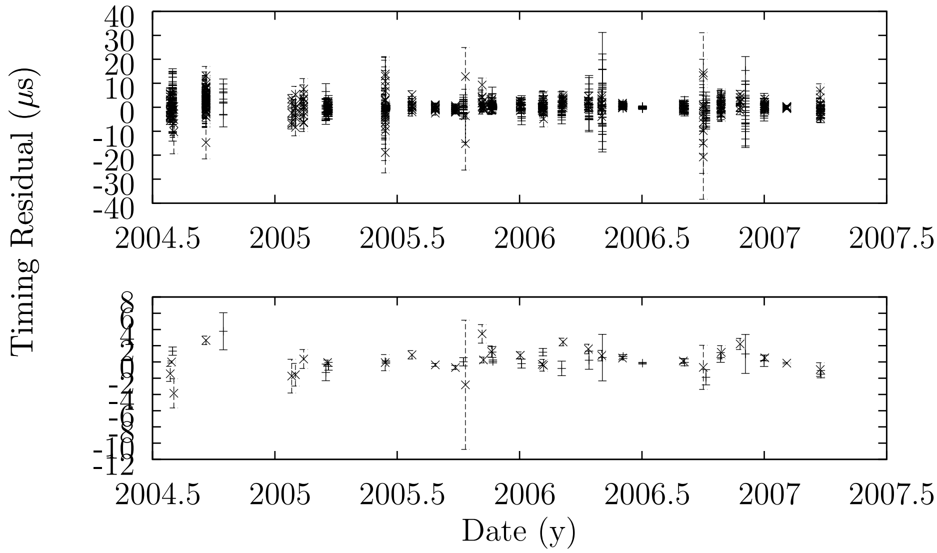

Figure 4.39: Timing residuals for PSR J1012+5307.

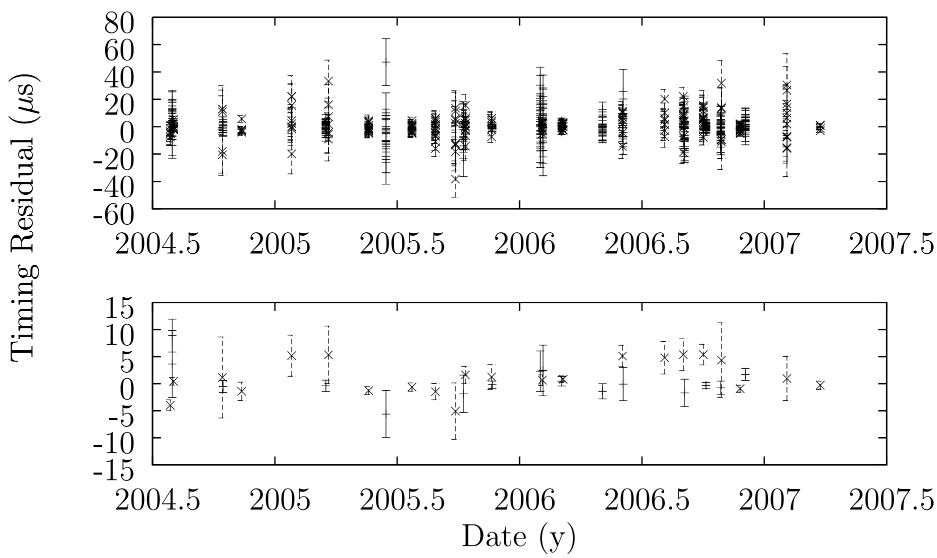

Figure 4.40: Timing residuals for PSR J1455−3330.

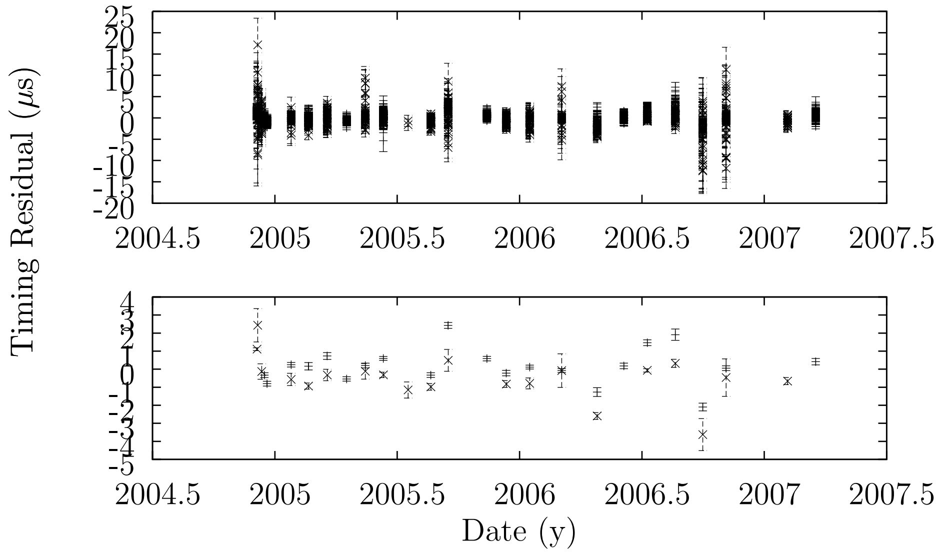

Figure 4.41: Timing residuals for PSR J1640+2224.

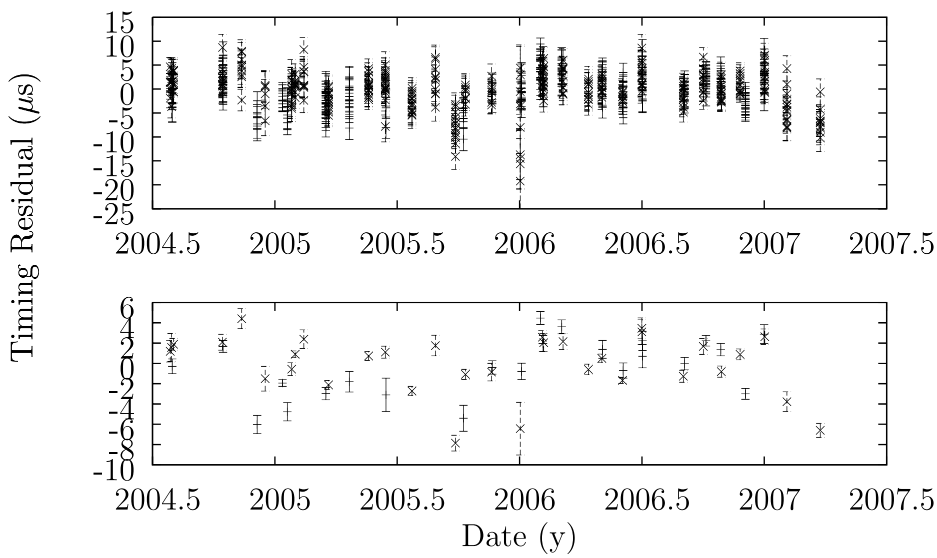

Figure 4.42: Timing residuals for PSR J1643−1224.

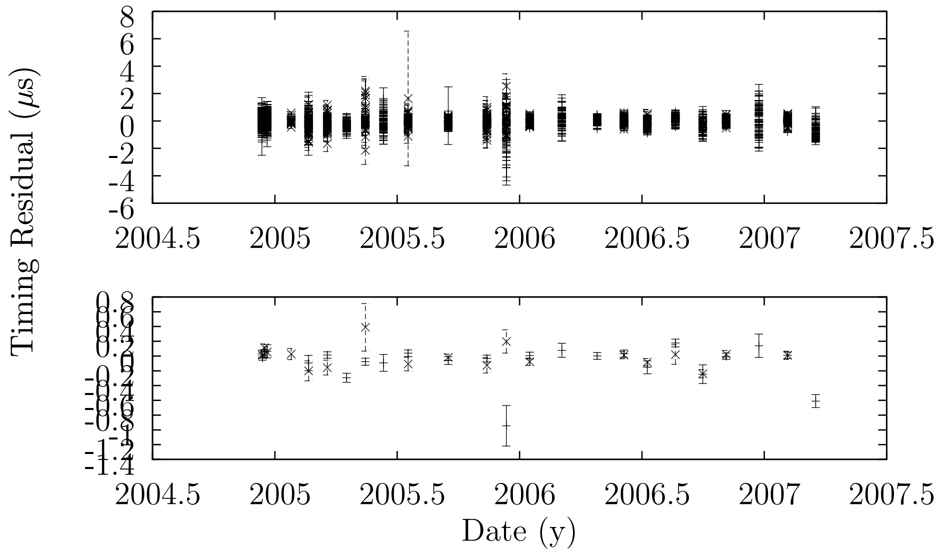

Figure 4.43: Timing residuals for PSR J1713+0747.

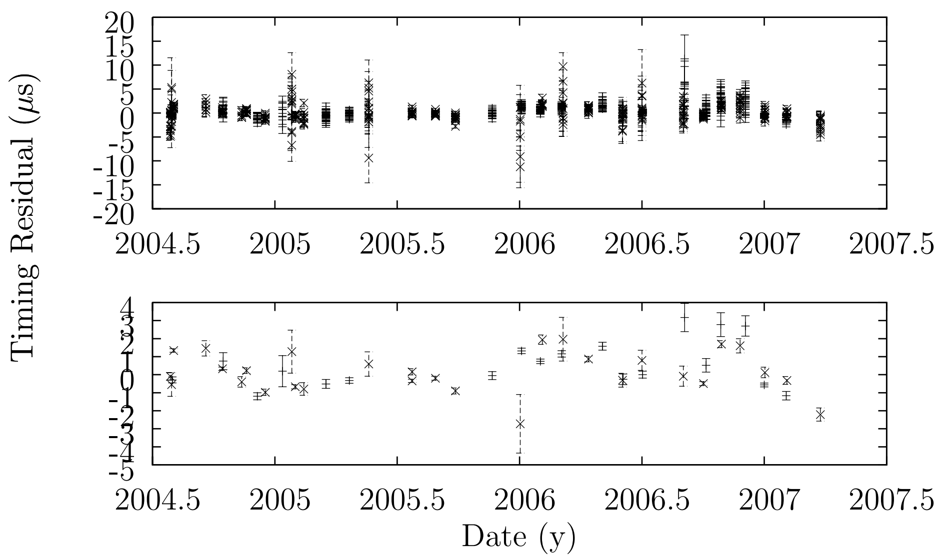

Figure 4.44: Timing residuals for PSR J1744−1134.

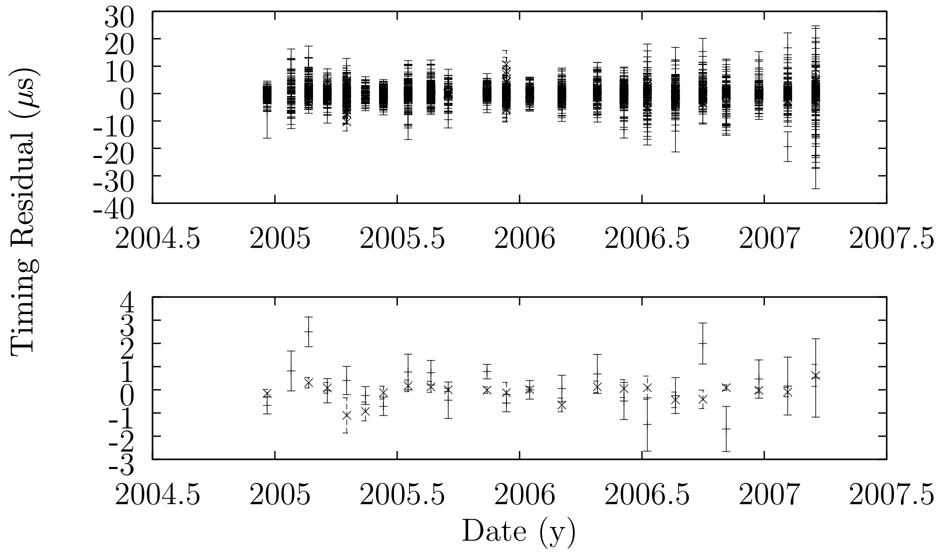

Figure 4.45: Timing residuals for PSR B1855+09.

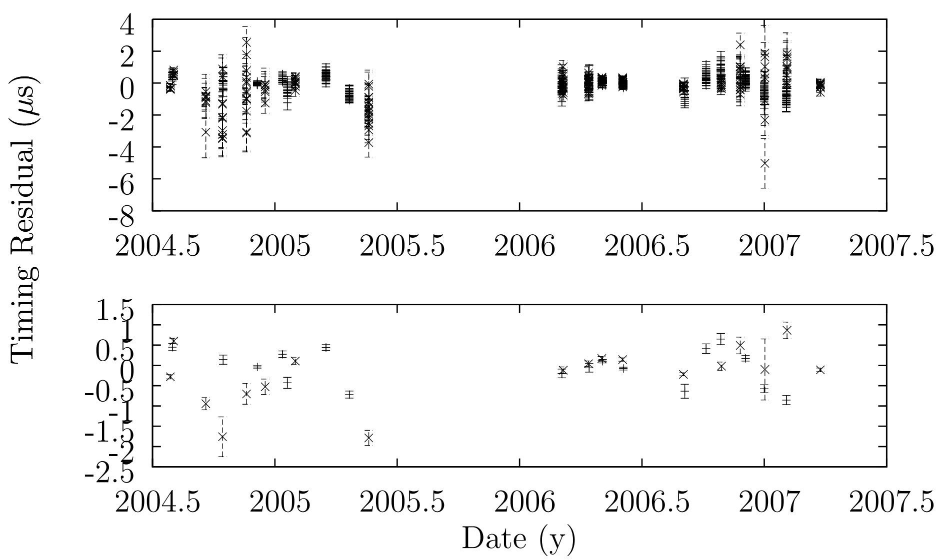

Figure 4.46: Timing residuals for PSR J1909−3744.

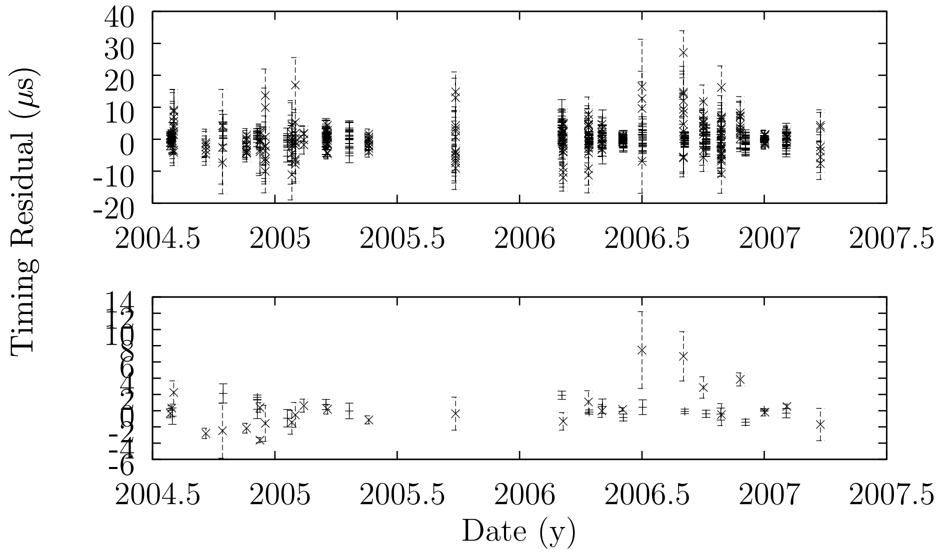

Figure 4.47: Timing residuals for PSR J1918−0642.

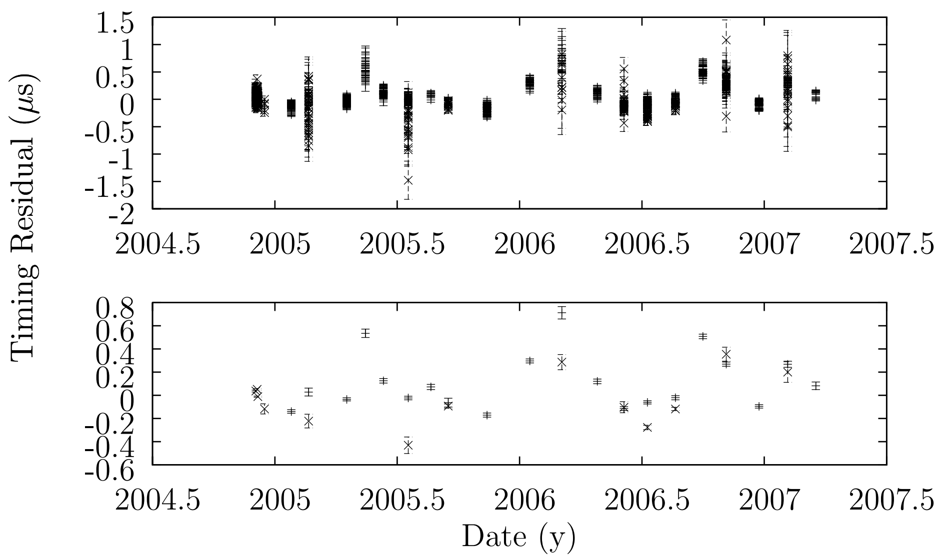

Figure 4.48: Timing residuals for PSR B1937+21.

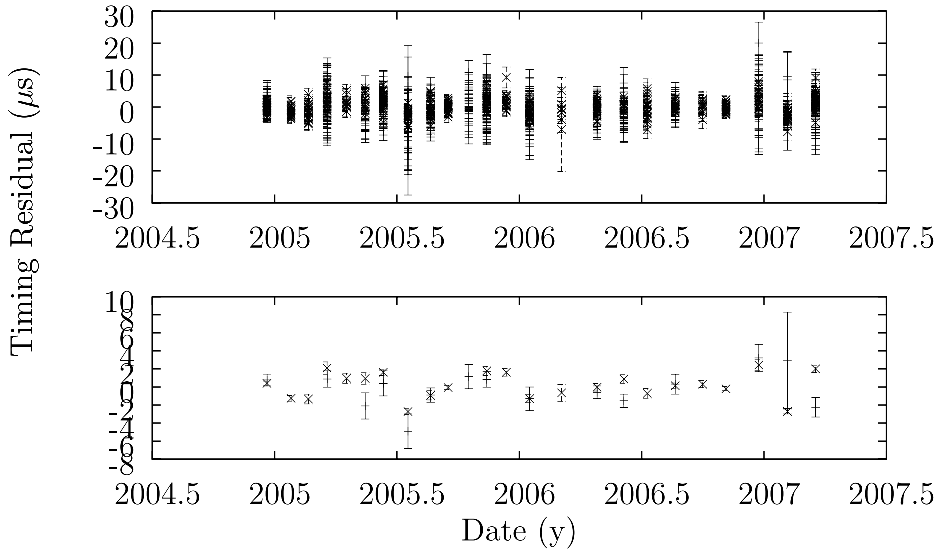

Figure 4.49: Timing residuals for PSR J2019+2425.

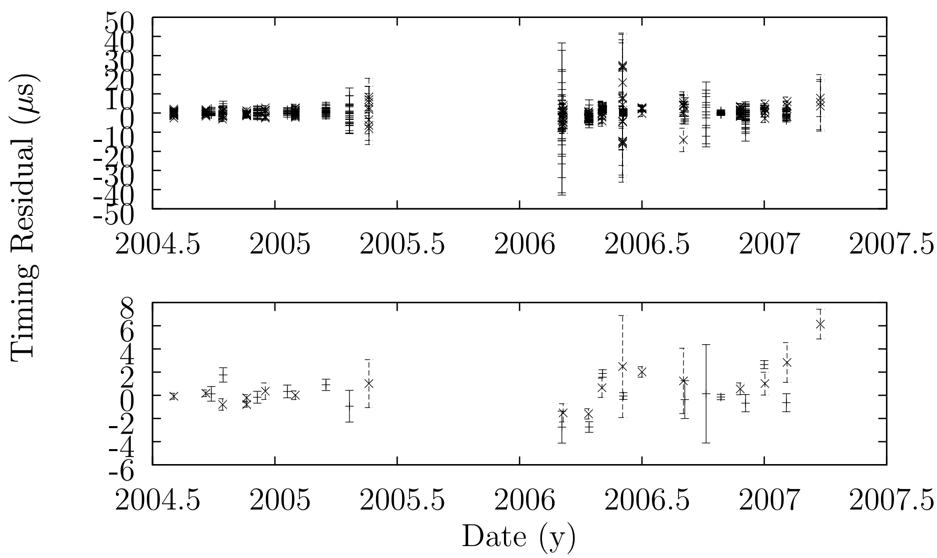

Figure 4.50: Timing residuals for PSR J2145−0750.

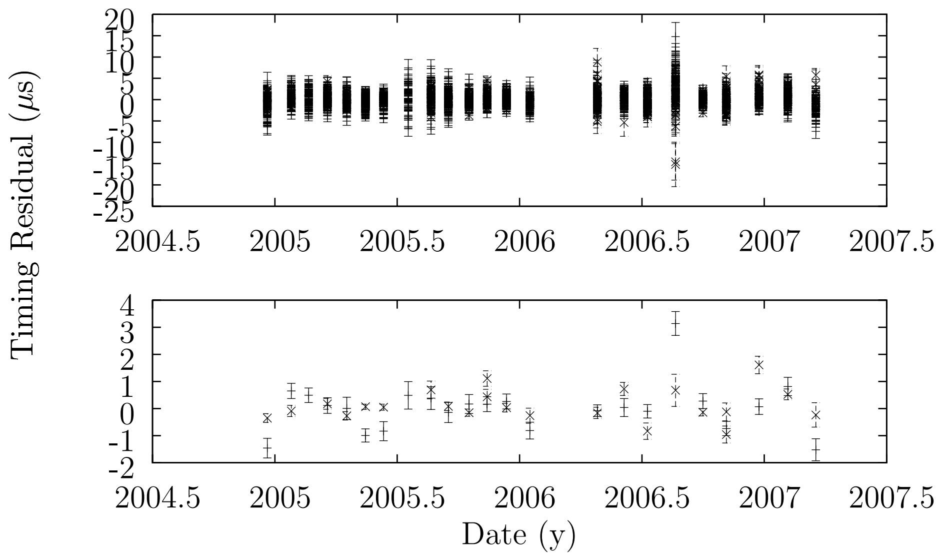

Figure 4.51: Timing residuals for PSR J2317+1439.

Chapter 5

Gravitational Radiation Data Analysis

5.1 Introduction

Direct detection of gravitational radiation (GW) is one of a few unsolved fundamental challenges in experimental physics. Major efforts are currently underway involving both ground- and space-based detectors for this purpose. These include the Laser Interferometer Gravitational-wave Observatory (LIGO), a currently operating ground-based laser experiment, and the planned Laser Interferometer Space Antenna (LISA), which would employ a set of satellites. It has long been known that pulsar timing presents a unique method for detecting GW as well (Detweiler, 1979; Hellings and Downs, 1983). Pulsar timing is potentially sensitive to GW with frequencies near $\nu_{GW} \sim 1/T$, where T is the experiment duration. As timing data typically spans 1–10 years, this puts $\nu_{GW} \sim 10^{-9}$ Hz. This frequency range is complementary to those explored by other experiments: LIGO operates in the 10^3 Hz range, and LISA is planned to explore the 10^{-3} Hz band. The GW strain sensitivity can be estimated to order of magnitude as $h \sim \delta t/T$, where δt is the accuracy of the timing data and can be as low as ~100 ns (see Chapter 4).

In this chapter, we present a summary of how gravity waves affect pulsar timing data, and discuss the strongest likely source: The merger of massive black hole (MBH; $M_{BH} \sim 10^{6-9}\ M_\odot$) binary systems throughout the history of the universe. Aside from being intrinsically interesting, the detection of GW would help constrain the overall MBH demographics and merger rate (Jaffe and Backer, 2003). Other sources that pulsar GW experiments offer unique constraints on are cosmic stings – theoretical remnants of phase transitions in the early universe (Damour and Vilenkin, 2005; Siemens et al., 2007; Hogan, 2006). We then present a newly developed analysis technique that compensates for the effect of the pulsar timing model fit on the GW signal. Finally, we will apply this method to the ~2.5 years of timing data presented in Chapter 4 and derive the resulting GW limits. The starting point is a brief discussion of the relevant concepts in general relativity (see also the standard texts by Schutz (1985) and Misner et al. (1973)).

In general relativity, the main quantity of interest is the metric tensor, $g_{\mu\nu}$. This can be represented as a 4-by-4 symmetric matrix whose dimensionless elements are functions of position in 4-dimensional space-time. The purpose of the metric tensor is to describe the geometry of space-time. For example, the frame-invariant distance between space-time events at the points x^μ and $x^\mu + dx^\mu$ is given by:[1]

$$ds^2 = dx_\mu dx^\mu = dx^\mu dx^\nu g_{\mu\nu}(x^\mu) \tag{5.1}$$

The metric tensor describes the action of gravity through the statement that any particle free of the influence of non-gravitational forces moves through space-time along a geodesic: a curve of minimum length, given the definition of length in Equation 5.1. In turn, the distribution of matter and energy determines the metric tensor through Einstein's equation.

For many applications, including GW, it is useful to work in a space which is nearly flat, aside from a small perturbation. In this case, we have $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, where $\eta_{\mu\nu}$ is the flat space (or Minkowski) metric, and $|h_{\mu\nu}| \ll 1$ everywhere. This leads to the weak-field (or linearized) Einstein equation:

$$\partial_\mu \partial^\mu \left(h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h^\alpha{}_\alpha \right) = -16\pi T_{\mu\nu} \tag{5.2}$$

Here, $T_{\mu\nu}$ is the stress-energy tensor, which describes the density and flow of energy and momentum at each point in space-time. Freely propagating wave solutions can be obtained by setting $T_{\mu\nu} = 0$, in which case Equation 5.2 reduces to the simple homogeneous wave equation.

[1]In this chapter, a "summation convention" on repeated indices is assumed. Greek indices $(\mu, \nu, \ldots)$ represent 4-dimensional tensor quantities, while Roman indices $(i, j, \ldots)$ represent 3-dimensional, flat-space quantities. We will also set $c = G = 1$ unless otherwise noted.

Since it is only derivatives of $h_{\mu\nu}$ that appear in both Equation 5.2 and the geodesic equations of motion, it makes sense that not all components of $h_{\mu\nu}$ are physically significant: There is a gauge freedom in determining the form of h. In this way, h is more analogous to the vector potential of electromagnetism than to the electric or magnetic field. For a plane wave travelling in a given direction, there are only two physical degrees of freedom (polarizations) in h. The usual gauge choice for h is called the "transverse-traceless" gauge, in which a plane wave travelling along the $\hat{\mathbf{z}}$ direction is represented as:

$$h_{\mu\nu}^{TT} = A_{\mu\nu} e^{ik_\mu x^\mu}, \qquad A_{\mu\nu} = A_+ e_{\mu\nu}^{+} + A_\times e_{\mu\nu}^{\times} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_+ & A_\times & 0 \\ 0 & A_\times & -A_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.3}$$

In this, A_+ and $A_\times$ give the amplitudes of the two polarizations, usually called the "plus" and "cross" polarizations respectively. These are represented by the basis tensors $e_{\mu\nu}^{+,\times}$. Figure 5.1 shows the effect of each polarization on a initially circular ring of particles, oriented transverse to the wave propagation direction. The equations also require $k_\mu k^\mu = 0$, which means that the waves travel at the speed of light, obeying the simple dispersion relation $\omega = k$, exactly the same as electromagnetic radiation.

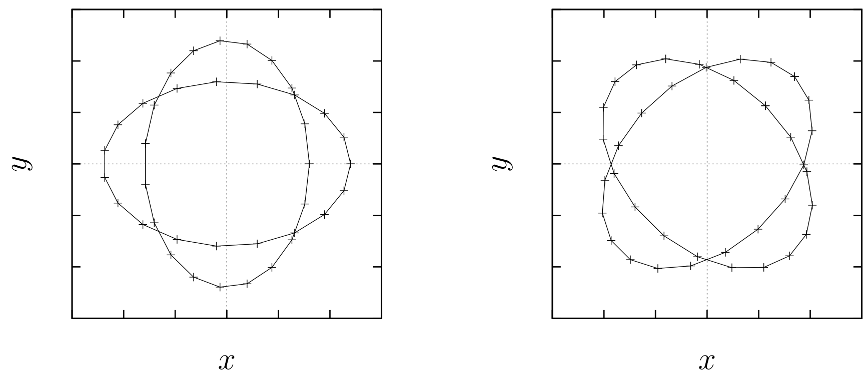

Figure 5.1: Illustration of GW polarizations.

5.2 Detecting GW Using Pulsar Timing

All modern GW detection schemes use electromagnetic radiation as a probe of the gravitational field. For our pulsar timing experiment, this is best thought of in the following way: The varying $g_{\mu\nu}$ along the line of sight path from a pulsar to Earth will either advance or retard the measured pulse time of arrival (TOA) by changing the radio pulse's time of flight. If this delay is changing as a function of time, it will appear in the timing residuals and can possibly be detected (Detweiler, 1979). The form of the delay can be calculated by using the fact that photons (since they are massless) travel along null geodesics, whose invariant path length is 0. This, combined with Equation 5.1 and the weak-field wave form of $g_{\mu\nu}$ gives:

$$0 = ds^2 = dt^2 - dx^2 - h_{ij} dx_i dx_j \tag{5.4}$$

We can get the time of flight to $o(h)$ by solving Equation 5.4 for dt, and integrating along the flat space (zeroth order) path from the pulsar to Earth.[2] If Earth is at the origin, the pulsar is at a distance d along the direction $\hat{\mathbf{n}}$, and the pulse arrives at Earth at time t_0, then the path can be parameterized as $x_i(r) = (d-r)n_i$, $t(r) = t_0 + r - d$, and the time of flight is found as:

$$\begin{aligned} T = \int dt &= \int_0^d \left(\frac{dx_i}{dr}\frac{dx_i}{dr} + h_{ij}\frac{dx_i}{dr}\frac{dx_j}{dr} \right)^{1/2} dr \\ &= \int_0^d (1 - h_{ij}n_i n_j)^{1/2}\, dr \\ &= d - \frac{1}{2}n_i n_j \int_0^d h_{ij} dr + o(h^2) \end{aligned} \tag{5.5}$$

The constant term d is simply the flat space (no GW) time of flight. We are interested in deviations about this value:

$$\Delta \equiv T - d = -\frac{1}{2}n_i n_j \int_0^d h_{ij}(x_i(r), t(r))dr \tag{5.6}$$

Since we know that $h_{ij}(x_i, t)$ satisfies the wave equation, we can proceed further than Equation 5.6. Any wave solution can be represented as a superposition of plane waves:

$$h_{ij}(x_i, t) = \int h_{ij}(\mathbf{k}) e^{ikt - ik_i x_i} d^3\mathbf{k} \tag{5.7}$$

Here $h_{ij}(\mathbf{k})$ is the amplitude of a wave propagating in the $\hat{\mathbf{k}}$ direction with angular frequency k. It should be noted that there are only 2 freely adjustable components of each $h_{ij}(\mathbf{k})$, as stated in §5.1. Basis tensors for these can be found by rotating the matrix in Equation 5.3. Also, $h_{ij}(\mathbf{k})$ is not dimensionless, but rather has dimension

[2] An alternate statement of the problem, which leads to identical results, is to calculate the time-varying proper distance between the pulsar and Earth.

L^3. Combining Equations 5.6 and 5.7 results in the following expression for Δ:

$$\Delta = \frac{i}{2} n_i n_j \int d^3\mathbf{k} \frac{h_{ij}(\mathbf{k})}{k + k_i n_i} \left(e^{ikt_0} - e^{ik(t_0 - d) - ik_i n_i d}\right) \tag{5.8}$$

Equation 5.8 merits some discussion. It describes, for a arbitrary collection of plane waves, the effect on the TOAs of a given pulsar. There are two important features of this expression: First, the terms in the parentheses represent the wave phase evaluated at the Earth (at the time of pulse reception) and at the pulsar (at the time of emission), respectively. This feature has been noted in various forms by many authors (Burke, 1975; Detweiler, 1979; Hellings and Downs, 1983) and is often used as motivation for ignoring the pulsar term on grounds that it will be uncorrelated between different pulsars. We will return to that question in §5.4. Second, we can see that Δ does not depend simply on the value of the metric tensor at Earth, as is sometimes stated. Rather, it comes from a weighted sum of all incident plane waves. This remains true even if our description is in terms of redshift $Z = d\Delta/dt_0$.

There are two particular forms of $h_{ij}(\mathbf{k})$ that we will discuss in detail. First, the single monochromatic plane wave is useful conceptually, and provides a connection to previous work. Then, in §5.4, a uniform stochastic background of waves with a red power spectrum, which is expected to be the dominant signal in reality. While not discussed here, other forms such as a spherical wave can be useful in certain cases (Jenet et al., 2004).

If we are dealing with a single plane wave as given in Equation 5.3, then $h_{ij}(\mathbf{k})$

contains a delta function, and the integral vanishes, leaving:

$$\begin{aligned}
\Delta(t_0) &= \frac{i}{2}\frac{n_i A_{ij} n_j}{k + k n_z}\left(1 - e^{-ik(1+n_z)d}\right)e^{ikt_0} \\
&= \frac{i}{2k}\frac{A_+(n_x^2 - n_y^2) + 2A_\times n_x n_y}{1 + n_z}\left(1 - e^{-ik(1+n_z)d}\right)e^{ikt_0} \\
&= \frac{i}{2k}\left(1 - \cos\theta\right)\left(A_+ \cos 2\phi + A_\times \sin 2\phi\right)\left(1 - e^{-ik(1+\cos\theta)d}\right)e^{ikt_0}
\end{aligned} \tag{5.9}$$

The final line in Equation 5.9 is a basic result first stated by Detweiler (1979).

It is also possible to write Equation 5.8 in a way that more explicitly shows the polarization dependence:

$$\Delta(t_0) = \frac{i}{2}\int d^3\mathbf{k}\frac{1}{k}\Big(\alpha_+(\hat{\mathbf{k}}, \hat{\mathbf{n}})A_+(\mathbf{k}) + \alpha_\times(\hat{\mathbf{k}}, \hat{\mathbf{n}})A_\times(\mathbf{k})\Big)\left(1 - e^{-ikd - ik_i n_i d}\right)e^{ikt_0} \tag{5.10}$$

In this expression, we have separated $h_{ij}(\mathbf{k})$ into its two polarization components. This formulation implicitly depends on a definition of the polarization directions, which we will leave unspecified here as it will eventually cancel out for unpolarized GW. Furthermore, each polarization term is split into an amplitude $A_{+,\times}$ and a angular factor $\alpha_{+,\times}$. The angular factors depend only on the directions $\hat{\mathbf{k}}$ and $\hat{\mathbf{n}}$ (and on the implicit polarization convention), and are defined as follows:

$$\alpha_{+,\times}(\hat{\mathbf{k}}, \hat{\mathbf{n}}) = \frac{n_j e_{jk}^{+,\times}(\hat{\mathbf{k}}) n_k}{1 + \hat{k}_i n_i} \tag{5.11}$$

This form will be used extensively in the following sections.

5.3 Effect of the Timing Model

So far we have derived an expression that describes the effect of a set of gravitational waves on a radio pulse's time of flight from a pulsar to Earth. We would like to measure these time deviations and use them as a GW detector. However, there are other effects that cause similar time deviations and need to be taken into account. By far the dominant effect is the motion (both rotational and translational) of the pulsar itself, relative to the Earth. These effects are collected into what is referred to as a timing model for a given pulsar (see also Chapters 2 and 4). The form of the timing model is a function giving the apparent rotational phase of the pulsar as a function of time:

$$\begin{aligned}\phi(t) &= \phi_0 + \nu t + \frac{1}{2}\dot{\nu}t^2 + \frac{\nu}{c}\left|\mathbf{r}_p - \mathbf{r}_e\right| + \cdots \\ &= \phi_0 + \nu t + \frac{1}{2}\dot{\nu}t^2 + \frac{\nu d}{c} - \frac{\nu}{c}\hat{\mathbf{n}}\cdot\mathbf{r}_e + \frac{\nu r_e^2}{2cd}\left(1 - \frac{1}{4}(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_e)^2\right) + \cdots\end{aligned} \tag{5.12}$$

Equation 5.12 is not meant to be exhaustive, but to illustrate many timing parameters common to all pulsars: ν is the intrinsic spin frequency of the pulsar; $\dot{\nu}$ is its intrinsic spindown rate; and $\mathbf{r}_p$ and $\mathbf{r}_e$ are position vectors to the pulsar and Earth relative to a inertial reference frame (for example the Solar system barycenter). The expansion of the vector distance shows terms that depend on the direction to the pulsar (sky position), and inversely on the distance (parallax). Additional terms not explicitly listed here include a description of any orbital motion of the pulsar, and relativistic delays due to the pulsar system and Solar system masses. In order to measure any GW induced timing fluctuations, we need to first subtract off these other effects.

While we know the motion of the Earth through the Solar system with high accuracy (see Standish, 1990), the pulsar's location and motion are determined from the timing measurements themselves. This is accomplished using a χ^2 fit of the data to a parameterized timing model. The differences between the data and the fitted model are known as the timing residuals, which we will analyze for the presence of GW. However, any GW signal that is degenerate with a timing model parameter will get absorbed by the fit, and hence not appear in the residuals. This process alters the frequency structure of the residuals and complicates the simple mapping of GW spectrum to residual spectrum. In order to make a meaningful GW measurement, this effect much be accounted for. Procedures for doing this have been a subject of much debate and uncertainty over the years.

A early analysis of the problem was done by Blandford et al. (1984) using a simplified analytic timing model, an approach that is very useful conceptually but not applicable to real data. Analyses of real data have usually proposed decomposing the timing residuals using sets of analytic basis functions that are approximately orthogonal to the fit functions (e.g., Kaspi et al., 1994). For example, cubic and higher order polynomials are orthogonal to the quadratic spin component in Equation 5.12, although they will have some overlap with the position terms. Recent analyses have dealt with this problem through Monte Carlo simulations (Jenet et al., 2006; Lommen et al., 2007), wherein artificial $h_{ij}(\mathbf{k})$ are computed based on assumed characteristics of the GW source population. These are converted to time shifts using Equation 5.8, and fed through the same fit procedure used on the actual data. The simulated residuals are then analyzed using a polynomial decomposition. The process is repeated to build up statistics on the results. While the simulation method will certainly give a correct answer (at least as far as the timing fit's effect is concerned – whether or not the simulated data is realistic is another issue), it does not lend much insight into what the fit is doing to the data.

We have developed a procedure that combines some of the best features of previous approaches. It is possible to compute the effect of the fit on a input signal directly from the fit design matrix, without reference to any particular data values. Since our approach incorporates the true timing model functions, it can be directly applied to analysis of real data. This method can be used to easily explore the effect of different timing model parameters without having to rely on Monte Carlo simulations. Furthermore, when combined with an assumed GW power spectrum, we can numerically compute a set of basis functions which are exactly orthogonal to the fit, and optimally capture the GW signal power. Our method also provides expected cross-correlation values between two sets of post-fit residuals, a entirely new result which will be important for the eventual detection of GW. The details of this method are presented in Appendix A, and we will apply it to the timing residuals of Chapter 4 in §5.5 below.

5.4 Stochastic Background

The dominant source of GW at nHz frequencies is expected to be the mergers of massive black holes (MBH) that occur over the history of the universe. These MBH binary systems are created following the merger of two host galaxies, each of which contains a MBH at its center. These MBHs "sink" towards each other via dynamical friction processes, eventually forming a tight binary that will radiate GW. The loss of energy to GW causes the orbit to decay, with the final result being the coalescence of the two black holes. The LISA project aims to detect the individual GW bursts caused by this final inspiral and coalescence. In the nHz frequency range, however, we are sensitive not to individual merger events, but to the stochastic background (GWB) created by the incoherent sum of many such events. The generation of GW through this process, and its dependence on cosmological and evolutionary parameters is explored in detail by Jaffe and Backer (2003). The resulting $h_{ij}(\mathbf{k})$ is expected to be uniform as a function of angle, and have a red power-law spectrum. Here we will present a mathematical description of the GWB and explore ways of measuring it using pulsar timing data: First using data from a single pulsar to place a upper limit on its amplitude; and second, correlating data from many pulsars to potentially detect it.

To begin, we can form a $h_{ij}(\mathbf{k})$ by summing over all MBH binary systems:

$$h_{ij}(\mathbf{k}) = \sum_{n=1}^{N_{BH}} A_n e^{i\phi_n} \left(e^{+}_{ij}(\hat{\mathbf{k}}_n) \cos 2\theta_n + e^{\times}_{ij}(\hat{\mathbf{k}}_n) \sin 2\theta_n \right) \delta(\mathbf{k} - \mathbf{k}_n) \tag{5.13}$$

Each MBH binary emits GW which have a strength (at Earth) of A_n, an apparent frequency k_n, and come from direction $\hat{\mathbf{k}}_n$. The polarization (θ_n) and phase (ϕ_n) angles depend on the orientation of the system. Equation 5.13 is based on several simplifying assumptions: First, that the binary system's evolution timescale is much longer than our observation time. This means we can take each binary's orbital frequency to be fixed, rather than changing as a function of time. Second, all systems have been assumed to be in circular orbits, which makes their GW emission monochromatic at twice the orbital frequency. Eccentric systems emit at higher harmonics as well.[3] Finally, we have assumed that all GW sources are far enough away that their emission can be treated as plane waves.

Following Jaffe and Backer (2003) we will approach this $h_{ij}(\mathbf{k})$ as a stochastic process: Each system in the sum has parameters which are drawn randomly and independently from a probability distribution $p(A, k)$. All angular parameters (θ, ϕ, $\hat{\mathbf{k}}$) are taken to be uniformly distributed (i.e., no preferred directions exist). From the construction of the background as a stochastic process, we can form the following expectation value:

$$\frac{1}{2}E\left\{h^*_{ij}(\mathbf{k})h_{ij}(\mathbf{k}')\right\} = N_{BH}\frac{\delta(\mathbf{k}-\mathbf{k}')}{4\pi k^2}\int A^2 p(A,k)dA = \frac{\delta(\mathbf{k}-\mathbf{k}')}{4\pi k^2}S_h(k) \qquad (5.14)$$

Equation 5.14 defines the strain power spectrum, $S_h(k)$, which has units of inverse frequency. To agree with the typical convention in the literature, we will define power spectra to be "one-sided," that is defined for positive frequencies only. A related quantity which is often used is the characteristic strain, $h_c(k) = \sqrt{kS_h(k)}$. Experimental limits on the stochastic background's strength are most often given as

[3]This simplification was also used by Jaffe and Backer (2003). Refer to that paper for detailed justification.

a fraction of the universe closure energy density (e.g., Jaffe and Backer, 2003, Eqn. 4):

$$\Omega_{GW}(\nu) = \frac{2\pi^2}{3H_0^2}\nu^3 S_h(\nu) \tag{5.15}$$

Since this expression depends on the Hubble constant H_0, GW limits are usually given as limits on $\Omega_{GW}h^2$, where the dimensionless number h is given by $H_0 = h \times 100$ km s^{-1} Mpc^{-1} $= h \times 1.02 \times 10^{-10}$ yr^{-1}. Out of these three quantities, we prefer to state limits in terms of h_c for the following reasons: Both S_h and h_c are directly measurable, while Ω_{GW} depends on cosmology. The numerical value of S_h will depend on the units of frequency, including a potentially ambiguous factor of 2π. In h_c, the units of frequency cancel out, leaving a single dimensionless number that characterizes the strain amplitude.

The key result of Jaffe and Backer (2003) is the prediction of a GW spectrum that has the shape $h_c \propto \nu^{-2/3}$, and an amplitude $\sim 10^{-15}$ at $\nu = 1$ yr^{-1}. The spectral slope of $-2/3$ is a robust result that has previously been noted by other authors as well (Rajagopal and Romani, 1995; Wyithe and Loeb, 2003; Enoki et al., 2004). This slope comes directly from two basic formulae regarding gravitational radiation from binary systems (e.g., Peters and Mathews, 1963; Misner et al., 1973): The RMS strain amplitude from a single binary system scales as $h_{rms} \propto \nu^{2/3}$, and the characteristic system evolution timescale due to GW losses scales as $\tau_{GW} \propto \nu^{-8/3}$. As the binary loses energy to GW, the orbit decays, increasing the GW frequency. Each system radiates more intensely during the high-frequency era of its life, however it spends more time radiating at lower frequencies. Since independent systems add in quadrature, for a stochastic background we get the previously stated result $h_c^2 \propto h_{rms}^2 \tau_{GW} \propto \nu^{-4/3}$. A measurement of this spectrum's amplitude would provide a mass-weighted estimate of the total number of MBH mergers throughout the Universe.

Another feature of the MBH merger spectrum is a low-frequency cutoff. As previously mentioned, dynamical friction[4] is necessary to reduce the size of the MBH binary orbit to a point where significant GW can be generated (Begelman et al., 1980). To put this another way, as long as $\tau_{GW} \gtrsim H_0^{-1}$, the binary will not merge in the lifetime of the Universe, and we will not observe GW from it. However, as long as the dominant energy loss mechanism is dynamical friction rather than GW, we will still not observe any GW from the system. The frequency at which GW emission takes over from dynamical friction will be the lowest observed GW frequency. Yu (2002) estimates this low-frequency cutoff as $10^{-4} - 10^{-2}$ yr^{-1} for many types of galaxy. A cutoff frequency that is much higher would reduce the likelihood of a pulsar GW detection (see §5.6). On the other hand, given a long enough data span, the spectrum turn-over could potentially be observed (or at least constrained) using pulsar data.

A second proposed source of nHz GW is cosmic strings (sometimes called cosmic superstrings). These theoretical objects are proposed one-dimensional remnants of phase transitions in the early Universe, and are predicted by some string theories (Damour and Vilenkin, 2005; Siemens et al., 2007; Hogan, 2006). Gravitational wave

[4] "Dynamical friction" refers to gravitational interaction between the MBH binary and other bodies, typically gas or large numbers of stellar-mass objects. These interactions tend to remove energy from the binary.

emission from cusps or kinks in these strings may form a stochastic GW background detectable by pulsar timing, with a predicted spectrum $h_c \propto \nu^{-7/6}$ (Damour and Vilenkin, 2005). Currently, pulsar timing offers the tightest experimental constraints on these theories, limiting the dimensionless string tension $G\mu$ to be less than $\sim 10^{-10}$ (Lommen et al., 2007; Hogan, 2006). However, the method by which pulsar timing data should be interpreted for these purposes is currently under debate (see Hogan (2006) for a discussion), as the string spectrum may be dominated by infrequent, high-amplitude bursts rather than continuous low-level emission as in the MBH case. Placing firmer limits on this type of background will require additional work on both the theoretical side and on the statistics used in data analysis.

To see the effect of a stochastic background on pulsar timing measurements, we can combine Equations 5.13 and 5.10 as follows:

$$\Delta(t) =$$
$$\sum_n \frac{iA_n e^{i\phi_n}}{2k_n} \left(\alpha_+(\hat{\mathbf{k}}_n, \hat{\mathbf{n}}) \cos 2\theta_n + \alpha_\times(\hat{\mathbf{k}}_n, \hat{\mathbf{n}}) \sin 2\theta_n\right) \left(1 - e^{-ik_n d - ik_{nl} n_l d}\right) e^{ik_n t} \quad (5.16)$$

In the frequency domain, $\Delta(\omega)$ has exactly the same form as Equation 5.16, with the final exponential $e^{ik_n t}$ replaced by $\delta(\omega - k_n)$. As a final remark, while we have phrased this discussion in terms of the MBH merger background, the equations are applicable to any set of sources that can be expressed as a sum of independent oscillators, and are as such very general.

5.4.1 Measurement using a Single Pulsar

It is possible to use timing data of a single pulsar to constrain the GW power spectrum. The starting point for this analysis is to compute the power spectrum of timing fluctuations, S_Δ:

$$E\left\{\Delta^*(\omega)\Delta(\omega')\right\} = \frac{N_{BH}}{3}\Phi(\omega d)\delta(\omega-\omega')\frac{1}{\omega^2}\int A^2 p(A,\omega)dA = \delta(\omega-\omega')S_\Delta(\omega) \quad (5.17)$$

The term $\Phi(x) = 1 - \frac{3}{2}\int_{-1}^{1}\mu^2\cos(x\mu)d\mu$ comes from integrating the polarization and phase terms over GW direction ($\hat{\mathbf{k}}$). It reduces the timing response at very small ω, for which the GW wavelength becomes larger than the pulsar distance (d). However, we will be working in the high-frequency regime where $\Phi(\omega d) \sim 1$.[5] We can combine Equations 5.14 and 5.17 to find the relation between the strain and timing power spectra:[6]

$$S_\Delta(\omega) = \frac{\Phi(\omega d)}{3\omega^2}S_h(\omega) \quad (5.18)$$

This simple relationship means that a measurement of the power spectrum of a pulsar's timing residuals, taking into account the effect of the timing fit discussed earlier, directly translates into a measurement of the GW spectrum. Alternately, we can assume a spectral shape (such as the expected $h_c \propto \nu^{-2/3}$), and measure the spectrum amplitude from the residuals. However, since there are many other sources of timing noise which may be present in the signal and indistinguishable from GW, this can at

[5] In principle, this is the only distinction between this experiment and LIGO or LISA. The interferometer experiments typically operate on GW wavelengths longer than their arm length.

[6] When the frequency ω in Equation 5.18 is expressed in typical units (i.e., Hz), a factor of $(2\pi)^{-2}$ must be included. Compare with Jenet et al. (2006) Eqn. 2.

best result in a upper limit on S_h.

5.4.2 Measurement using Several Pulsars

We can use Equations 5.14, 5.10 and 5.11 to compute the expected cross correlation between two sets of timing data:

$$\rho_{12}(\tau) = E\left\{\Delta_1(t_0-\tau)\Delta_2(t_0)\right\} =$$
$$N_{BH}\int \frac{A^2}{8k^2}p(A,k)\gamma(\hat{\mathbf{k}},\hat{\mathbf{n}}_1,\hat{\mathbf{n}}_2)\left(1-e^{ikd_1(1+\mu_1)}\right)\left(1-e^{-ikd_2(1+\mu_2)}\right)e^{ik\tau}\frac{dA\,d^3\mathbf{k}}{4\pi k^2} \quad (5.19)$$

Here, Δ_1 and Δ_2 are the timing fluctuations due to GW for two pulsars, which are located at positions $d_1\hat{\mathbf{n}}_1$ and $d_2\hat{\mathbf{n}}_2$ relative to the Earth. We have also defined $\mu_1 \equiv \hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_1$ and $\mu_2 \equiv \hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_2$. The angular factor γ is a function only of the orientation of $\mathbf{k}$ and the pulsar directions, and is given as:

$$\gamma(\hat{\mathbf{k}},\hat{\mathbf{n}}_1,\hat{\mathbf{n}}_2) = \alpha_+(\hat{\mathbf{k}},\hat{\mathbf{n}}_1)\alpha_+(\hat{\mathbf{k}},\hat{\mathbf{n}}_2) + \alpha_\times(\hat{\mathbf{k}},\hat{\mathbf{n}}_1)\alpha_\times(\hat{\mathbf{k}},\hat{\mathbf{n}}_2) \quad (5.20)$$

When the terms inside the parentheses in Equation 5.19 are multiplied out, there are four factors. Three of these are exponentials that are rapidly oscillating (since $kd \gg 1$). These will average out to zero when the k integral is performed. It is the first, constant term that is responsible for the correlation. This represents the contribution from the "Earth end" of the time shift integral for each pulsar. When the three oscillatory cross-terms are dropped, the angular and radial parts of the integral separate, and we are left with the following:

$$\begin{aligned} \rho_{12}(\tau) &= \frac{N_{BH}}{8} \int A^2 p(A,k) e^{ik\tau} dAdk \int \gamma(\hat{\mathbf{k}}, \hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2) \frac{d\Omega_k}{4\pi} \\ &= \frac{1}{2} \int S_\Delta(\omega) e^{i\omega\tau} d\omega \int \gamma(\hat{\mathbf{k}}, \hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2) \frac{3d\Omega_k}{4\pi} \end{aligned} \tag{5.21}$$

Here the second line comes from inserting the definition of S_Δ from Equation 5.17, and relabelling k as ω. The first integral is the inverse Fourier transform of $S_\Delta(\omega)$, otherwise known as the autocorrelation function $C_\Delta(\tau)$. The angular integral is much more complicated and will result in the expected correlation between the two sets of time shifts depending on the angular separation of the two pulsars. We will call this function $\zeta(\beta)$, where $\beta_{12} \equiv \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2$ is the cosine of the 1–2 angular separation:

$$\rho_{12}(\tau) = \frac{1}{2} C_\Delta(\tau) \zeta(\beta_{12}) \tag{5.22}$$

This normalization sets $\zeta(1) = 1$, and highlights the fact that the correlated power is at most 1/2 of the total GW power in a pulsar signal, due to the "pulsar term" of the time shift being lost in the cross-correlation. A evaluation of the angular integral was performed by Hellings and Downs (1983, HD), giving the following result:

$$\frac{\zeta(\beta)}{3} = \frac{1-\beta}{2} \log(\frac{1-\beta}{2}) - \frac{1-\beta}{12} + \frac{1}{3} \tag{5.23}$$

This function is plotted in Figure 5.2. Demonstrating a correlation of timing residual pairs versus separation angle that has this shape would be a unambiguous detection of GW. This method of combining data from many pulsars is known as a pulsar timing array.

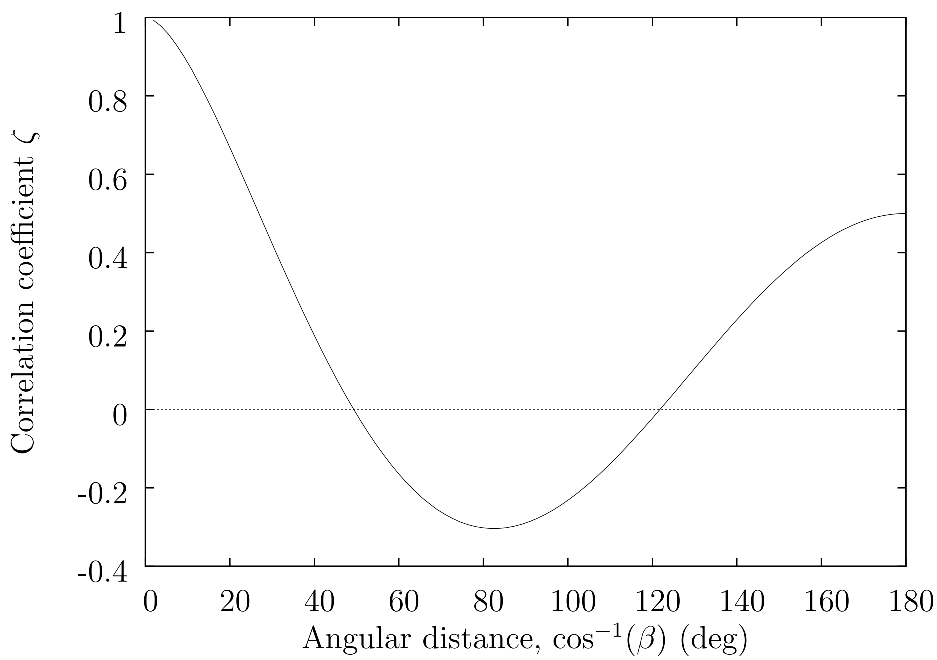

Figure 5.2: Expected timing correlation as a function of angular separation for a pair of pulsars.

5.5 Analysis of Timing Data

In this section, we use the timing results presented in Chapter 4, and the analysis techniques developed in this chapter to place upper limits on the stochastic gravitational wave background. This is divided into two sections, corresponding to the single-pulsar and multiple-pulsar cases discussed above. A summary of past results is presented first.

Since the original HD paper, all published GW pulsar timing limits have been done using the single-pulsar method. A summary of these results is presented in Table 5.1. The HD limit was made prior to the discovery of millisecond pulsars (MSPs), using data from slow (or "normal") pulsars,[7] and as such is much less restrictive than the later single-pulsar MSP results, due to the factor of $\sim 10^3$ improvement in timing accuracy.

While we have been referring to these as "single-pulsar" limits, they are sometimes in fact results of data from several pulsars. More accurate names for the two methods might be "incoherent" versus "coherent" for the cross-correlation analysis. In the incoherent method, the residual power spectrum of each pulsar is measured separately, and these results can then be combined into a single limit on S_h, with weights depending on the quality of the different pulsar timing records. However, the final answer is almost always dominated by the one best pulsar in the set, typically PSR B1855+09, since it is very stable on long time scales. Recently, better results

[7] In fact, the first MSP was discovered only one month after the HD paper was submitted (Backer et al., 1982).

Source(s)	C/I	T (yr)	$\Omega_{GW}(1/T)h^2$	$h_c(1\ \mathrm{yr}^{-1})$	Reference
B1133+16, B1237+25, B1604–00, B2045–16	C	12	$< 1 \times 10^{-4}$	$< 9.1 \times 10^{-13}$	Hellings and Downs (1983)
B1855+09, B1937+21	I	8	$< 6 \times 10^{-8}$	$< 1.9 \times 10^{-14}$	Kaspi et al. (1994)
B1855+09, ...	I	8	$< 2 \times 10^{-8}$	$< 1.1 \times 10^{-14}$	Jenet et al. (2006)
J1713+0747, B1855+09	I	20	$< 2 \times 10^{-9}$	$< 4.9 \times 10^{-15}$	Lommen et al. (2007)

Table 5.1: Summary of previous pulsar GW limits. The columns list the timespan of the data used (T), and the resulting $\Omega_{GW}(1/T)$ limit. We have converted these to $h_c(1\ \mathrm{yr}^{-1})$ assuming a $h_c \propto \nu^{-2/3}$ spectrum. C = coherent (multi-pulsar) method, I = incoherent (single-pulsar) method.

have been obtained by focusing on J1713+0747 (Lommen et al., 2007). The original MSP, B1937+21, is very bright and fast, and therefore allows very precise timing, but shows significant instability on long time scales (Kaspi et al., 1994), reducing its utility for low-frequency GW detection.

5.5.1 Incoherent Analysis

In this analysis, we do not have a long enough span of data to get single-pulsar limits comparable to those just listed. However, we will present an analysis of 2.2 years of data from the best pulsar presented in Chapter 4, J1713+0747, as a demonstration of the new methods developed here. Applying these concepts to the long existing datasets on several pulsars will be a topic of future work.

As mentioned previously, there has historically been a large amount of uncertainty and debate regarding how to translate pulsar timing data into GW limits (see Kaspi et al., 1994; Thorsett and Dewey, 1996; Lommen, 2001; Jenet et al., 2006). The problems arise from the fact that the timing model fit, discussed earlier, alters the statistics of the GW signal in non-obvious ways. We have developed a method that takes this into account, and will hopefully simplify future analyses. Our method, described in detail in Appendix A, combines an assumed timing power spectrum $S_\Delta(\omega)$ or autocorrelation function $C_\Delta(\tau)$ together with the fit design matrix. The result is a set of optimal orthonormal basis functions (eigenvectors or components, $v_j^{(i)}$) with which to decompose the timing residuals (r_j), and an associated set of eigenvalues (λ_i) that give the expected GW power in each basis function:

$$\begin{aligned} r_j &= \sum_i c_i v_j^{(i)} \\ E\left\{c_i^{(GW)2}\right\} &= A^2 \lambda_i \end{aligned} \tag{5.24}$$

The factor A gives the amplitude of h_c, which we want to estimate. These are illustrated in Figures 5.3 and 5.4 for our timing data on J1713+0747. We have assumed a GW power spectrum of the form $S_\Delta(\omega) \propto \omega^{-13/3}$, corresponding to the expected $h_c \propto \omega^{-2/3}$ MBH merger prediction.

For this data, we can see that nearly all the expected GW power is contained in the first eigenvector. With longer datasets, several eigenvectors may be usable (see plots in Appendix A). Furthermore, over long time spans, other red noise processes such as intrinsic pulsar timing noise may appear in the data. This will increase the measured values in the low eigenvectors above what is expected from measurement

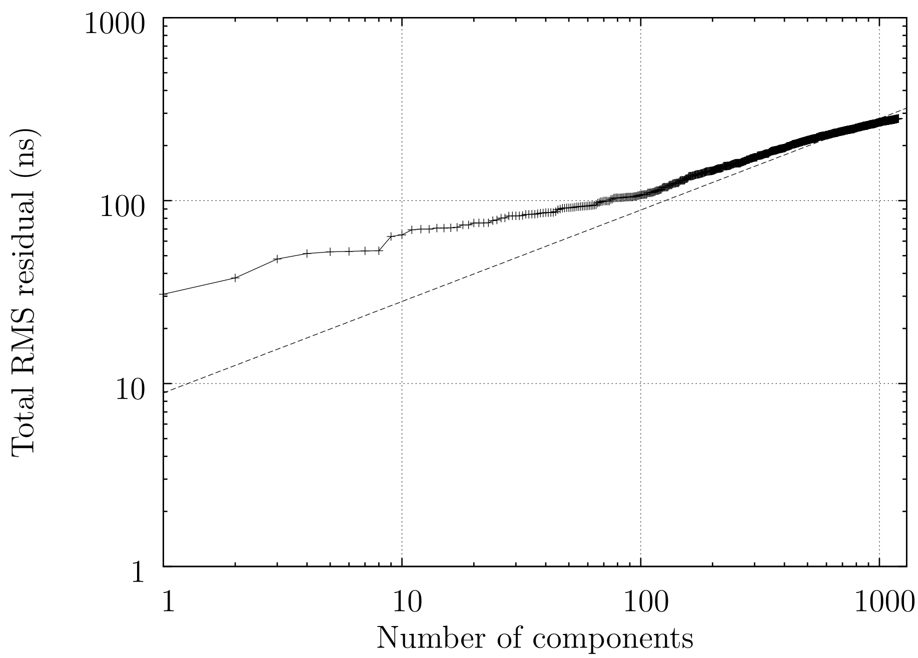

Figure 5.3: Cumulative RMS residual versus number of components used to decompose the timing residuals of J1713+0747. The value reaches the total weighted RMS residual, 0.26 μs, at the right hand side. The line is proportional to $\sqrt{N}$, the expected behavior of white noise.

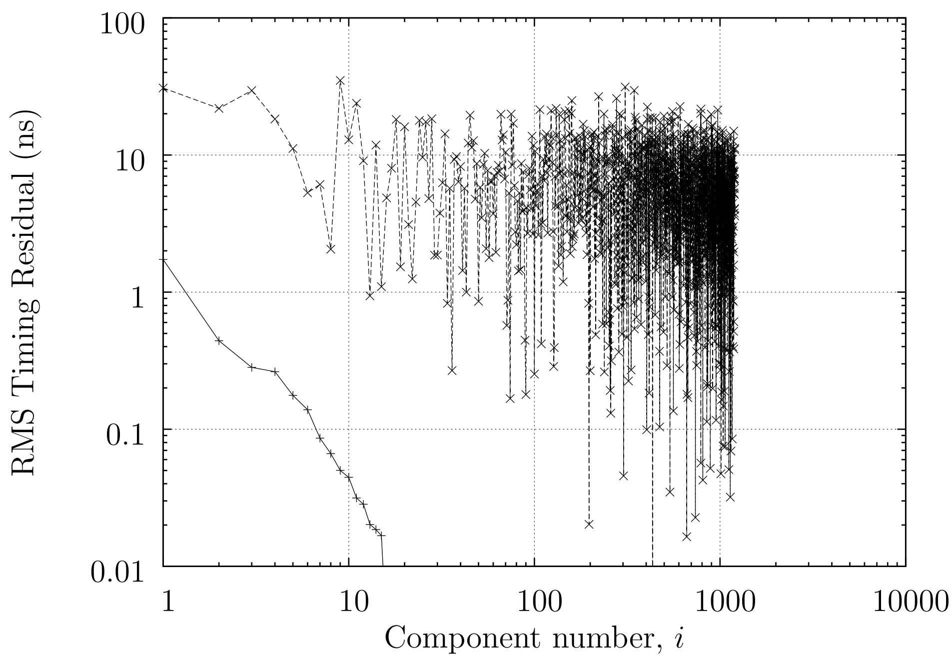

Figure 5.4: RMS timing residual due to each component. The top line shows the measured $|c_i|$ from J1713+0747. The bottom line shows the expected values $A\lambda_i^{1/2}$ for a $h_c(\nu) = 10^{-15}(\nu/1\ \mathrm{yr}^{-1})$ GWB.

noise alone. In this case, a better limit may be obtained by discarding or down-weighting these values. This framework should provide a useful tool with which to explore these issues. For our purposes here, however, the only remaining question is how to convert our one measured value and expectation into a statistical statement about the GW amplitude.

Each measured coefficient c_i is a sum of several contributions, which we will categorize as GW and "noise": $c_i = c_i^{(GW)} + c_i^{(N)}$. Each of these contributions is drawn from its own probability distribution, of which we have varying amounts of knowledge. We can safely assume that both distributions are zero-mean. We also know the variance of $c_i^{(GW)}$ as a function of i, up to a overall scale factor A that we want to estimate. So we would like to determine the probability distribution $P(A|c_i)$, that is the probability of A having a certain value given our measurement(s) of c_i. Unfortunately in this situation, we are lacking a large amount of other necessary information: We don't have a explicit form, or higher order moments, for the GW source distribution $p(A, k)$ discussed earlier, although this could potentially come from further theoretical work. More importantly, we can assume almost nothing about the noise distribution. We may be able to estimate the measurement noise contribution,[8] but given various timing systematics and intrinsic pulsar timing noise, the statistics at low i are most likely different from those at high i where "white" measurement noise dominates.

[8]Refer to Jenet et al. (2006) and Lommen et al. (2007) for a clever technique on how to do this.

In this uncertain situation, we propose the following simple analysis: Assume both $c^{(GW)}$ and $c^{(N)}$ are drawn from zero mean normal distributions with variance $A^2\lambda$ and σ_N^2 respectively. Then c will also be normally distributed with variance $\sigma_c^2 = A^2\lambda + \sigma_N^2$. Given our measured value c_0, we can say with 95% confidence that $\sigma_c < 2c_0$. Formally this comes from assuming a uniform prior for $\log(\sigma_c)$ (i.e., no preferred scale) and applying Bayes' theorem. Since $A^2\lambda < \sigma_c^2$, this also means our limit on A becomes:

$$A < \frac{2c_0}{\lambda_0^{1/2}} \tag{5.25}$$

For our measured values from J1713+0747, this translates to $h_c(1\ \mathrm{yr}^{-1}) < 3.5\times10^{-14}$. The result can also be expressed as $h^2\Omega_{GW}(1/2.2\ \mathrm{yr}) < 4.5\times10^{-7}$. This method can be extended to the case of multiple c_i, potentially from multiple pulsars, by defining a separate noise variance $\sigma_{N,i}^2$ for each c_i, then marginalizing over the $\sigma_{N,i}^2$ parameters.

5.5.2 Coherent Analysis

We will now combine all the timing data from Chapter 4 in a cross-correlation analysis. The basic plan for the analysis is to estimate a cross-correlation amplitude and uncertainty for each pair of pulsars. We then fit these values versus separation angle to a scaled Equation 5.23. The scale factor can be converted to a limit on the GWB amplitude. This approach is fundamentally different from that of the last section in that it is possible for a significant detection to occur, as there are no other effects expected to exactly mimic the GW angular correlation function.

Given sets of timing residuals $r_i^{(a,b)}$ for pulsars a and b, and associated measurement uncertainties $\sigma_i^{(a,b)}$, we estimate the cross-correlation by summing over all pairs of data points as follows:

$$\rho_{ab} = \sum_{i,j} \frac{r_i^{(a)} r_j^{(b)} S_{ij}}{(\sigma_i^{(a)} \sigma_j^{(b)})^2} \left(\sum_{i,j} \frac{S_{ij}^2}{(\sigma_i^{(a)} \sigma_j^{(b)})^2} \right)^{-1} \tag{5.26}$$

Here, S_{ij} is the expected correlation value for each pair of points (i, j), computed from the assumed GW power spectrum and fit design matrices (see §A.4). This includes only the time part of the correlation ($\frac{1}{2}C_\Delta(\tau)$ term), not the angular dependence. When expressed in this way, the measured values of ρ_{ab} should have the form:

$$\rho_{ab} = A^2 \zeta(\beta_{ab}) \tag{5.27}$$

Again, A gives the amplitude of h_c, relative to the reference amplitude used in the definition of the S_{ij}. In our case, we choose to normalize A so that $h_c(\nu) = A \times 10^{-15}(\nu/1\ \mathrm{yr}^{-1})^{-2/3}$. The uncertainty of each ρ_{ab} point, σ_{ab}, is computed from the spread of the individual terms in the sum of Equation 5.26 about their mean value. These ρ_{ab}, σ_{ab} pairs are then fit to Equation 5.27 to determine the value of A^2, using a standard χ^2 procedure.

The results of this fit are shown in Figure 5.5. The derived values are $A^2 = 92.8$ and $\sigma_{A^2} = 278.3$. The fit reduced-χ^2 value of 13.2 has been incorporated into the value of σ_{A^2}. Since this value is consistent with $A^2 = 0$, and $A^2 < 0$ is unphysical, we will convert it into a upper limit on A^2 by determining A^2_{max} such that 95% of the inferred normal distribution for A^2 falls between 0 and A^2_{max}. Essentially, this means truncating the distribution at $A^2 = 0$ then re-normalizing it. This results in a 95%

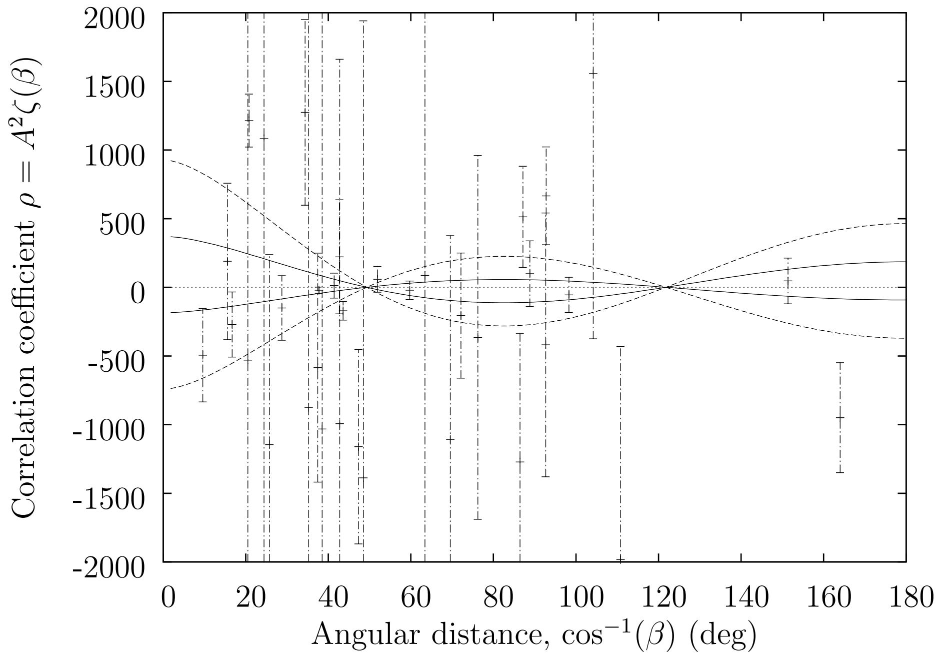

Figure 5.5: Measured cross-correlation value versus angular separation for the set of millisecond pulsars. The lines show ± 1- and ± 3-σ deviation from the derived value of A^2. In this plot, $A = 1$ is the expected value for a $h_c(\nu) = 10^{-15}(\nu/1\ \mathrm{yr}^{-1})^{-2/3}$ GWB. Points with uncertainty much larger than the plot scale are not shown here.

confidence limit of $A^2 < 608.2$, which corresponds to an upper limit on $h_c(1\ \mathrm{yr}^{-1})$ of 2.46×10^{-14}, or $h^2\Omega_{GW}(1/2.2\ \mathrm{yr}) < 2.2 \times 10^{-7}$.

5.6 Discussion

This limit is somewhat better than we were able to do with J1713+0747 alone using the same span of data. However, we can see from Figure 5.5 that almost all the weight for the final result comes from a handful of data points. These points come from correlations involving the few best pulsars in the set: J1713+0747, B1937+21, and J1909−3744. This will almost always be the case as long as we are in a regime of placing limits rather than making a detection. Some of the issues involved in actual detection were explored by Jenet et al. (2005), who conclude that it is important to keep observing pulsars covering a wide range of angular separations, rather than just focusing on the best few. We tend to agree: For a detection to be believable, it would need to show the predicted behavior over the full range of the angular separation plot. Furthermore, the "best" pulsars are also typically the ones whose timing is not signal-to-noise limited, and so allocating more intense observations to these may not bring the expected improvement unless systematic errors are able to be reduced (see Chapter 4).

However, the most important factor needed to improve these limits is simply time. All proposed sources of nHz GWB are expected to have a red power spectrum. Since the timing fit filters out frequencies less than $1/T$, this means the transmitted GW power will increase as the observation span grows. Over our ∼2.5 years of data, the RMS GW-caused residual was predicted to be only 1–3 ns (this varies per each object due to the slightly different observation pattern and timing models). For the MBH GW spectrum, this number will grow like $T^{5/3}$. We can use this dependence, and the measured values from the previous section to predict how our h_c limit will improve versus time. This is shown in Figure 5.6. If timing errors integrate down versus time as expected for white noise, we will gain an additional $T^{-1/4}$ in sensitivity.[9] A slightly more pessimistic estimate assumes that systematic errors will prevent us from achieving this, and does not include this factor.

From Figure 5.6, we can see that we expect to meet the current best 20-year limit with the addition of 2-4 years more data on this same set of pulsars. Getting to the expected detection regime ($h_c < 10^{-15}$) will take another 10 years. However, there are ways to increase T aside from simply taking more data. Timing records exist for many MSPs covering the last 5–10 years, and extend up to 20 years for a few. The number of pulsars declines the further back in time one looks, and the quality of data tends to decline as well, due to advances in instrumentation made along the way. Detecting GW will most likely come from combining this large, heterogeneous data set into a single measurement. This is a much larger task than the analysis presented here, as it involves connecting timing data from multiple instruments and observatories. Furthermore, while most pulsar data is "publicly available in principle," it is in fact not

[9] This is not $T^{-1/2}$ because h_c is an amplitude rather than a power measurement.

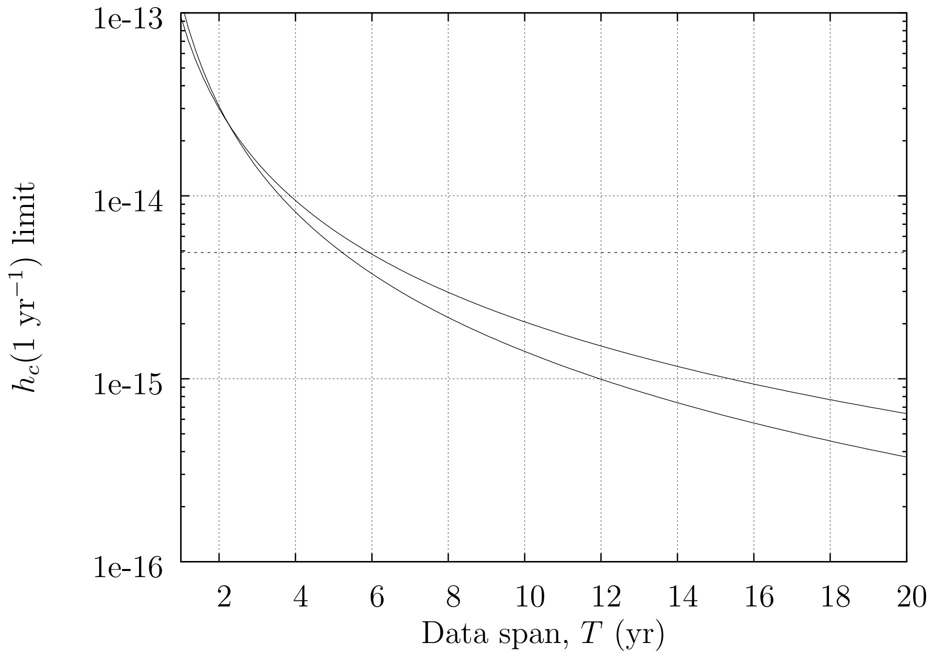

Figure 5.6: Expected pulsar timing limit on $h_c(1\ \mathrm{yr}^{-1})$ as a function of time. The lower curve is based on noise integrating down, while the upper only includes the GW power growth versus time. These are normalized to the measured limit from §5.5.2. The horizontal line shows the current best experimental limit, using 20 yr of data on B1855+09 (Lommen et al., 2007).

easily accessible and is stored in a large variety of formats at various sites. However, given the interest in solving this problem, several collaborations have recently been formed to work out these issues. These include the Australian Parkes Pulsar Timing Array (PPTA) group, the European PTA (EPTA) group, and most recently, our own USPTA group. Current prospects appear good for achieving significant improvements over the next few years.

Chapter 6

A Long-Term Timing Study of PSR B1937+21

6.1 Introduction

The dispersion measure (DM) of a pulsar probes the column density of free electrons along the line of sight (LOS).[1] Observed DM variations over time scales of several weeks to years sample structures in the electron plasma over length scales of 10^{10} to 10^{12} m. Diffraction of pulsar signals is the result of scattering by structures on scales below the Fresnel radius, $\sim 10^8$ m. The DM as well as the scattering measure (SM) variability along the LOS to the Crab pulsar was first reported by Isaacman

[1] A version of this chapter was previously published as "Interstellar Plasma Weather Effects in Long-term Multifrequency Timing of Pulsar B1937+21" by R. Ramachandran, P. Demorest, D. C. Backer, I. Cognard, and A. Lommen, in *The Astrophysical Journal*, Volume 645, pp. 303–313, July 2006.

and Rankin (1977), who reported that the DM variability poorly correlated with the SM variability. Helfand et al. (1980) inferred an upper limit for DM variations of a few parts per thousand for several pulsars. In an earlier study of PSR B1937+21, Cordes et al. (1990) measured a DM change of $\Delta\mathrm{DM} \sim 0.003$ pc cm^{-3} over a period of 1000 days. The work of Phillips and Wolszczan (1991) presented the variations of DM observed along the LOS to a few pulsars. They connected these variations to those on diffractive scales, and derived an electron density fluctuation spectrum slope of 3.85 ± 0.04 over a scale range of $10^7 - 10^{13}$ m. Backer et al. (1993) report on further DM variability and show that the amplitude of the variations known at that time are consistent with a scaling by the square root of DM. Another important investigation by Kaspi et al. (1994) studied DM variations of the millisecond pulsars PSR B1937+21 and B1855+09 during calendar years 1984-1993. In addition to establishing a secular variation in DM over this time interval, they show that the underlying density power spectrum has a spectral index of 3.874±0.011, which is close to what we would expect if the density fluctuations are described by Kolmogorov turbulence. An "anomalous" dispersion event towards the Crab pulsar was reported by Backer et al. (2000), in which they report a DM jump as large as 0.1 pc cm^{-3}.

In this chapter, we present results of several long term monitoring programs on PSR B1937+21. Our data, which includes that of Kaspi et al. (1994), spans calendar years 1983-2004. These data sets have been taken with four different telescopes,

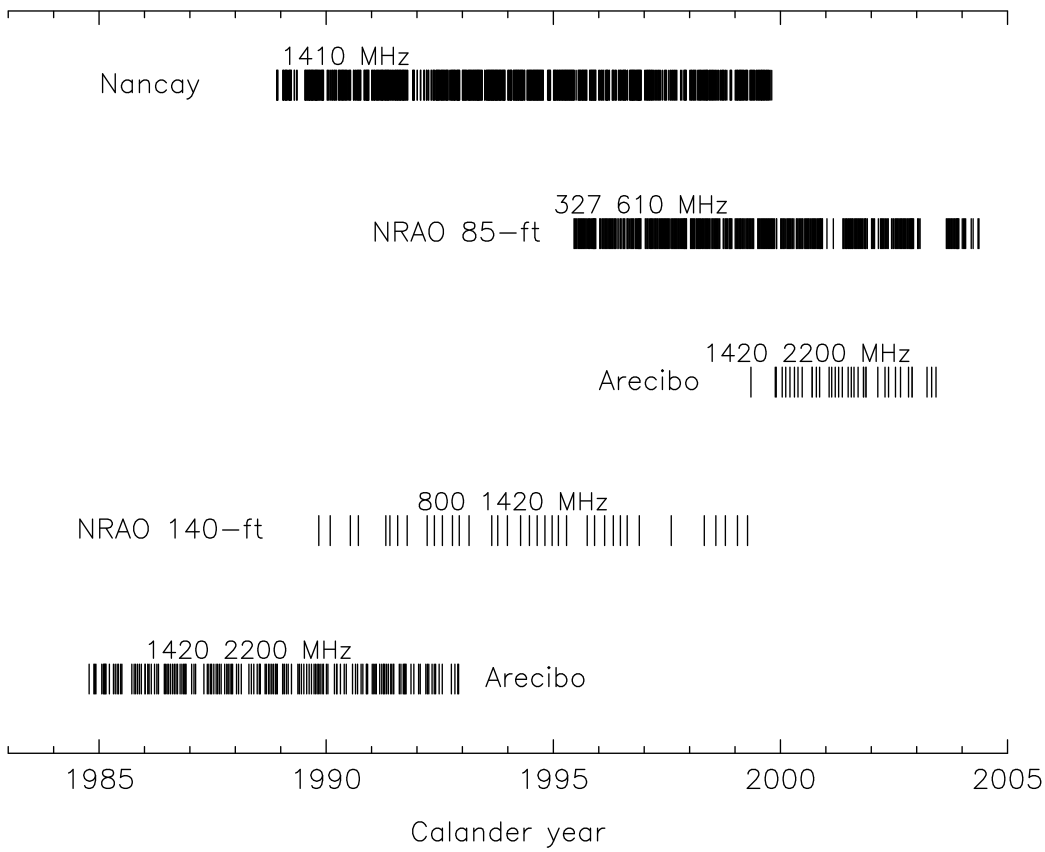

Figure 6.1: Summary of our data sample. See text for details.

the NRAO[2] Green Bank 42 m (140 foot) and 26 m (85 foot) telescopes, the NAIC[3] Arecibo telescope and the Nançay telescope, at frequency bands of 327, 610, 800, 1400 and 2200 MHz. After giving the details of our observations in §6.2, we describe our analysis methods in §6.3. This is followed in §6.4 by a discussion of the distribution of scattering material along the LOS. As we describe, the knowledge of temporal and angular broadening of the source, proper motion, and scintillation-based velocity estimates enables us to at least qualitatively study the distribution of scattering matter as well as properties of its wavenumber spectrum.

We have measured some of the basic refractive scintillation parameters from our observations, and these are discussed in §6.5. The frequency dependence of the refractive scintillation time scale and the modulation index indicate a caustic-dominated regime that results from a large inner scale in the spectrum.

We have detected DM variations as a function of time and frequency. We determine the phase structure function of the medium with the knowledge of the time dependent DM variations, which is consistent with a Kolmogorov distribution of density fluctuations between scale sizes of about 1 and 100 A.U. These are summarized in §6.6 and §6.7.

PSR B1937+21 is known for its short term timing stability. However, the achievable long term timing accuracy is suspected to be seriously limited by the interstellar

[2]The National Radio Astronomy Observatory (NRAO) is owned and operated by Associated Universities, Inc under contract with the US National Science Foundation.

[3]The National Astronomy and Ionosphere Center is operated by Cornell University under contract with the US National Science Foundation.

scattering properties. With our sensitive measurements, we are in a position to quantify these errors. In §6.8, we describe in detail various sources of these errors and quantify them. The quantitative study of these and other types of systematic errors present in pulsar timing data is important for assessing the future of precision pulsar timing. In particular, the likelihood of achieving the first direct detection of gravitational radiation (see Chapter 5) will depend on at what level these systematic effects dominate timing results. Further study of these effects in the set of "best" millisecond pulsars will be a topic of future work.

6.2 Observations

We have used five different primary data sets for this analysis. The first set is the 1984–1992 Arecibo pulse timing and dispersion measurements obtained by Kaspi, Taylor, and Ryba (1994, hereafter KTR94). Their observations were performed with their Mark II backend (Rawley, 1986; Rawley et al., 1988) and later their Mark III backend (Stinebring et al., 1992) at two different radio frequency bands, 1420 and 2200 MHz. Their analysis methods are described in KTR94.

The second data set is from 800 and 1400 MHz observations at the NRAO 140 foot telescope in Green Bank, WV. The Spectral Processor backend, a hardware fast Fourier transform (FFT) device, was used. Details of the observations and analysis are contained in an earlier report on dispersion measure variability (Backer et al., 1993).

The third data set consists of observations at 327 and 610 MHz using the 26 m (85 foot) pulsar monitoring telescope at NRAO's Green Bank site. Room temperature (uncooled) receivers at the two bands are mounted off-axis. At 327 MHz the total bandwidth used was 5.5 MHz, and 16 MHz was used at 610 MHz. The two orthogonally polarized signals were split into 32 frequency channels in a hybrid analog/digital filter bank in the GBPP (Green Bank–Berkeley Pulsar Processor). Dispersion effects were removed in the GBPP in real-time with a coherent (voltage) deconvolution algorithm. At the end of the real-time processing, folded pulse profiles were recorded for each frequency channel and polarization. Further details of the backend and analysis can be found in Backer et al. (2000). PSR B1937+21 was observed for about 2 h per day starting in mid-1995.

The fourth data set comes from a bi-monthly precision timing program that includes B1937+21 at the Arecibo Observatory, which we started in 1999 after the telescope upgrade. Signals at 1420 and 2200 MHz were recorded using the Arecibo–Berkeley Pulsar Processor (ABPP) backend, which is identical to the GBPP. Our typical observing sessions at 1420 and 2200 MHz had bandwidths of 45 MHz and 56 MHz, respectively, and integration times of approximately 10 minutes per session.

The fifth data set is from a pulsar timing program that has been ongoing since October 1989 with the large decimetric radio telescope located at Nançay, France. The Nançay telescope has a surface area of 7000 m^2, which provides a telescope gain of 1.6 K Jy^{-1}. Observations are performed with dual linear feeds at frequencies

1280, 1680 and 1700 MHz. Then the signal is dedispersed by using a swept frequency oscillator (at 80 MHz) in the receiver intermediate frequency (IF) chain. The pulse spectra are produced by a digital autocorrelator with a frequency resolution of 6.25 kHz. Cognard et al. (1995) describe in detail the backend setup and the analysis procedure.

A small amount of additional data from the Effelsberg telescope was used in our profile analysis. At Effelsberg the EBPP backend, a copy of the GBPP/ABPP, was used.

6.3 Basic Analysis

We first present several results from the analysis of these data sets: a description of the frequency-dependent profile template used for timing; spin and astrometric timing parameters from high frequency data; and pulse broadening, flux densities and dispersion measure as functions of time. In §6.4 we proceed to interpret these results, then return to finer details regarding dispersion measure variations in §6.6.

Our basic data set consists of average pulse profiles obtained approximately every 5 minutes in each of the radio frequency bands, 327, 610, 800, 1420 and 2200 MHz. Figure 6.1 provides a graphical summary of observation epochs versus date. For data sets corresponding to all frequencies except 327 MHz, times of arrival (TOAs) were computed by cross-correlating these average profiles with a template profile. The template profile at a given frequency was made by using multiple Gaussian fits to very high signal-to-noise ratio averages at that frequency; the interactive program BFIT, which is based on M. Kramer's original program FIT was used. These fit parameters are listed in Table 6.1. Column (1) in the Table gives the radio frequency and the backend name is in column (2). Column (3) gives the width of component 1 (w_1; its location is taken to be 0° and its amplitude is set to 1.0); columns (4)-(6) and columns (7)-(9) give the location (l), width (w) and amplitude (h) values for components 2 and 3, respectively. The location and width are given in units of longitudinal degrees, where 360° indicates one full rotation cycle. The results of this analysis can be compared with that of Foster et al. (1991) which are given on the line at 1000 MHz [4] There is reasonable agreement for all values except h_2, which we suspect was erroneously entered in Table 4 of Foster et al. (1991). In our analysis, templates corresponding to arbitrary frequencies are produced by spline interpolation of the component parameters as functions of frequency.

We used the Arecibo (1420 and 2200 MHz) TOAs, and the GBT 140 foot (800 and 1420 MHz) TOAs to fit for pulsar spin (rotation frequency f, first time derivative $\dot{f}$, and second time derivative $\ddot{f}$) and astrometric (position R.A. and Decl., proper motion $\mathrm{PM}_{\mathrm{R.A.}}$, μ_α, along right ascension, and $\mathrm{PM}_{\mathrm{Decl.}}$, μ_δ, along declination) parameters. All TOAs were referred to the UTC time scale kept by the National Institute of Standards and Technology (NIST) via GPS satellite comparison. We removed the effects of variable dispersion from this fitting procedure with weekly estimation of DMs

[4] The widths w_1 and w_3 are inverted in Table 4 of Foster et al. (1991).

ν (MHz) (1)	Backend (2)	w_1 (deg) (3)	l_2 (deg) (4)	w_2 (deg) (5)	h_2 (rel) (6)	l_3 (deg) (7)	w_3 (deg) (8)	h_3 (rel) (9)
327	GBPP	8.7	0	0	0	186.9	12.0	0.51
430	ABPP	10.3	7.0	2.5	0.19	186.8	12.4	0.53
610	GBPP	9.4	7.1	2.9	0.34	187.3	10.8	0.60
863	EBPP	8.9	7.7	5.2	0.38	187.7	10.9	0.55
1000[a]	SP	9.2	8.6	3.9	0.56	187.9	11.3	0.54
1419	ABPP	8.5	8.5	3.7	0.37	187.5	10.1	0.45
1689	EBPP	9.1	9.2	3.9	0.37	187.4	10.5	0.40
2200	GBPP	9.4	9.8	3.2	0.29	187.6	10.4	0.36
2379	ABPP	10.0	9.8	3.0	0.29	187.1	10.8	0.33

Table 6.1: Template fit parameters at various frequencies.

[a] From Foster et al. (1991).

and subsequent extrapolation of the dual frequency data to infinite frequency prior to parameter estimation. The nature of achromatic timing noise makes it particularly difficult to determine a precise timing model. As one adds additional higher derivatives of rotation frequency (e.g., a third derivative), the best fit parameters change by amounts much larger than the nominal errors reported by the package that we used, TEMPO. The results are listed in Table 6.2. The errors presented in the table incorporate the range of variation of each parameter as additional derivative terms are included. In comparison to KTR94, the derived proper motion values are marginally different. We attribute this difference to the variable influence of timing noise. An important point that needs to be stressed here is that there is no reason for us to assume that the higher derivative terms of rotation period (e.g., $\ddot{f}$ or higher) have anything to do with the radiative braking index. They are most likely dominated by

Parameter	Value
PSR	1937+21
R.A. (J2000)	$19^{\mathrm{h}}39^{\mathrm{m}}38\overset{\mathrm{s}}{.}561$ (1)
Decl. (J2000)	$21^{\circ}\ 34'59\overset{''}{.}136$ (6)
$\mathrm{PM}_{\mathrm{R.A.}}$(mas yr^{-1})	0.04 (20)
$\mathrm{PM}_{\mathrm{Decl.}}$(mas yr^{-1})	-0.45 (6)
f (Hz)	641.9282626021 (1)
$\dot{f}$ (10^{-15} Hz s^{-1})	-43.3170 (6)
$\ddot{f}$ (10^{-26} Hz s^{-2})	1.5 (3)
PEPOCH^a (MJD)	47500.000000
START^b (MJD)	45985.943
FINISH^c (MJD)	52795.286
EPHEM^d	DE405
CLK^e	UTC (NIST)

Table 6.2: Timing model parameters for PSR B1937+21.

^{a} Reference date for position and frequency.
^{b} Fit start date.
^{c} Fit end date.
^{d} Solar system ephemeris version used (JPL).
^{e} Clock scale used.

some intrinsic instabilities of the star itself, or some other perturbation on the star.

Extension of dispersion measurement to 327 MHz requires removal of the time-variable broadening of the intrinsic pulse profile owing to multipath propagation in the interstellar medium. We deconvolved the effect of interstellar scattering following precepts first introduced by Rankin et al. (1970). We assume that the interstellar temporal broadening is quantified in terms of convolution of a Gaussian function and a truncated exponential function. If there is only one scattering screen along the LOS, the assumption of a truncated exponential function will suffice to represent the scatter broadening. However, since the scattering may arise from material distributed all along the LOS, a more realistic representation is approximated by a truncated exponential function "smoothed" (convolved) with a Gaussian function. The intrinsic pulse profile was estimated by extrapolation of parameters from the higher frequency profiles. In the deconvolution procedure, we minimized the normalized χ^2 value by varying the width of the Gaussian w_g and the decay time scale of the truncated exponential τ_e, while keeping the intrinsic pulse profile fixed. The pulse scatter broadening is quantified as $\tau_{\rm sc} = (w_g^2 + \tau_e^2)^{1/2}$. We repeated this for average profiles obtained at every epoch to obtain the $\tau_{\rm sc}$ measurement. In our fits, the average value of w_g came to about 74 μs, whereas the corresponding value for τ_e was about 85 μs. The measurement of $\tau_{\rm sc}$ versus time at 327 MHz is plotted in Figure 6.2. This quantity has a mean value of 120 μs, an RMS variation of 20 μs, and a fluctuation timescale of $\sim$ 60 days. We explain these variations as the result of refractive modulation of this inherently diffractive parameter in discussion below. The estimated RMS variation at the next higher frequency in our data set, 610 MHz, is $\sim$2.5 μs, using a frequency dependence of $\tau_{\rm sc} \propto \nu^{-4.4}$. This is too small to allow fitting at this frequency band.

In the strong scintillation regime, time dependent variations in the observed flux occur in two distinct regimes, *diffractive* and *refractive*. The diffractive effects are dominated by structures smaller than the Fresnel scale, and appear on short time scales and over narrow bandwidths. In our observations diffractive modulations are strongly suppressed due to integrating over larger bandwidths and timescales. On the other hand refractive effects occur on timescales of days and are correlated over

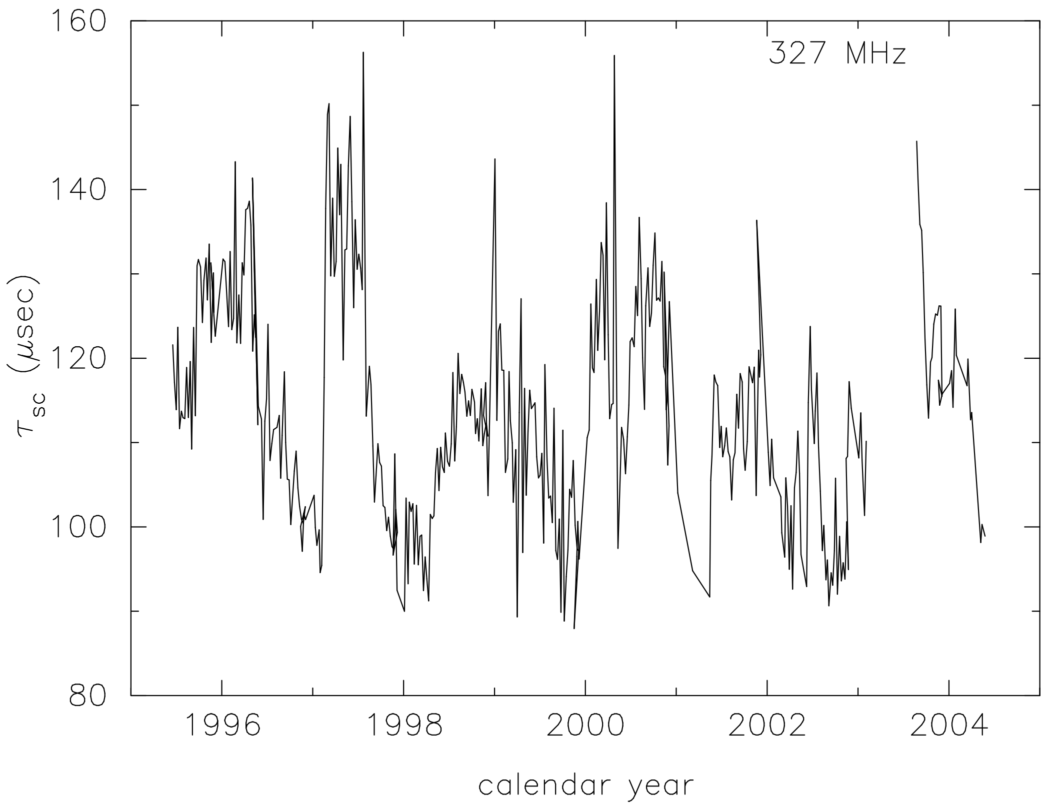

Figure 6.2: Measured temporal pulse broadening timescale (τ_{sc}) as a function of time at 327 MHz.

wide bandwidths. We have analyzed our best data sets, the densely sampled data at 327 and 610 MHz from Green Bank and at 1410 MHz from Nançay, for flux density variations as a function of time. The data are presented in Figure 6.3.

In analyzing the low frequency flux data from Green Bank, we have not adopted a rigorous flux calibration procedure. While there is a pulsed calibration noise source installed in this system, equipment changes and the nature of the automated observing have led to large gaps in the calibration record. Rather than dealing with a mix of calibrated and uncalibrated data, or lose a large fraction of the data, we decided not to apply any calibration. Instead, we normalize our data by assuming the system temperature is constant. In order to see what effect this has on our results, we did two tests.

First, we analyzed observations of PSR B1641−45, taken with the same system, over a similar time range. This pulsar is known to have a very long refractive timescale, $T_{\rm ref} > 1800$ days (Kaspi and Stinebring, 1992), so it can be used as a flux calibrator. We find our data to have a modulation index of $m = 0.10$. This immediately puts a upper limit of 10% on any systematic gain and/or system temperature variations. Since modulation adds in quadrature, and we observe modulation indices of $m \sim 0.4$ for PSR B1937+21, gain fluctuations represent at most a small fraction of the observed modulation.

We also considered the possibility that gain variations could influence our measurement of $T_{\rm ref}$. This might happen if they occur with a characteristic timescale longer than 1 day. In order to test this, we analyzed observations of the Crab pulsar, PSR B0531+21, again taken with the same system over the same time range. The refractive parameters of this pulsar were studied in detail by Rickett and Lyne (1990). It makes a good comparison since it has modulation index of $m = 0.4$ at 610 MHz, very similar to PSR B1937+21. Applying the structure function analysis (see §6.5) to this data gives $T_{\rm ref} = 11$ days at 610 MHz, and $T_{\rm ref} = 63$ days at 327 MHz, consistent with the previously published results and a scaling law of $T_{\rm ref} \propto \nu^{-2.2}$.

The procedure that we have adopted to calibrate our data set from the Nançay telescope is described in detail by Cognard et al. (1995).

6.4 Distribution of Scattering Material Along the Line of Sight

Several authors have shown how the scattering parameters of a pulsar can be used to assess the distribution of scattering material along the LOS (Gwinn et al., 1993; Deshpande and Ramachandran, 1998; Cordes and Rickett, 1998). This results from the varied dependences of the scattering parameters on the fractional distance of scattering material along the LOS. PSR B1937+21 is viewed through the local spiral arm as well as the Sagittarius arm which are both potential sites of strong scattering. The parameters employed in this analysis are: the temporal pulse broadening by scattering ($\tau_{\rm sc}$; or its conjugate parameter $\Delta\nu$, the diffractive scintillation bandwidth), the diffractive scintillation time scale ($T_{\rm diff}$), the angular broadening from scattering (θ_H), the proper motion of the pulsar (μ_α, μ_δ), and a distance estimate of the pulsar (D).

Let us first compare θ_H and $\tau_{\rm sc}$, which are the result of multiple scattering along the LOS, and express them as (Blandford and Narayan, 1985)

$$\tau_{\rm sc} = \frac{1}{2cD}\int_0^D x(D-x)\,\psi(x)\,{\rm d}x \tag{6.1}$$

$$\theta_H^2 = \frac{4\ln 2}{D^2}\int_0^D x^2\psi(x)\,{\rm d}x. \tag{6.2}$$

In these equations, x is the coordinate along the LOS, with the pulsar at $x = 0$ and the observer at $x = D$, and $\psi(x)$ is the mean scattering rate. If the scattering material is uniformly distributed along the LOS, then the relation between the two quantities can be expressed as $\theta_H^2 = 16\ln 2\,(c\tau_{\rm sc}/D)$. With a distance of 3.6 kpc to the pulsar (according to the distance model of Cordes and Lazio (2002)), and the average pulse broadening time scale of 120 μs (from the present work), we obtain an estimate of the angular broadening, θ_τ, of 12 mas. This is in modest agreement with the measured value of 14.6 ± 1.8 mas, given the uncertainty in the distance to the pulsar and the simple assumption that the scattering material is uniformly distributed along the LOS.

Next, we formulate two approaches to estimation of the velocity of the LOS with respect to the scattering medium, and use these approaches to assess the location and extent of the medium. The transverse velocity of the pulsar based on the measured proper motion (Table 6.2) and assumed distance of $D = 3.6$ kpc (Cordes and Lazio,

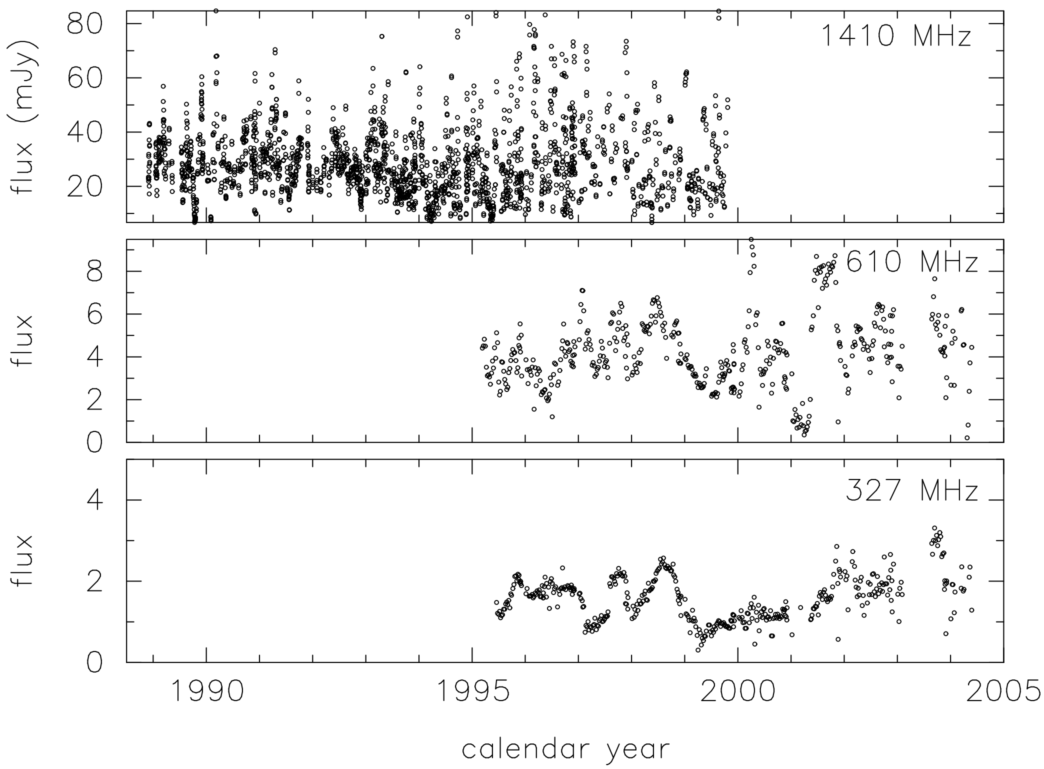

Figure 6.3: Measured flux density as a function of time. The top panel corresponds to the radio frequency of 1410 MHz with the data obtained from the Nançay telescope, the middle and the bottom panel to 610 and 327 MHz with the data obtained from the Green Bank 85 foot telescope.

2002) is 9 km s^{-1}. This value is the velocity of the pulsar with respect to the solar system barycenter. With the assumed "flat rotation curve" linear velocity of the Galaxy of 225 km s^{-1}, and the Sun's peculiar velocity of 15.6 km s^{-1} in the Galactic coordinate direction of $(l, b) = (48\overset{\circ}{.}8, 26\overset{\circ}{.}3)$ (Murray, 1983), the transverse velocity of the pulsar in its LSR (V_p) is 80 km s^{-1}.

The *scintillation velocity* ($V_{\rm iss}$), which is an estimate of the velocity of the *diffraction pattern* at the location of the Earth, is estimated from the decorrelation bandwidth ($\Delta\nu$) and the diffractive scintillation time scale ($T_{\rm diff}$). Gupta et al. (1994) conclude that

$$V_{\rm iss} \;=\; 3.85 \times 10^4 \sqrt{\frac{D\, z\, \Delta\nu}{(1-z)}\, \frac{1}{T_{\rm diff}\; \nu_{\rm GHz}}} \quad {\rm km\ s^{-1}} \qquad (6.3)$$

where $\nu_{\rm GHz}$ is the observing frequency in GHz, D is in kpc, $\Delta\nu$ is in MHz, and $T_{\rm diff}$ is in seconds. The variable z gives the fractional distance to the scattering screen, with $z = 0$ referring to the observer's position, and $z = 1$ the pulsar's position. The value of decorrelation bandwidth can be computed by the relation $\Delta\nu = 1/(2\pi\tau_{\rm sc})$. When the effective scattering screen is midway along the LOS ($z = 0.5$), $V_{\rm iss} = V_p$, and when the screen is at the location of the pulsar ($z = 1.0$), $V_{\rm iss} = \infty$. While doing this, an important assumption is that the pulsar proper motion is dominant over contributions from differential Galactic motion, solar peculiar velocity, and the Earth's annual orbital modulation. In the case of PSR B1937+21, this assumption is not justified. The effective scattering screen, which is located somewhere along the LOS, has a Galactic motion whose component along the LOS direction is different from that of the pulsar or the Sun. In order to correct for this effect, let us calculate the LOS velocity across the effective scattering screen at a fractional distance z from the observer:

$$V_{\perp} = (1-z)V_{\rm iss} = 3.85 \times 10^4 \frac{\sqrt{Dz(1-z)\Delta\nu}}{T_{\rm diff}\nu} {\rm km\ s^{-1}} \tag{6.4}$$

Then, let us assume that the scattering along the LOS can be adequately expressed by a single thin screen, at a distance zD from the observer. In this case, Equation 6.2 can be expressed as

$$\tau_{\rm sc} = \frac{\psi_{\circ}}{2c}\, D\, z\, (1-z) \tag{6.5}$$

$$\theta_H^2 = 4\, \ln 2\, (1-z)^2 \psi_{\circ} \tag{6.6}$$

Here, $\psi_{\circ}$ gives the mean scattering rate corresponding to the effective thin screen. The transverse velocity of the LOS across the scattering screen can be expressed independently as

$$\begin{aligned} \mathbf{V}'_{\perp} &= (1-z)\mathbf{V}_{\odot} + z\mathbf{V}_p - \mathbf{V}_G(zD\hat{n}) \\ &= \mathbf{V}_{\odot} + zD\mu - \mathbf{V}_G(zD\hat{n}), \end{aligned} \tag{6.7}$$

where $\mathbf{V}_{\odot}$ is the transverse component of the Sun's velocity, $\mathbf{V}_p$ is the pulsar's transverse velocity, and $\mathbf{V}_G$ is the transverse component of the Galactic rotation at a distance zD along the LOS. All the vectors in Equation 6.7 must be taken relative to the same inertial frame of reference, for example the LSR at Earth. We have for the moment neglected to include the effect of Earth's orbital velocity.

Equations 6.4 and 6.7 give two independent estimates of the line-of-sight velocity across the effective scattering screen and therefore allow us to solve for the value of z given D. With $D = 3.6$ kpc, we find $z = 0.7$. The LOS velocity is 51 km s^{-1}. The assumed value of $T_{\rm diff}$ is 78 s at 327 MHz (scaled from 444 s at 1400 MHz of Cordes et al. (1990)), and the value of $\Delta\nu$ is 0.0013 MHz calculated from $\tau_{\rm sc} = 120$ μs.

To summarize, the measured value of $\theta_H = 14.6 \pm 1.8$ mas and the estimated value of θ_τ are consistent with each other, suggesting a uniformly distributed scattering medium. On the other hand, comparison of the velocities $V_\perp$ and $V'_\perp$ suggest a thin-screen at $z \sim 2/3$. As Deshpande and Ramachandran (1998) demonstrate, this solution is equivalent to having a uniformly distributed scattering medium. Therefore, we conclude that the line of sight to PSR B1937+21 can be described adequately by a uniformly distributed scattering matter.

The Earth's orbital velocity around the Sun will modulate the observed scintillation speed, and therefore the diffractive scintillation time scale, with a one year periodicity. The amplitude of this modulation will depend on the effective z of the diffracting material, and so monitoring could provide an estimate of the effective screen location. If the effective screen is close to the Earth, then the modulation is strong, and if it is located close to the pulsar, then it is negligible. Figure 6.4 demonstrates this effect. The ordinate and abscissa give the LOS velocity across the effective scattering screen along the galactic longitude and latitude, respectively. For an assumed distance of 3.6 kpc, the straight line shows the expected centroid velocity of $\mathbf{V}'_\perp$. The left most end of the line (origin of the plot) corresponds to $z = 0$, and the right most end corresponds to $z = 1$. The annual modulation of $\mathbf{V}'_\perp$, shown as the two ellipses, correspond to $z = 0.5$ and $z = 2/3$. We have no way of identifying this annual modulation in our data, as we are insensitive to diffractive effects in our data set.

Another measurement that could help us is the direct measurement of distance to this object via annual parallax. Chatterjee et al. (2005, private communication), from their preliminary Very Long Baseline Array (VLBA) based parallax measurements, report that the distance to PSR B1937+21 is $2.3^{+0.8}_{-0.5}$ kpc, if they force the proper motion value to be the same as that of our timing based measurements (Table 6.2). In the coming year, accuracy of their measurements will improve with further sensitive observations.

6.5 Refractive Scintillation

6.5.1 Parameter Estimation

We determine refractive scintillation parameters from the data presented in Figure 6.3 following the structure function approach used in previous studies (Stinebring et al., 2000; Kaspi and Stinebring, 1992; Rickett and Lyne, 1990). We define the structure function D_F for flux time series $F(t)$ as

$$D_F(\delta t) = \frac{\langle [F(t) - F(t + \delta t)]^2 \rangle}{\langle F(t) \rangle^2}, \tag{6.8}$$

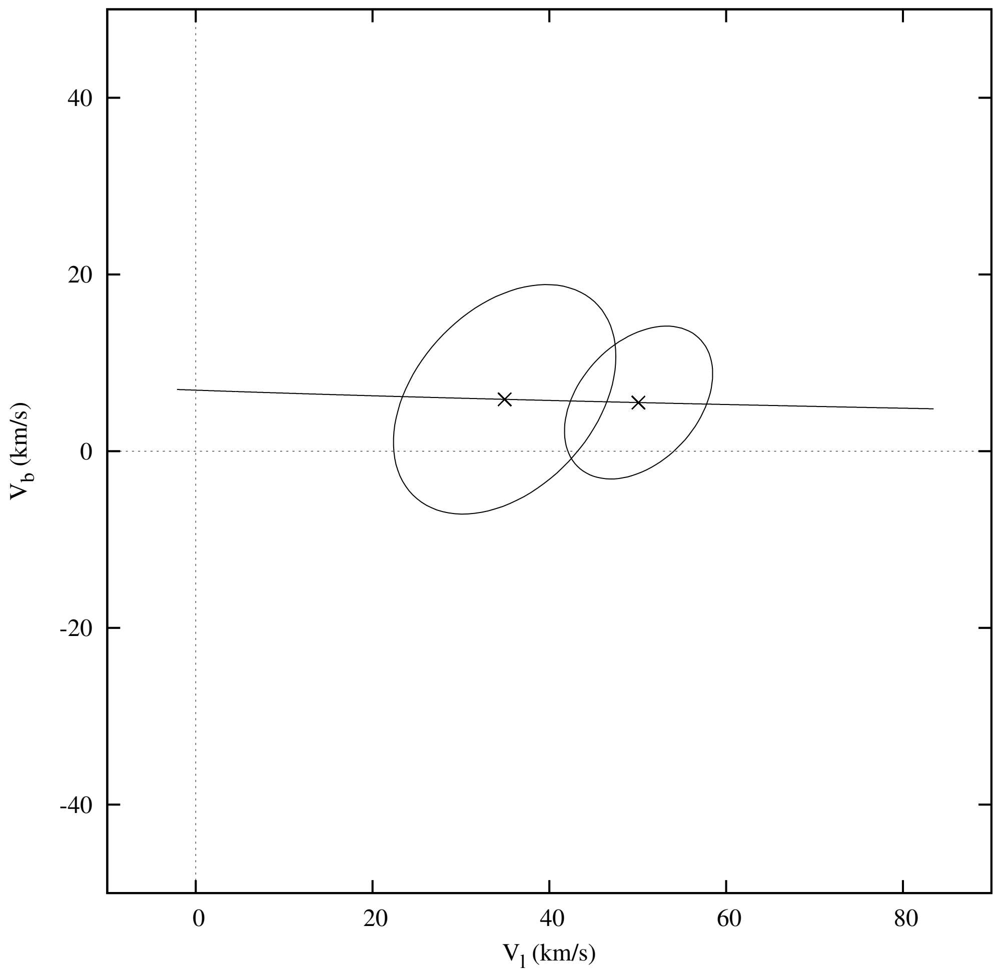

Figure 6.4: Estimated line of sight transverse velocity across the effective scattering screen at a distance of zD from the Sun. The velocity is resolved into the components along galactic longitude (V_l) and latitude (V_b). We have assumed a distance of 3.6 kpc to the pulsar from the Sun. Any point on the line indicates a combination of V_l and V_b corresponding to a value of z, with the left most end for $z = 0$ and the right most end for $z = 1$. The line itself does not include the Earth's orbital velocity contribution. The annual modulation due to Earth's motion in its orbit is shown as the ellipses. The two ellipses correspond to the scenarios of $z = 0.5$ and $z = 2/3$.

where δt is a time delay. Since our flux measurements occur at discrete and unevenly spaced time intervals, we compute the flux difference for all possible lags, then average results into logarithmically spaced bins.

The flux structure function typically has a form described by Kaspi and Stinebring (1992), a flat, noise dominated section at small lags, then a power-law increase which finally saturates at a value D_s at large lags. In practice, the saturation regime may have large ripples in it, an effect of the finite length of any data set. In addition, the measured flux structure function is offset from the "true" flux structure function due to a contribution from uncorrelated measurement errors. At 327 and 610 MHz, we estimate this noise term from the short-lag (noise regime) values. At 1410 MHz (from Nançay), we use the individual flux error bars to get the noise level. After subtracting the noise value, we fit the result to a function of the form

$$D_F(\delta t) = \begin{cases} D_s(\delta t/T_s)^\alpha, & 0 < \delta t < T_s \\ D_s, & \delta t > T_s \end{cases} \tag{6.9}$$

In this fit, the power law slope α, the saturation timescale T_s, and the saturation value D_s are all free parameters. The flux structure function data and fits are shown in Figure 6.5.

As shown by Rickett and Lyne (1990), the refractive parameters can be measured from the flux structure function using the following relationships: The modulation index m is given by $m = \sqrt{D_s/2}$, and the refractive scintillation timescale $T_{\rm ref}$ is given by $D_F(T_{\rm ref}) = D_s/2$. All the measured parameters, including those measured

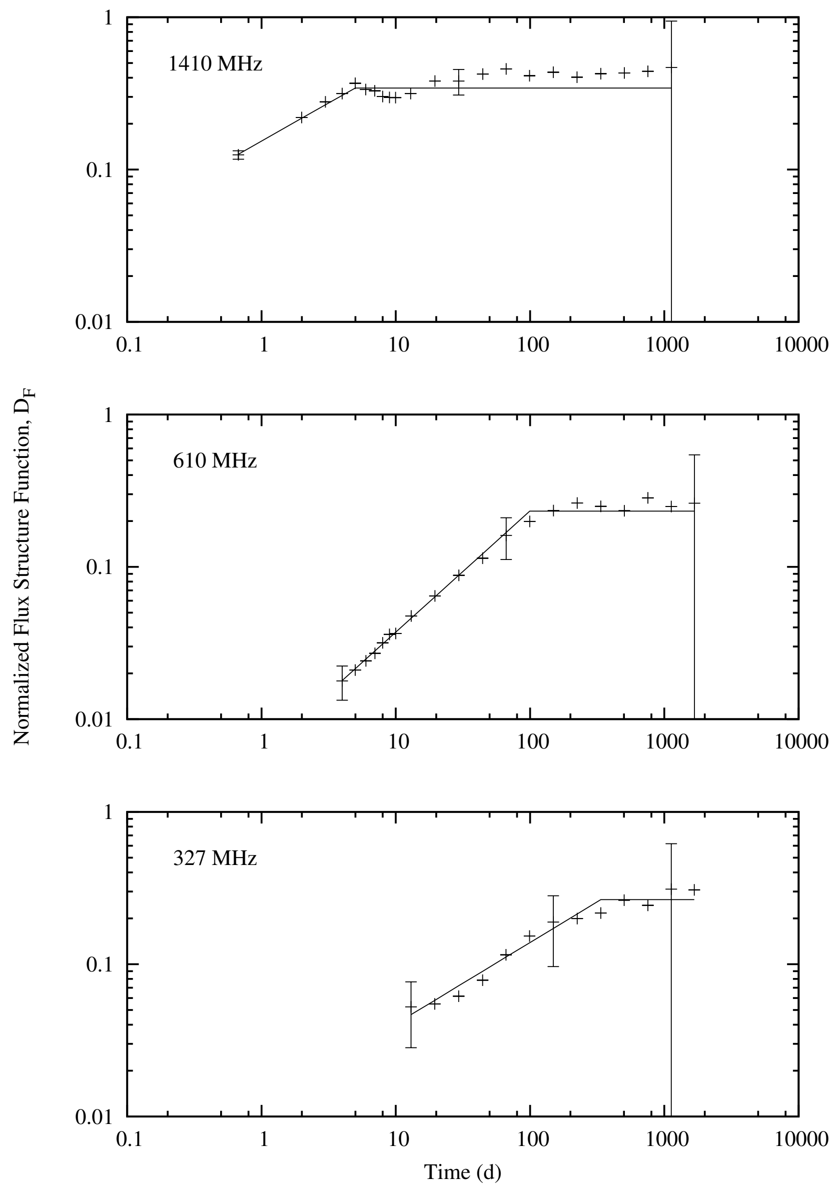

Figure 6.5: Structure function of normalized flux variations at 1410 (*Top*), 610 (*middle*), and 327 MHz (*bottom*). The 1410 MHz data was obtained from the Nançay telescope. The saturation value of the structure function at larger lag values indicates the observed modulation index. Error bars are shown on only a few points, to preserve clarity. See text for details.

by earlier investigators are summarized in Table 6.3.

Based on a propagation model through a simple power-law density fluctuation spectrum, we expect to see refractive variations in the flux measurements on a timescale $T_{\rm ref} \sim 0.5\theta_H D/V_\perp$, where $V_\perp$ is the line of sight velocity across the effective scattering screen. For the sake of argument, if we assume an effective scattering screen at $z = 0.5$, then $V_\perp \sim 40$ km s^{-1}. With $\theta_H = 14.6$ mas, the expected refractive scintillation time scale is $\sim$3 yr at 327 MHz. This is more than an order of magnitude in excess of the measured value. Furthermore, if the density fluctuations in the medium are distributed according to the Kolmogorov power law distribution, then the expected frequency scaling law is $T_{\rm ref} \propto \lambda^{2.2}$. While our measured values do follow the $\lambda^{2.2}$ relation between 610 and 1420 MHz, the measured scaling is linear between 327 and 610 MHz. Our observed modulation index (m) values are also considerably larger than predicted, and show a "flatter" wavelength dependence, as listed in Table 6.3. We address this issue in detail in the following section.

6.5.2 Nature of the Spectrum: Inner Scale Cutoff

The three disagreements with a simple model summarized in §6.5.1 force us to explore a few aspects of the electron density power spectrum that may possibly explain what we observe. The effects of *caustics* on the observed scintillations have been explored by several earlier investigators, most notably Goodman et al. (1987) and Blandford and Narayan (1985). In particular, if the power law scale distribution in the medium is truncated at an inner scale that is considerably larger than the diffractive scale, as they show, the observed flux variations are dominated by caustics. This is of great interest to us, as this seems to explain all the discrepancies that we note in our observed refractive parameters. For instance, as Goodman et al. (1987) show, if the inner scale cutoff is a considerable fraction of the Fresnel scale, then the observed fluctuation spectrum of flux is dominated by fluctuation frequencies that are lower than the diffractive frequencies, but significantly higher than that expected from refractive scintillation. This is what we observe. Moreover, as they note, the observed wavelength dependence of the refractive time scale, as well as that of the modulation index is expected to be "shallower" than the expected values of $\lambda^{2.2}$ and $\lambda^{-0.57}$, respectively.

A shallow frequency dependence of the modulation index has been reported by others (Coles et al., 1987; Kaspi and Stinebring, 1992; Gupta et al., 1994; Stinebring et al., 2000). While Kaspi and Stinebring (1992) find that the observed refractive quantities matched well with the predicted values for five objects, three other objects, especially PSR B0833–45, each have a significantly shorter measured $T_{\rm ref}$ and greater modulation index than expected. This is very similar to our situation here with PSR B1937+21.

Stinebring et al. (2000) concluded that the 21 objects that they analyzed fell into two groups. The first group followed the frequency dependence predicted by a Kolmogorov spectrum with the inner cutoff scale far less than the diffractive scales

("Kolmogorov-consistent group"). The second group, which they called the "super-Kolmogorov group", is consistent with a Kolmogorov spectrum with an inner scale cutoff at $\sim 10^8$ m. The observed modulation indices were consistently greater than that of the Kolmogorov predictions, as we have seen in our measurements of PSR B1937+21. This group includes pulsars such as PSRs B0833–45 (Vela), B0531+21 (Crab), B0835–41, B1911–04 and B1933+16. An important physical property that binds them all is that, with the exception of one, all objects have a strong *thin-screen* scatterer somewhere along the LOS. This is either a supernova remnant (or a plerion) as in the case of Vela and Crab pulsars, a HII region as in the case of B1942–03 and B1642–03, or a Wolf-Reyet star as in the case of B1933+16 (see Prentice and Ter Haar, 1969; Smith, 1968). Although our measurements show that pulsar PSR B1937+21 is consistent with the characteristics of the super-Kolmogorov group, as we describe in §6.4, we find no compelling evidence for the presence of any dominant scatterer somewhere along the LOS.

To summarize, while some investigators have reported agreement of the measured refractive properties with the theoretical expectations from a Kolmogorov spectrum with an infinitesimally small inner scale, there are a considerable number of cases in which the observed properties differ significantly from the theoretical predictions. These other cases can be explained by invoking spectra with large inner scale cutoffs, including the case in which the cutoff approaches the Fresnel radius and leads to a caustic-dominated regime. From Gupta et al. (1994) and Stinebring et al. (2000), the modulation index can be specified as a function of the inner cutoff scale as

$$m = 0.85 \left(\frac{\Delta\nu}{\nu}\right)^{0.108} \left(\frac{r_i}{10^8\mathrm{m}}\right)^{0.167} D_{\mathrm{kpc}}^{-0.0294}. \tag{6.10}$$

With the known value of $\Delta\nu$ at 327 MHz of 1.33 kHz, the distance to the pulsar of 3.6 kpc, and the observed modulation index of 0.39, the inner scale cutoff, r_i, comes to 1.3×10^9 m.

6.6 DM Variations

We turn now to the dispersion measure variations presented in Figure 6.6 that sample density variations on transverse scales much larger than those involved with diffractive and refractive effects. The most striking feature in Figure 6.6 is the large secular decline from 71.040 pc cm^{-3} in 1985 to 71.033 pc cm^{-3} in 1991 and then to 71.022 pc cm^{-3} by late 2004. These long-term secular variations are many times greater than the RMS fluctuations of $\sim 10^{-3}$ pc cm^{-3} on short time scales. An important question that arises is whether these variations are the result of a spectrum of electron-density turbulence, or whether there might be a contribution from the smooth gradient of a cloud, or clouds along the LOS. We look at this question from two angles. First we present a phase structure function analysis of the dispersion measure data and estimate a power-law index of the electron density spectrum. Then we estimate the probability that such a spectrum would produce a 22 yr realization that was so strongly dominated by the large, monotonic changes mentioned above.

We write the power spectrum of electron-density fluctuations as

$$P(q) = C_n^2 q^{-\beta}, \qquad q_\circ < q < q_i \tag{6.11}$$

where β is the power law index, $q_\circ$ and q_i are the spatial frequencies corresponding to the outer and the inner boundary scale, within which this power law description is valid, and C_n^2 is the amplitude, or strength, of the fluctuations. A quantity that is closely related to the density spectrum that can be quantified by observable variables is the phase structure function, $D_\phi(b)$, with $b = 2\pi/q$. This is defined as the mean square geometric phase between two straight line paths to the observer, with a separating distance of b between them in the plane normal to the observer's sight line. The phase structure function and the density power spectrum are related by a transform (Rickett, 1990; Armstrong et al., 1995),

$$D_\phi(b) = \int_0^\infty 8\pi^2\lambda^2 r_e^2 dz' \int_0^\infty q[1 - J_0(bqz'/z)]dq \times P(q=0) \tag{6.12}$$

Here, r_e is the classical electron radius (2.82×10^{-15} m), and $J_\circ$ is the Bessel function. Under the conditions that we have assumed, $D_\phi(b)$ is also a power law (Rickett, 1990; Armstrong et al., 1995), given by

$$D_\phi(b) = \left(\frac{b}{b_{\mathrm{coh}}}\right)^{\beta-2} \tag{6.13}$$

where b_{coh} is the spatial coherence scale (which is defined by the relation $D_\phi(b_0) = 1$). Dispersion measure can be written as

$$\mathrm{DM} = 2.410 \times 10^{-16} \left[\frac{(\nu_1^2 - \nu_2^2)}{\nu_1^2\nu_2^2}\right]\left(\frac{\Delta\phi}{f}\right) \mathrm{pc\ cm}^{-3}, \tag{6.14}$$

$\tau_{\rm sc}$ (μs)	θ_H (mas)	$T_{\rm diff}$ (s)	$T_{\rm ref}$ observed	$T_{\rm ref}$ expected	m observed	m expected	ν (MHz)
$120^\dagger$	14.6^b	–	73 days	3 y^f	0.33	0.14^c	327
38^e	–	100^e	–	–	–	–	430
–	–	–	43.9 days	6 mon^g	0.39	0.2	610
–	–	260^a	3 days^d	45 days^g	0.45	0.33^c	1400
0.17^e	–	444^e	–	–	–	–	1400

Table 6.3: Measured and expected scintillation parameters.

†Has a time dependent RMS variation of 20 μs
^{a}Cordes et al. (1986)
^{b}Gwinn et al. (1993)
^{c}Romani et al. (1986); Kaspi and Stinebring (1992)
^{d}Lestrade et al. (1998) give the value as 13 days
^{e}Cordes et al. (1990)
^{f}Calculated with $T_{\rm ref} \sim \theta_H D/2V_\perp$
^{g}Extrapolated with $T_{\rm ref} \propto \lambda^{2.2}$

where $\Delta\phi$ is the difference in the arrival phases $(\phi_2 - \phi_1)$ of the pulse at the two barycentric radio frequencies (Hz) ν_1 and ν_2, with f being the barycentric apparent rotation frequency (Hz) of the pulsar. With this linear relation between DM and geometric phase difference, the structure function can be written as (KTR94)

$$D_\phi(b_0) = \left(\frac{2\pi}{\nu}\frac{\rm Hz}{2.410\times10^{-16}\ \rm pc\ cm^{-3}}\right)^2 \times \left\langle[\mathrm{DM}(b+b_0) - \mathrm{DM}(b)]^2\right\rangle . \tag{6.15}$$

Here, the angular brackets indicate ensemble averaging. The transformation between the spatial coordinate b (and the spatial delay b_0) and the time coordinate t (or time delay τ) is simply given by $b = V_\perp t$, where $V_\perp$ is the transverse velocity of the LOS across the effective scattering screen.

With the understanding that any difference in DM that we compute for a time baseline from Figure 6.6 corresponds to a point in the phase structure function, we

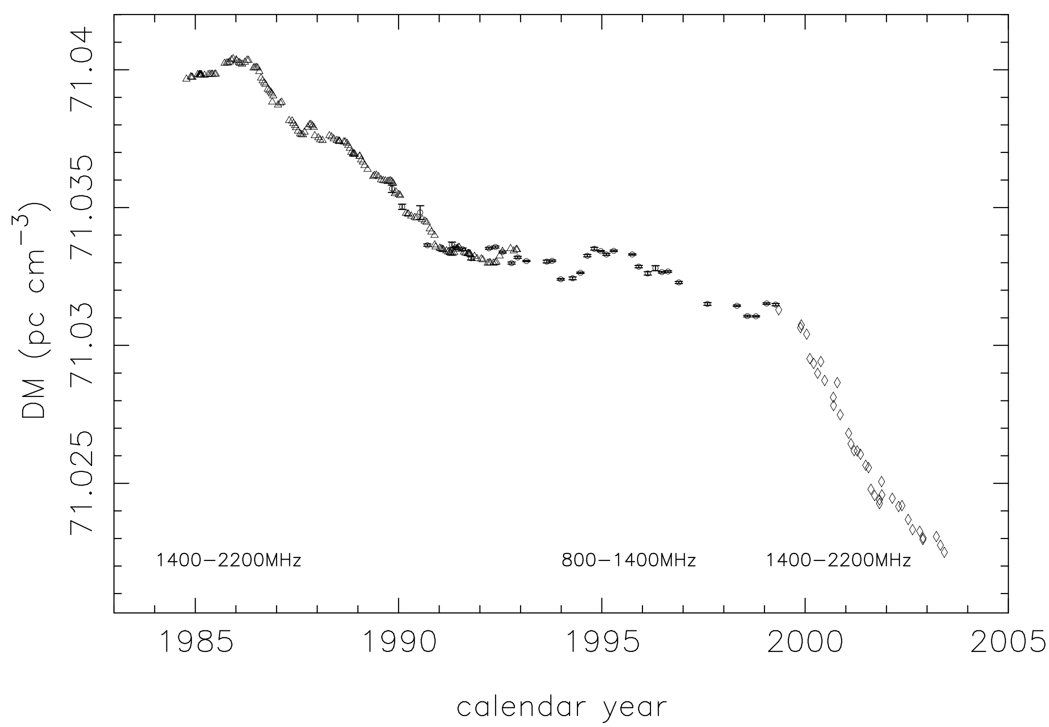

Figure 6.6: Dispersion Measure as a function of time. Open triangles give the measurements of KTR94 at 1400 and 2200 MHz, open circles are from our Green Bank 140-ft telescope measurements at 800 and 1400 MHz, and the open diamond symbols indicate our measurements from the Arecibo Observatory, at 1420 and 2200 MHz. All error bars indicate RMS errors.

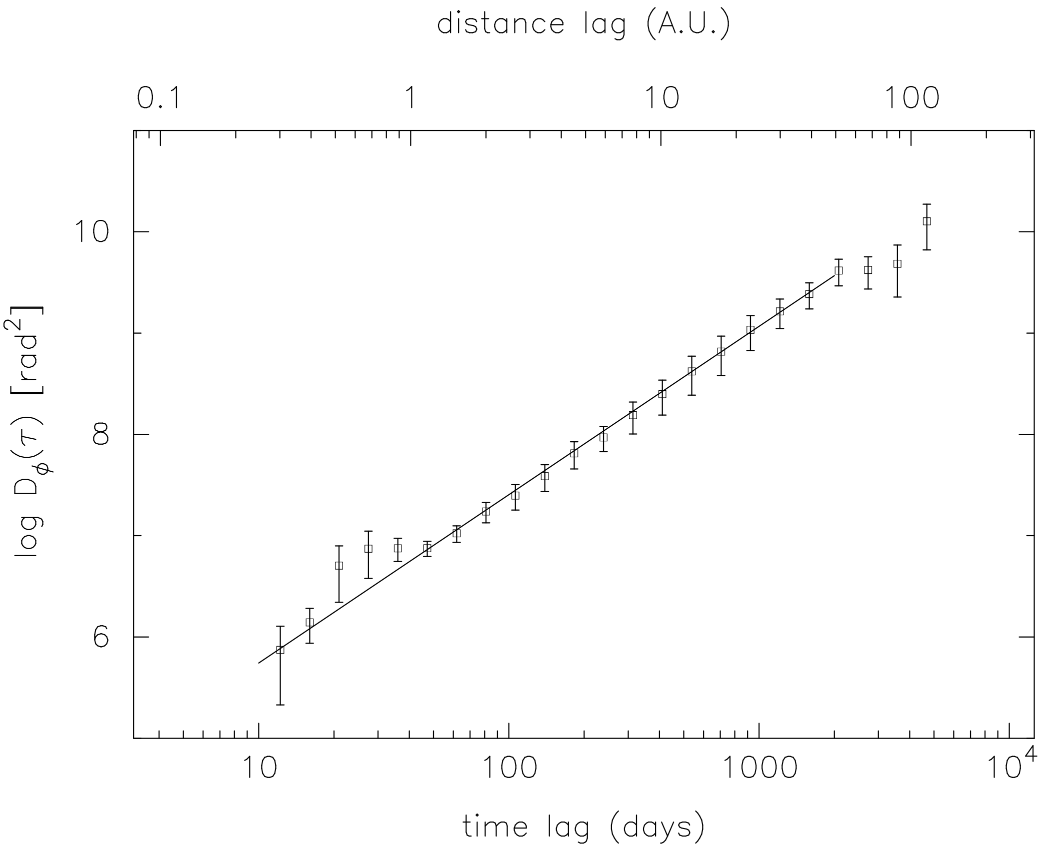

Figure 6.7: The phase structure function derived with the help of Equation 6.15 from the data displayed in Figure 6.6. Solid line represents the best fit for the data in the time range of 30 days to 2000 days. For translating the time delay range into a space delay, we have assumed a sightline transverse velocity of 40 km sec^{-1}, which is half of the pulsar's peculiar velocity in its LSR. We have assumed an effective screen at a fractional distance of 0.5. See text for details.

can derive the phase structure function on the basis of Equation 6.15. This is given in Figure 6.7.

There are several important points in Figure 6.7. The solid line gives the best fit line for the data in the time interval of 5-2000 days. The derived values of the intercept and the power law index (β) are,

$$\begin{aligned} \text{intercept} &= 4.46 \pm 0.09 \\ \beta &= 3.66 \pm 0.04 \end{aligned} \tag{6.16}$$

The value of β is remarkably close to the value expected from a Kolmogorov power law distribution ($\beta = 11/3$). We are using the terminology "intercept" only to indicate the value of $\log[D_\phi(\tau)]$ at $\tau = 1$ day. Here, a cautionary remark is warranted: Given fact that the low spatial frequencies dominate the DM variations, closely spaced data points are not statistically independent. We have estimated the error in each bin of the structure function as

$$\sigma_s = \frac{\sigma_D}{\sqrt{N_i}},$$

where σ_D is the root mean square deviation with respect to the mean phase structure function value in a bin, $D_\phi(\tau)$, and N_i is the number of independent samples in the bin. This is estimated as the smaller of (T/τ) and the actual number of samples that have gone into the estimation of $D_\phi(\tau)$. Here, T is the time span of the data. By assuming that the transverse speed of the sightline across the effective scattering screen is ~ 40 km s^{-1} (half of pulsar's velocity in its LSR), we can translate the delay range between which this slope is valid, from 0.2 to 50 A.U.

The time delay value corresponding to the phase structure function value of unity is, by definition, the coherent diffractive time scale ($T_{\rm diff}$) at the corresponding radio frequency, with the assumption that the scattering material is uniformly distributed along the LOS. From the fit parameters given in Equation 6.16, this delay is 180 s. This should be compared with the measured $T_{\rm diff}$ value of 444 ± 28 s tabulated in Table 6.3. If we interpret the inner scale cutoff value of $r_i \sim 1.3 \times 10^9$ m as the scale size below which the slope (β) of the density fluctuation spectrum changes to a value greater than that given in Equation 6.16, then the fact that the measured $T_{\rm diff}$ value of 444 s being significantly greater than 180 s is understandable. In the limiting case, in which the slope of the density irregularity power spectrum changes to the critical value of $\beta = 4$ below the inner scale cutoff value, the expected $T_{\rm diff}$ value is about 1100 s. This makes it very important to measure the exact frequency dependence of the diffractive parameters like temporal scatter broadening and diffractive scintillation time scale. To the best of our knowledge, Cordes et al. (1990) show the most complete multifrequency measurements of the diffractive scintillation parameters of this pulsar. However, their measurements are not accurate enough to distinguish between such small variations in slope.

While our analysis of DM variability suggests a Kolmogorov spectrum at AU scales, we are struck by the long term near-monotonic decrease of DM and wonder if we might be seeing the effects of smooth gradients in large scale galactic structures that are not part of a turbulent cascade. We performed a Monte Carlo simulation to investigate this. In each realization of the simulation, we generated, with a different random number seed, a screen of density fluctuations. We assumed that the random fluctuations at each spatial frequency are drawn from Gaussian distributions whose width as a function of frequency follows a power law of index $-11/3$. Assuming that the screen is located at the mid point along the sight line, we let the pulsar drift with its transverse velocity, and computed the implied DM as a function of time.

We developed a procedure similar to that of Deshpande (2000) to compare the observed DM(t) curve with the simulated ones. From the observed DM(t) curve, we computed the parameter $\Delta\mathrm{DM} = [\mathrm{DM}(t) - \mathrm{DM}(t - \tau_\circ)]$, where $\tau_\circ$ is the time delay. Our aim is to compare the distribution of this parameter in very short delays and very large delays. As we can see in Figure 6.7, the structure function describes a well defined slope between the delay range of ∼30 and ∼2000 days. We defined two delay bins, 30–60 and 1300–2000 days, within which we monitored the distribution function of the quantity ΔDM. From this, we could infer that the distribution at the bin of 1300–2000 days had a span of $\sim 20\sigma_s$, where σ_s is the RMS deviation of the distribution at the delay range of 30–60 days. That is, the ΔDM values that we see at the largest delays are as high as 20 times that of the typical deviations at short delays. We performed the same procedure on the simulated set of data to quantify the likelihood of such deviations. Out of 1024 simulated screens, we found that such large deviations were possible ∼7% of the times. This is perhaps not surprising, as with such a steep spectrum, it is obvious that most of the power is in large scales (smaller spatial frequencies), and hence they tend to dominate our ΔDM measurements. We conclude that while monotonic changes of this magnitude are somewhat rare, the observations are still consistent with a turbulent cascade spectrum of density fluctuations.

6.7 Frequency Dependence of DM

Dispersion depends on the column density of electrons. In a uniform medium radio wave propagation senses the average density in a tube whose width is set by the Fresnel radius $\sqrt{z(1-z)\lambda D}$. In a turbulent medium, frequency-dependent multipath propagation can expand this tube considerably. With refraction, the center of the tube wanders from the geometric LOS. Indeed there may be a number of wave propagation tubes, each with different relative gain. The consequence is that DM and related effects will show radio frequency dependence in the following ways:

1. The effective DM depends on frequency.

2. The DM variations observed at lower radio frequencies will be much “smoother” than those at higher frequencies, as the apparent angular size of the source acts as a *smoothing function* on the measured DM variations.

3. Since the apparent size of the source is larger at low frequencies, some features of the ISM that are visible at lower frequencies may be invisible at higher frequencies.

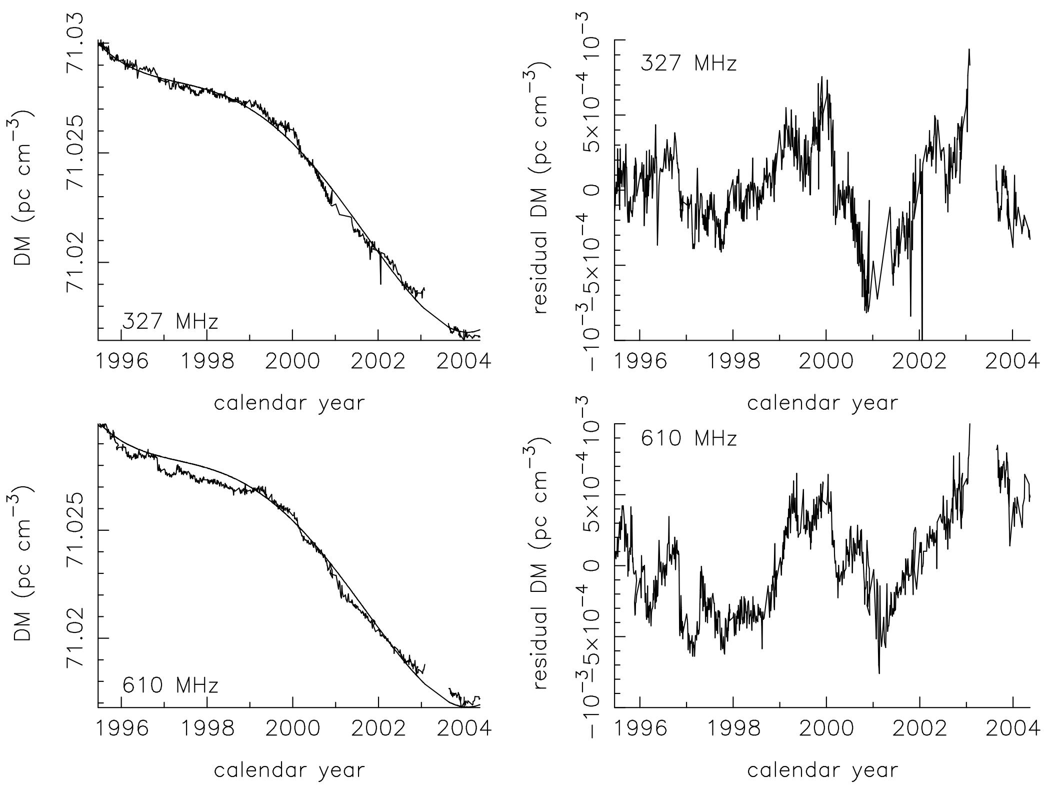

Figure 6.8: Dispersion Measure variations at 327 and 610 MHz. The two left hand side panels give DM as a function of time. The 327 MHz best fit line, produced by a fourth order polynomial fit, is given as a solid line in both 327 and 610 MHz plots. The residual DM, which is the difference between the actual DM and the best fit line, is given in the right hand side panels. See text for details.

We can explore these effects by assuming that the timing residuals at 327 and 610 MHz, which are relative to the timing model derived at higher frequencies that included removal of DM variations, are due to DM variations. The smoothing effect of scattering could be revealed by a spectral analysis. The slow variations were removed to prewhiten the spectrum that would otherwise be severely contaminated. The 327 MHz data were fit to a fourth order polynomial and the result was subtracted from both data sets. The two right side panels in Figure 6.8 give the residual DM values after subtracting the best fit curve from the actual DM curve. The resulting spectral comparison fails to have sufficient signal to clearly demonstrate increased smoothing at 327 MHz relative to 610 MHz. Higher signal-to-noise ratio is required.

An important source of systematic error that can affect our analysis here is the effect of scattering on the derived DM as a function of time at a given frequency. At 327 MHz, as we described in §6.3, we perform an elaborate procedure to fit for the scatter broadening of the pulse profile in order to compute the "true" TOA of the profile. However, we do not follow this procedure at 610 MHz (or any other higher frequency). The error due to this can be quantified easily from Figure 6.2. The temporal scatter broadening value varies by an RMS value of 19.6 μs. With the wavelength dependence of $\tau_{\rm sc} \propto \lambda^{4.4}$, the expected RMS variation at 610 MHz is 1.3 μs. The equivalent DM perturbation at 610 MHz with respect to infinite frequency is $\sim 10^{-4}$ pc cm^{-3}.

6.8 Achievable Timing Accuracy

In this section, we quantitatively estimate errors introduced by various scintillation related effects. For PSR B1937+21, although a typical observation with a highly sensitive telescope such as Arecibo allows us to achieve a TOA accuracy of a few tens of nanoseconds, the ultimate long-term timing residuals (difference between the data and the timing model) are much larger than this. In general, they are a combination of frequency-independent "intrinsic timing noise" from the pulsar itself, and the frequency-dependent effects, such as those we are addressing here. With 18 yr of data at 800, 1400 and 2200 MHz, Lommen (2001) quantifies the timing residual, after fitting for position, proper motion, rotation frequency and its time derivative (see also KTR94). A large fraction of the leftover residuals is presumably due to intrinsic timing noise. As we have mentioned before, we have absorbed a good part of this by fitting for the second time derivative of the rotation frequency, $\ddot{f}$ (see Table 6.2). In this section, our aim is to quantify possible timing errors from various "chromatic" effects related to interstellar scintillation.

6.8.1 Fluctuation of Apparent Angular Size

The temporal variability of pulse broadening, $\tau_{\rm sc}$, (as shown in Figure 6.2) means that the apparent angular broadening of the source, θ_H, is also changing as a function of time. Since $\tau_{\rm sc} \propto \theta_H^2$, with a RMS variation in $\tau_{\rm sc}$ of 19.6 μs at 327 MHz, the corresponding variation in θ_H comes to $\sim$8% of the mean value. This change occurs with typical time scales of ∼67 days, which is the time scale with which $\tau_{\rm sc}$ changes. Since we have only one epoch of θ_H measurement, we have no way of observationally verifying the mean value or the time scale of its variation.

6.8.2 Image Wandering

Because of non-diffractive scintillation that "steers" the direction of rays ("refractive focusing"), the apparent position of the pulsar is expected to change as a function of time. This is an important and significant effect, as it introduces a TOA residual as a function of time, depending on the instantaneous position of source on the sky. Several authors have investigated this effect in the past (Cordes et al., 1986; Romani et al., 1986; Rickett and Coles, 1988; Fey and Mutel, 1993; Lazio and Fey, 2001). For a Kolmogorov spectrum of irregularities ($\beta = 11/3$) with infinitesimally small inner scale cutoff, Cordes et al. (1986) predict the value of RMS image wandering as

$$\langle \delta\theta^2 \rangle^{1/2} = 0.18 \text{mas} \left(\frac{D_{\rm kpc}}{\lambda_{\rm cm}} \right)^{-1/6} \theta_H^{2/3} \tag{6.17}$$

For an assumed distance to PSR B1937+21 of 3.6 kpc, this comes to 2 mas at 327 MHz (wavelength $\lambda = 92$ cm). The value of 2 mas is still significantly less than the apparent angular size of the source, 14.6 mas, measured by Gwinn et al. (1993). However, for a spectrum with a steeper slope or with a significantly larger inner scale cutoff (as in our case), the value of $\langle \delta\theta^2 \rangle^{1/2}$ is expected to be much larger, perhaps comparable to the value of θ_H.

In order to estimate the timing errors introduced by this image wandering, we need an estimate of scattering measure (SM) and C_n^2 along the LOS to this pulsar. Following Cordes et al. (1991)

$$\begin{aligned}\mathrm{SM} &= \int_0^D C_n^2(x)\mathrm{d}x = \left(\frac{\theta_H}{128\ \mathrm{mas}}\right)^{5/3} \nu_{\mathrm{GHz}}^{11/3} \\ &= 292\left(\frac{\tau_{\mathrm{sc}}}{D_{\mathrm{kpc}}}\right)^{5/6} \nu_{\mathrm{GHz}}^{11/3}.\end{aligned} \tag{6.18}$$

Here, τ_{sc} is specified in seconds, and SM is specified in units of kpc $\mathrm{m}^{-20/3}$. Assuming a distance of 3.6 kpc, $\tau_{\mathrm{sc}} = 120\ \mu$s, and $\nu = 0.327$ GHz, the value of SM comes to $\sim 8.8 \times 10^{-4}$ kpc $\mathrm{m}^{-20/3}$. Assuming that the scattering material is uniformly distributed along the LOS, $C_n^2 \sim 2.4 \times 10^{-4}\ \mathrm{m}^{-20/3}$.

Then, for a Kolmogorov spectrum, the RMS timing residual due to the image wandering can be written as (Cordes et al., 1986)

$$\sigma_{\delta t_\theta} = (26.5\mathrm{ns})\nu^{-49/15} D^{2/3} \left(\frac{C_n^2}{10^{-4}\ \mathrm{m}^{-20/3}}\right)^{4/5}. \tag{6.19}$$

With the computed value of C_n^2 and a distance of 3.6 kpc, this amounts to 4.8 μs at 327 MHz. Given the frequency dependence, this effect can be minimized by timing the pulsar at higher frequencies. For instance, at frequencies of 1.0 and 2.2 GHz, this error translates to 125 and 2 ns, respectively. However, given the significantly large value of the inner scale cutoff, the RMS timing error that we have computed may well be a lower limit, and it is likely to be higher. In addition, the fact that the exact source position due to this effect is unknown at any given time makes it very difficult to compensate for this effect.

6.8.3 Barycentric Position Errors

As we saw above, due to refractive effects, the apparent position of the source wanders in the sky. This leads to yet another timing error, as follows: While translating the TOA at the observatory to the solar system barycenter, we use a source position which is shifted away from the actual apparent position at the time of observation. This introduces an error, which can be quantified as (Foster and Cordes, 1990)

$$\Delta t_{\rm bary} = \frac{1}{c}(\mathbf{r_e} \cdot \hat{n})(1 - z)\Delta\theta_r(\lambda), \tag{6.20}$$

where c is the velocity of light, the dot product term gives the projected extra path length travelled by the ray due to Earth's annual cycle around the Sun, and $\Delta\theta_r(\lambda)$ is the positional error due to image wandering. Since this last term is a function of frequency, the error accumulated is different at different frequencies.

An object in the ecliptic plane with an RMS image wandering angle of 2 mas would have $\Delta t_{\rm bary} \sim 2$ μs. For PSR B1937+21 at 327 MHz, this error amounts to ~ 0.8 μs. At frequencies of 1.0 and 2.2 GHz, this error translates to 85 and 17 ns, respectively.

6.8.4 Frequency Dependent DM

An important issue that arises due to the frequency dependent DM variation is the timing accuracy. Bright MSPs like PSR B1937+21 are known for the accuracy to which one can measure the pulse TOA. Given this, one wishes to eliminate any additional error that is incurred due to systematic effects. Between 327 and 610 MHz (the two curves in Figure 6.8), the typical relative fluctuation of DM that we see is about 5×10^{-4} pc cm^{-3}. As discussed before, this difference arises because the signal at each radio frequency takes a slightly different path through the interstellar medium, due to the various refractive effects. At 610 MHz, this relative DM fluctuation corresponds to some 6 μs. That is, at 610 MHz an unaccounted residual of up to 6 μs is incurred due only to effective DM errors. Even if the behavior of the pulse emission is extremely stable, at low frequencies interstellar scattering limits our timing capabilities.

Because dispersion delay scales as ν^{-2}, one should be able to reduce the above effect by going to higher frequencies. For instance, at 2.2 GHz, the DM-limited TOA error for PSR B1937+21 will be ~ 0.5 μs. This is not necessarily encouraging, as a timing residual error of 0.5 μs is large when compared to the accuracy that we can achieve in quantifying the TOAs (a few tens of ns) for this pulsar.

To summarize, although one takes into account time dependent DM changes while analyzing the data, in order to achieve high accuracy timing, it is necessary to consider a frequency-dependent DM as well. This, and the other frequency-dependent timing effects discussed here, should be weighed against the pulsar flux versus frequency, and equipment limitations, to determine a optimal set of observing frequencies.

6.9 Concluding Remarks

We have presented in this paper a summary of over twenty years of timing of PSR B1937+21. These observations have been done over frequencies ranging from 327 MHz to 2.2 GHz with four different telescopes.

Given the agreement between the measured apparent angular broadening and that estimated by the temporal broadening, and the measured proper motion velocity and that estimated by the knowledge of scintillation parameters, we conclude that the scattering material is uniformly distributed along the sightline.

There are three significant discrepancies between the expected values and the measured refractive parameters:

1. The measured flux variation time scale is about an order of magnitude shorter than what is expected from the knowledge of the observed apparent angular broadening.

2. The flux variation time scale is observed to be directly proportional to the wavelength, whereas it is expected to vary as proportional to $\lambda^{2.2}$ (for a Kolmogorov spectrum).

3. The flux modulation index is observed to have a wavelength dependence that is much "shallower" than the expected value.

These three discrepancies consistently imply that the optics are "caustic-dominated". This would mean that the density irregularity spectrum has a large inner scale cutoff,

1.3×10^9 m. Our extrapolation of the phase structure function from the regime sampled by DM variations to the diffractive regime seems to indicate that the expected $T_{\rm diff}$ value is considerably shorter than the measured value. This is in favor of the above conclusion. Accurate measurements of frequency dependence of diffractive parameters is much needed.

In general, millisecond pulsars are known for their timing stability. Potentially, we may achieve adequate accuracy in timing some of these pulsars to understand some of the most important questions related to the gravitational background radiation, or the internal structure of these neutron stars. However, our analysis here shows that interstellar scattering could be an important and significant source of timing error. As we have shown, although PSR B1937+21 is known to produce short term TOA errors as low as 10–20 ns with sensitive observations, the long term error is far larger than this. After fitting for $\ddot{f}$ (which absorbs most of the achromatic timing noise), the best accuracy that we can achieve for this pulsar is 0.9 μs at 1.4 GHz, and about 0.5 μs at 2.2 GHz (Lommen, 2001). It appears that almost all of this error can be accounted for by various effects that we have discussed in §6.8. In general, for millisecond pulsars with substantial DM, even if achromatic timing noise is small, the interstellar medium may be a major source of timing noise.

Bibliography

J. W. Armstrong, B. J. Rickett, and S. R. Spangler. Electron density power spectrum in the local interstellar medium. ApJ, 443:209–221, April 1995.

D. C. Backer and S. T. Sallmen. Polarimetry, Peculiar Mode and X-Ray Alignment of PSR B1821-24. AJ, 114:1539–+, October 1997.

D. C. Backer, S. R. Kulkarni, C. Heiles, M. M. Davis, and W. M. Goss. A millisecond pulsar. Nature, 300:615–618, December 1982.

D. C. Backer, S. Hama, S. van Hook, and R. S. Foster. Temporal variations of pulsar dispersion measures. ApJ, 404:636–642, February 1993.

D. C. Backer, M. R. Dexter, A. Zepka, D. Ng, D. J. Werthimer, P. S. Ray, and R. S. Foster. A Programmable 36-MHz Digital Filter Bank for Radio Science. PASP, 109: 61–68, January 1997.

D. C. Backer, T. Wong, and J. Valanju. A Plasma Prism Model for an Anomalous Dispersion Event in the Crab Pulsar. ApJ, 543:740–753, November 2000.

M. C. Begelman, R. D. Blandford, and M. J. Rees. Massive black hole binaries in active galactic nuclei. Nature, 287:307–309, September 1980.

R. Blandford and R. Narayan. Low-frequency variability of pulsars. MNRAS, 213: 591–611, April 1985.

R. Blandford and S. A. Teukolsky. Arrival-time analysis for a pulsar in a binary system. ApJ, 205:580–591, April 1976.

R. Blandford, R. W. Romani, and R. Narayan. Arrival-time analysis for a millisecond pulsar. *Journal of Astrophysics and Astronomy*, 5:369–388, December 1984.

M. M. Blaskiewicz. *Observational constraints on pulsars: Location of the emission region and pulse shape stability on decade time scales*. PhD thesis, Cornell University, 1991.

M. C. Britton. Radio Astronomical Polarimetry and the Lorentz Group. ApJ, 532: 1240–1244, April 2000.

W. L. Burke. Large-Scale Random Gravitational Waves. ApJ, 196:329–334, March 1975.

I. Cognard and D. C. Backer. A Microglitch in the Millisecond Pulsar PSR B1821-24 in M28. ApJ, 612:L125–L127, September 2004.

I. Cognard, G. Bourgois, J.-F. Lestrade, F. Biraud, D. Aubry, B. Darchy, and J.-P.

Drouhin. High-precision timing observations of the millisecond pulsar PSR 1937+21 at Nancay. A&A, 296:169–+, April 1995.

W. A. Coles, B. J. Rickett, J. L. Codona, and R. G. Frehlich. Refractive scintillation in the interstellar medium. ApJ, 315:666–674, April 1987.

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965.

J. M. Cordes and T. J. W. Lazio. NE2001.I. A New Model for the Galactic Distribution of Free Electrons and its Fluctuations. *ArXiv Astrophysics e-prints*, July 2002.

J. M. Cordes and B. J. Rickett. Diffractive Interstellar Scintillation Timescales and Velocities. ApJ, 507:846–860, November 1998.

J. M. Cordes, A. Pidwerbetsky, and R. V. E. Lovelace. Refractive and diffractive scattering in the interstellar medium. ApJ, 310:737–767, November 1986.

J. M. Cordes, A. Wolszczan, R. J. Dewey, M. Blaskiewicz, and D. R. Stinebring. Timing and scintillations of the millisecond pulsar 1937+214. ApJ, 349:245–261, January 1990.

J. M. Cordes, M. Ryan, J. M. Weisberg, D. A. Frail, and S. R. Spangler. The Galactic distribution of free electrons. Nature, 354:121–124, November 1991.

T. Damour and N. Deruelle. *Ann. Inst. H. Poincare*, 43:107, 1985.

T. Damour and A. Vilenkin. Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. Phys. Rev. D, 71(6): 063510–+, March 2005.

A. A. Deshpande. The small-scale structure in interstellar Hi: a resolvable puzzle. MNRAS, 317:199–204, September 2000.

A. A. Deshpande and R. Ramachandran. Improving pulsar distances by modelling interstellar scattering. MNRAS, 300:577–582, October 1998.

S. Detweiler. Pulsar timing measurements and the search for gravitational waves. ApJ, 234:1100–1104, December 1979.

M. Enoki, K. T. Inoue, M. Nagashima, and N. Sugiyama. Gravitational Waves from Supermassive Black Hole Coalescence in a Hierarchical Galaxy Formation Model. ApJ, 615:19–28, November 2004. doi: 10.1086/424475.

A. L. Fey and R. L. Mutel. 608 MHz VLBI observations of the compact double radio source 2050+364 - Constraints on interstellar scattering. ApJ, 404:197–205, February 1993.

R. S. Foster and J. M. Cordes. Interstellar propagation effects and the precision of pulsar timing. ApJ, 364:123–135, November 1990.

R. S. Foster, L. Fairhead, and D. C. Backer. A spectral study of four millisecond pulsars. ApJ, 378:687–695, September 1991.

M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. special issue on "Program Generation, Optimization, and Platform Adaptation".

T. Gold. Rotating Neutron Stars as the Origin of the Pulsating Radio Sources. Nature, 218:731–+, May 1968.

J. J. Goodman, R. W. Romani, R. D. Blandford, and R. Narayan. The effects of caustics on scintillating radio sources. MNRAS, 229:73–102, November 1987.

Y. Gupta, B. J. Rickett, and A. G. Lyne. Refractive Interstellar Scintillation in Pulsar Dynamic Spectra. MNRAS, 269:1035–+, August 1994.

C. R. Gwinn, N. Bartel, and J. M. Cordes. Angular broadening of pulsars and the distribution of interstellar plasma fluctuations. ApJ, 410:673–685, June 1993.

R. J. Hanisch, A. Farris, E. W. Greisen, W. D. Pence, B. M. Schlesinger, P. J. Teuben, R. W. Thompson, and A. Warnock, III. Definition of the Flexible Image Transport System (FITS). A&A, 376:359–380, September 2001.

T. H. Hankins and B. J. Rickett. Pulsar signal processing. In *Methods in Computational Physics. Volume 14 - Radio astronomy*, pages 55–129, 1975.

C. Heiles, P. Perillat, M. Nolan, D. Lorimer, R. Bhat, T. Ghosh, M. Lewis, K. O'Neil, C. Salter, and S. Stanimirovic. Mueller Matrix Parameters for Radio Telescopes and Their Observational Determination. PASP, 113:1274–1288, October 2001.

D. J. Helfand, J. H. Taylor, P. R. Backus, and J. M. Cordes. Pulsar timing. I - Observations from 1970 to 1978. ApJ, 237:206–215, April 1980.

R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. ApJ, 265:L39–L42, February 1983.

G. B. Hobbs, R. T. Edwards, and R. N. Manchester. TEMPO2, a new pulsar-timing package - I. An overview. MNRAS, 369:655–672, June 2006.

C. J. Hogan. Gravitational waves from light cosmic strings: Backgrounds and bursts with large loops. Phys. Rev. D, 74(4):043526–+, August 2006.

A. W. Hotan, M. Bailes, and S. M. Ord. High-precision baseband timing of 15 millisecond pulsars. MNRAS, 369:1502–1520, July 2006.

A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. New York: J. Wiley, 2001.

R. Isaacman and J. M. Rankin. The Crab Nebula pulsar - Variability of dispersion and scattering. ApJ, 214:214–232, May 1977.

J. D. Jackson. *Classical Electrodynamics*. New York: J. Wiley, 3rd edition, 1999.

A. H. Jaffe and D. C. Backer. Gravitational Waves Probe the Coalescence Rate of Massive Black Hole Binaries. ApJ, 583:616–631, February 2003.

F. A. Jenet and S. B. Anderson. The Effects of Digitization on Nonstationary

Stochastic Signals with Applications to Pulsar Signal Baseband Recording. PASP, 110:1467–1478, December 1998.

F. A. Jenet, A. Lommen, S. L. Larson, and L. Wen. Constraining the Properties of Supermassive Black Hole Systems Using Pulsar Timing: Application to 3C 66B. ApJ, 606:799–803, May 2004.

F. A. Jenet, G. B. Hobbs, K. J. Lee, and R. N. Manchester. Detecting the Stochastic Gravitational Wave Background Using Pulsar Timing. ApJ, 625:L123–L126, June 2005.

F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest, R. T. Edwards, A. W. Hotan, J. M. Sarkissian, and S. M. Ord. Upper Bounds on the Low-Frequency Stochastic Gravitational Wave Background from Pulsar Timing Observations: Current Limits and Future Prospects. ApJ, 653:1571–1576, December 2006.

V. M. Kaspi and D. R. Stinebring. Long-term pulsar flux monitoring and refractive interstellar scintillation. ApJ, 392:530–542, June 1992.

V. M. Kaspi, J. H. Taylor, and M. F. Ryba. High-precision timing of millisecond pulsars. 3: Long-term monitoring of PSRs B1855+09 and B1937+21. ApJ, 428: 713–728, June 1994.

M. Kramer, K. M. Xilouris, F. Camilo, D. J. Nice, D. C. Backer, C. Lange, D. R.

Lorimer, O. Doroshenko, and S. Sallmen. Profile Instabilities of the Millisecond Pulsar PSR J1022+1001. ApJ, 520:324–334, July 1999.

M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo. Tests of General Relativity from Timing the Double Pulsar. *Science*, 314:97–102, October 2006.

T. J. W. Lazio and A. L. Fey. Multifrequency Very Long Baseline Array Observations of the Compact Double B2 2050+36: Constraints on Interstellar Scattering Revisited. ApJ, 560:698–706, October 2001.

J.-F. Lestrade, B. J. Rickett, and I. Cognard. Interstellar modulation of the flux density and arrival time of pulses from pulsar B1937+214. A&A, 334:1068–1084, June 1998.

A. N. Lommen. *Precision Multi-Telescope Timing of Millisecond Pulsars: New Limits on the Gravitational Wave Background and other results from the Pulsar Timing Array*. PhD thesis, University of California, Berkeley, 2001.

A. N. Lommen, P. B. Demorest, D. C. Backer, I. H. Stairs, and D. J. Nice. *In preparation*, 2007.

D. R. Lorimer and M. Kramer. *Handbook of Pulsar Astronomy*. Handbook of pulsar astronomy, by D.R. Lorimer and M. Kramer. Cambridge observing handbooks for research astronomers, Vol. 4. Cambridge, UK: Cambridge University Press, 2004, December 2004.

A. G. Lyne and F. Graham-Smith. *Pulsar astronomy.* Pulsar astronomy, 3rd ed., by A.G. Lyne and F. Graham-Smith. Cambridge astrophysics series. Cambridge, UK: Cambridge University Press, 2006 ISBN 0521839548., 2006.

C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation.* San Francisco: W.H. Freeman and Co., 1973.

C. A. Murray. *Vectorial Astrometry.* Bristol: Adam Hilger, 1983.

P. C. Peters and J. Mathews. Gravitational radiation from point masses in a keplerian orbit. *Phys. Rev.*, 131(1):435–440, Jul 1963.

J. A. Phillips and A. Wolszczan. Time variability of pulsar dispersion measures. ApJ, 382:L27–L30, November 1991.

A. J. R. Prentice and D. Ter Haar. On H II regions and pulsar distances. MNRAS, 146:423–+, 1969.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C. The art of scientific computing.* Cambridge: University Press, 2nd edition, 1992.

M. Rajagopal and R. W. Romani. Ultra–Low-Frequency Gravitational Radiation from Massive Black Hole Binaries. ApJ, 446:543–+, June 1995.

R. Ramachandran and M. Kramer. Unusual profile variations in pulsar PSR J1022+1001 - Evidence for magnetospheric "return currents"? A&A, 407:1085–1095, September 2003.

R. Ramchandran, P. Demorest, D. C. Backer, I. Cognard, and A. Lommen. Interstellar Plasma Weather Effects in Long-term Multi-frequency Timing of Pulsar B1937+21. ApJ, 641(2), April 2006.

J. M. Rankin, J. M. Comella, H. D. Craft, D. W. Richards, D. B. Campbell, and C. C. Counselman. Radio Pulse Shapes, Flux Densities, and Dispersion of Pulsar NP 0532. ApJ, 162:707–+, December 1970.

L. A. Rawley. Timing millisecond pulsars. *Ph.D. Thesis*, 1986.

L. A. Rawley, J. H. Taylor, and M. M. Davis. Fundamental astrometry and millisecond pulsars. ApJ, 326:947–953, March 1988.

B. J. Rickett. Radio propagation through the turbulent interstellar plasma. ARA&A, 28:561–605, 1990.

B. J. Rickett and W. A. Coles. The influence of propagation through the irregular interstellar plasma on VLBI observations. In M. J. Reid and J. M. Moran, editors, *IAU Symp. 129: The Impact of VLBI on Astrophysics and Geophysics*, pages 287–294, 1988.

B. J. Rickett and A. G. Lyne. Intensity variations of the Crab pulsar and refractive interstellar scintillation. MNRAS, 244:68–75, May 1990.

R. W. Romani, R. Narayan, and R. Blandford. Refractive effects in pulsar scintillation. MNRAS, 220:19–49, May 1986.

G. B. Rybicki and W. H. Press. Interpolation, realization, and reconstruction of noisy, irregularly sampled data. ApJ, 398:169–176, October 1992.

B. F. Schutz. *A first course in general relativity*. Cambridge: University Press, 1985.

X. Siemens, V. Mandic, and J. Creighton. Gravitational-Wave Stochastic Background from Cosmic Strings. *Physical Review Letters*, 98(11):111101–+, March 2007.

L. F. Smith. The distribution of Wolf-Rayet stars in the Galaxy. MNRAS, 141: 317–+, 1968.

E. M. Splaver, D. J. Nice, I. H. Stairs, A. N. Lommen, and D. C. Backer. Masses, Parallax, and Relativistic Timing of the PSR J1713+0747 Binary System. ApJ, 620: 405–415, February 2005.

E. M. Standish, Jr. The observational basis for JPL's DE 200, the planetary ephemerides of the Astronomical Almanac. A&A, 233:252–271, July 1990.

D. R. Stinebring, V. M. Kaspi, D. J. Nice, M. F. Ryba, J. H. Taylor, S. E. Thorsett, and T. H. Hankins. A flexible data acquisition system for timing pulsars. *Review of Scientific Instruments*, 63:3551–3555, July 1992.

D. R. Stinebring, T. V. Smirnova, T. H. Hankins, J. S. Hovis, V. M. Kaspi, J. C. Kempner, E. Myers, and D. J. Nice. Five Years of Pulsar Flux Density Monitoring: Refractive Scintillation and the Interstellar Medium. ApJ, 539:300–316, August 2000.

J. H. Taylor. Pulsar timing and relativistic gravity. *Philosophical Transactions of the Royal Society of London*, 341:117–134, 1992.

S. E. Thorsett and R. J. Dewey. Pulsar timing limits on very low frequency stochastic gravitational radiation. Phys. Rev. D, 53:3468–3471, March 1996.

J. H. Van Vleck and D. Middleton. The spectrum of clipped noise. *Proceedings of the IEEE*, 54:2–19, January 1966.

M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33:387–392, April 1985.

J. S. B. Wyithe and A. Loeb. Low-Frequency Gravitational Waves from Massive Black Hole Binaries: Predictions for LISA and Pulsar Timing Arrays. ApJ, 590: 691–706, June 2003.

Q. Yu. Evolution of massive binary black holes. MNRAS, 331:935–958, April 2002.

Appendix A

Effect of the Timing Model Fit

In this section, we investigate the effect of a linear χ^2 fit on the spectral content of a signal. This is useful in the analysis of pulsar timing data for gravitational wave signals. We follow this with a discussion of how to optimally detect such a signal, and present a principal components-based approach for doing do.

A.1 Properties of Residuals

Linear χ^2 fitting is a analysis procedure which aims to reproduce a set of data as a linear combination of several basis functions. This is a very standard technique which is described in many texts (for example Press et al., 1992). Here we present a review, and establish terminology for the following sections.

The goal is to minimize a goodness of fit statistic, χ^2, versus a set of free parameters a_i:

$$\chi^2 = \sum_{i=0}^{N-1} \frac{\left(y_i - \sum_{j=0}^{M-1} a_j f_j(\mathbf{x}_i)\right)^2}{\sigma_i^2} \tag{A.1}$$

Here y_i are the measured data points, $\mathbf{x}_i$ are the independent variables associated with each data point, $f_j(\mathbf{x})$ are the fit basis functions, and σ_i is the measurement uncertainty of each data point. There are N data points and M free parameters. Minimizing this expression versus a_i results in the following matrix equation:

$$\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{a} = \mathbf{A}^T\mathbf{W}^2\mathbf{y} \tag{A.2}$$

The N-by-M matrix $\mathbf{A}$ is known as the design matrix, whose columns are the basis functions f_j evaluated at the points $\mathbf{x}_i$. $\mathbf{W}$ is a N-by-N diagonal weighting matrix with $W_{ii} = \sigma_i^{-1}$. $\mathbf{y}$ is a N element column vector of data, and $\mathbf{a}$ is a M element column vector of fitted parameters. The best-fit values for a_i are determined by solving Equation A.2 for $\mathbf{a}$.

Once a solution to Equation A.2 has been found, we can form the fitted data series, $\mathbf{y}_f = \mathbf{A}\mathbf{a}$, and the post-fit residuals, $\mathbf{r} = \mathbf{y} - \mathbf{y}_f$. Substituting in the solution for $\mathbf{a}$ gives:

$$\mathbf{r} = \mathbf{y} - \mathbf{A}\mathbf{a} = \mathbf{y} - \mathbf{A}(\mathbf{A}^T\mathbf{W}^2\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}^2\mathbf{y} = \mathbf{R}\mathbf{y} \tag{A.3}$$

The key insight here is that the residuals are a linear function of the data, obtained by applying the linear operator (N-by-N matrix) $\mathbf{R}$. This greatly simplifies the analysis since it means we can study the properties of $\mathbf{R}$ independent of the value of the data - $\mathbf{R}$ depends only on the fit basis functions (f_j), the sampling pattern ($\mathbf{x}_i$)

and the statistical weights (σ_i). We will state without proof a short list of generic properties of $\mathbf{R}$:

1. If $\mathbf{W} \propto \mathbf{I}$ (uniform weights), then $\mathbf{R}$ is symmetric. Otherwise, $\mathbf{R}^T = \mathbf{W}^{-1}\mathbf{R}\mathbf{W}$.

2. It is a projection: $\mathbf{R}^2 = \mathbf{R}$.

3. As such, it is also singular, with rank $N - M$. The null space of $\mathbf{R}$ is the subspace of signals spanned by the fit basis functions f_j.

The numerical computation of $\mathbf{R}$ is greatly simplified through the use of the singular value decomposition (SVD; see Press et al. (1992)). This has two benefits: First, determining the SVD of $\mathbf{WA}$ offers protection against the matrix inverse in Equation A.3 failing due to degenerate fit basis functions. This is a standard step in many χ^2 fit implementations. Second, once the SVD has been performed, $\mathbf{R}$ assumes an especially simple form. The SVD of a N-by-M matrix $\mathbf{B}$ is given by $\mathbf{B} = \mathbf{USV}^T$, where $\mathbf{U}$ is a N-by-M column-orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}$), $\mathbf{V}$ is a M-by-M orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}$), and $\mathbf{S}$ is a M-by-M diagonal matrix of singular values. When $\mathbf{WA}$ is expressed in this way, $\mathbf{R}$ takes the form

$$\mathbf{R} = \mathbf{I} - \mathbf{W}^{-1}\mathbf{U}\mathbf{U}^T\mathbf{W}. \tag{A.4}$$

A.2 Frequency Domain Analysis

So far we have said nothing about time or frequency. We will now specialize to the case where one of the independent variables is time, and the behavior of the signals $\mathbf{y}$

and $\mathbf{r}$ versus time is our primary interest. We would like to explore what the fitting process does to the spectral content of the signal. To start, we define the double Fourier transform of $\mathbf{R}$:

$$T(\omega_1, \omega_2) = \sum_{i,j} e^{-i\omega_2 t_i} R_{ij} e^{i\omega_1 t_j} \tag{A.5}$$

This function describes the effect of the fit on a pure tone: Given a input at frequency ω_1, the residuals will contain a signal at ω_2 with a relative (complex) amplitude of $T(\omega_1, \omega_2)$. It should be noted that aliasing effects due to discrete sampling and finite data length are included in T. We can use T to connect the input (data) and output (residual) signals in the frequency domain. Given a input with frequency content $y(\omega)$ we can find the output $r(\omega)$ as:

$$r(\omega) = \int y(\omega')T(\omega, \omega')d\omega' \tag{A.6}$$

We are interested in considering the input $\mathbf{y}$ as a stationary stochastic process with a given power spectrum $S_y(\omega)$. For stationary processes, the power spectrum completely describes the second order statistics:

$$E\left\{y^*(\omega_1)y(\omega_2)\right\} = S_y(\omega_1)\delta(\omega_1 - \omega_2) \tag{A.7}$$

In the time domain, this becomes $E\left\{y_i y_j\right\} = C_y(t_i - t_j)$, where the autocorrelation function $C_y(\tau)$ is related to $S_y(\omega)$ through a inverse Fourier transform. Combining Equations A.6 and A.7 gives the following expression for the statistics of $\mathbf{r}$:

$$E\left\{r^*(\omega_1)r(\omega_2)\right\} = \int S_y(\omega')T^*(\omega', \omega_2)T(\omega', \omega_1)d\omega' \tag{A.8}$$

In general, Equation A.8 will not have the simple form of Equation A.7. This means the effect of the fit process cannot be completely described by a simple one-dimensional transfer (or transmission) function. However, the 1-D function $P_r(\omega) = E\{r^*(\omega)r(\omega)\}$ does have a valid interpretation as the expected periodogram of the residuals. For the simple case of white noise input ($S_y(\omega) = 1$), we get the result $P_r(\omega) = T(\omega, \omega)$. This function is useful in illustrating how the fit process will selectively absorb certain frequencies, as shown in Figure A.1. This figure can be compared to the same function computed analytically by Blandford et al. (1984).

A.3 Optimal Basis

As discussed in the previous section and Chapter 5, we are interested in measuring (or at least placing limits on) a stochastic signal which has gone through a χ^2 fit procedure. The signal is expected to have a red spectrum with a power-law dependence on frequency, and will be combined with (possibly much stronger) white measurement noise. This raises the question of how to optimally detect such a signal. The detection procedure needs to be weighted towards low frequencies where the red noise spectrum is strongest, and also must avoid being influenced by the fit procedure. Two methods suggested in the past are the periodogram (Lommen, 2001; Jenet et al., 2004) and orthogonal polynomials (Jenet et al., 2006; Kaspi et al., 1994). Quantitative interpretation of a periodogram is complicated by the fact that sine waves are heavily affected by the fit: Different frequencies recieve different weights,

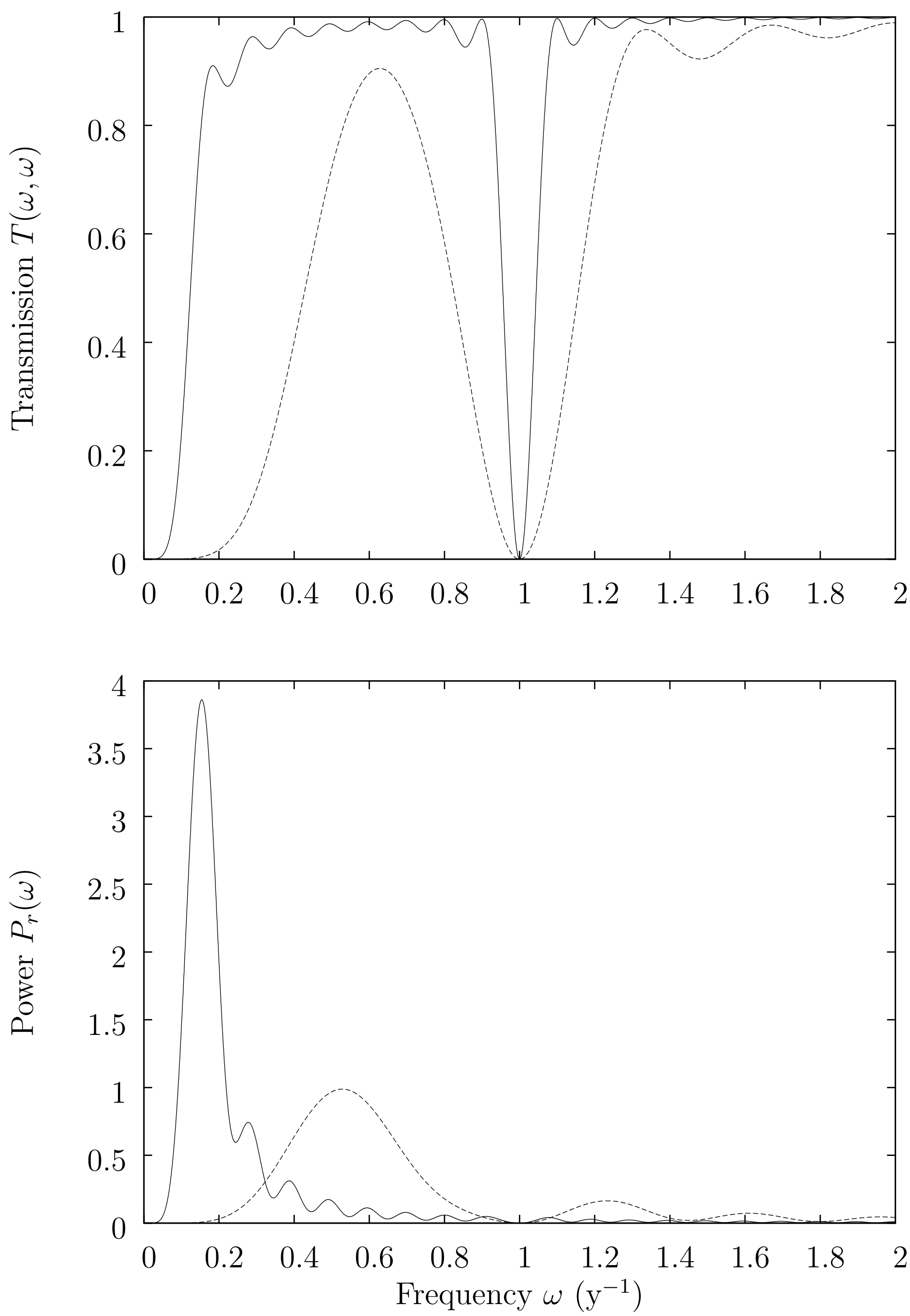

Figure A.1: Example timing model transmission function for 3 year (dashed) and 10 year (solid) data span (top). Expected periodogram for a $\alpha = 13/3$ signal (bottom).

and due to the action of $T(\omega_1, \omega_2)$ discussed above, they are no longer statistically independent. A orthogonal polynomial decomposition of the residuals is much better: Low-order polynomials are weighted towards low frequencies, and since the timing model includes a quadratic spin function, keeping only cubic and higher terms makes the decomposition approximately orthogonal to the fit.

It is possible to take the orthogonal polynomial approach to the next step: Given our assumptions about the input (the theoretically expected power spectrum), can we find a orthonormal basis which optimally represents the signal? It turns out that this is exactly the question answered by principal components analysis (PCA; see Chapter 2). Diagonalizing the expected covariance matrix of the residuals gives a orthonormal set of basis functions (eigenvectors) which are guaranteed to be orthogonal to the fit functions. The associated eigenvalues tell how much power is absorbed by each basis function. This makes it possible to determine how many basis functions to use, and how to weight each one, without having to rely on Monte Carlo approaches. It is also a convenient tool for investigating how different power spectra and fit parameters affect the final answer.

The starting point for this analysis is the expected covariance matrix of the input signal, $\mathbf{\Sigma_y}$. The elements of $\mathbf{\Sigma_y}$ are given by $\Sigma_{yij} = C_y(t_i - t_j)$. For a power-law spectrum $S_y(\omega) \propto \omega^{-\alpha}$, $C_y(\tau)$ can be computed as:

$$C(\tau) = (\alpha - 1)(\omega_0\tau)^{\alpha-1} \int_{\omega_0\tau}^{\infty} u^{-\alpha} \cos u du \tag{A.9}$$

Here, ω_0 is the spectrum's low-frequency cutoff, and the autocorrelation function has been normalized so that $C(0) = 1$. This integral can be computed numerically to populate the covariance matrix. The covariance matrix of the residuals is found by multiplying by $\mathbf{R}$: $\mathbf{\Sigma_r} = \mathbf{R}\mathbf{\Sigma_y}\mathbf{R}^T$. We then determine the optimal basis by diagonalizing $\mathbf{\Sigma_r}$.

This process is illustrated in Figures A.2 through A.4. These plots were generated using a "toy" timing model which includes only a quadratic spin function and a 1-year sine wave. Data points were taken to be equally spaced at 10 per year, and equally weighted. The spectral index used was $\alpha = 13/3$, corresponding to a characteristic strain $h_c \propto \omega^{-2/3}$ (Jaffe and Backer, 2003). The method is equally applicable to actual timing models once the appropriate R-matrices have been computed. This is done in the analysis presented in Chapter 5.

A.3.1 Note on Units

In Equation A.9 we presented a normalized autocorrelation function. It is necessary to consider how this should relate to physical units. The value $C(0)$ gives the total power in the input signal. In terms of the power spectrum, this is:

$$C(0) = \int_{\omega_0}^{\infty} S(\omega) d\omega \tag{A.10}$$

Normalizing $C(0)$ to a constant such as 1 means that the physical value of the total power will change depending on ω_0, the spectrum's low-frequency cutoff. Since all frequencies below $\sim 1/T_e$ (where T_e is the experiment duration) will be filtered out by the fit, it is more useful to consider a normalization where the spectrum amplitdue,

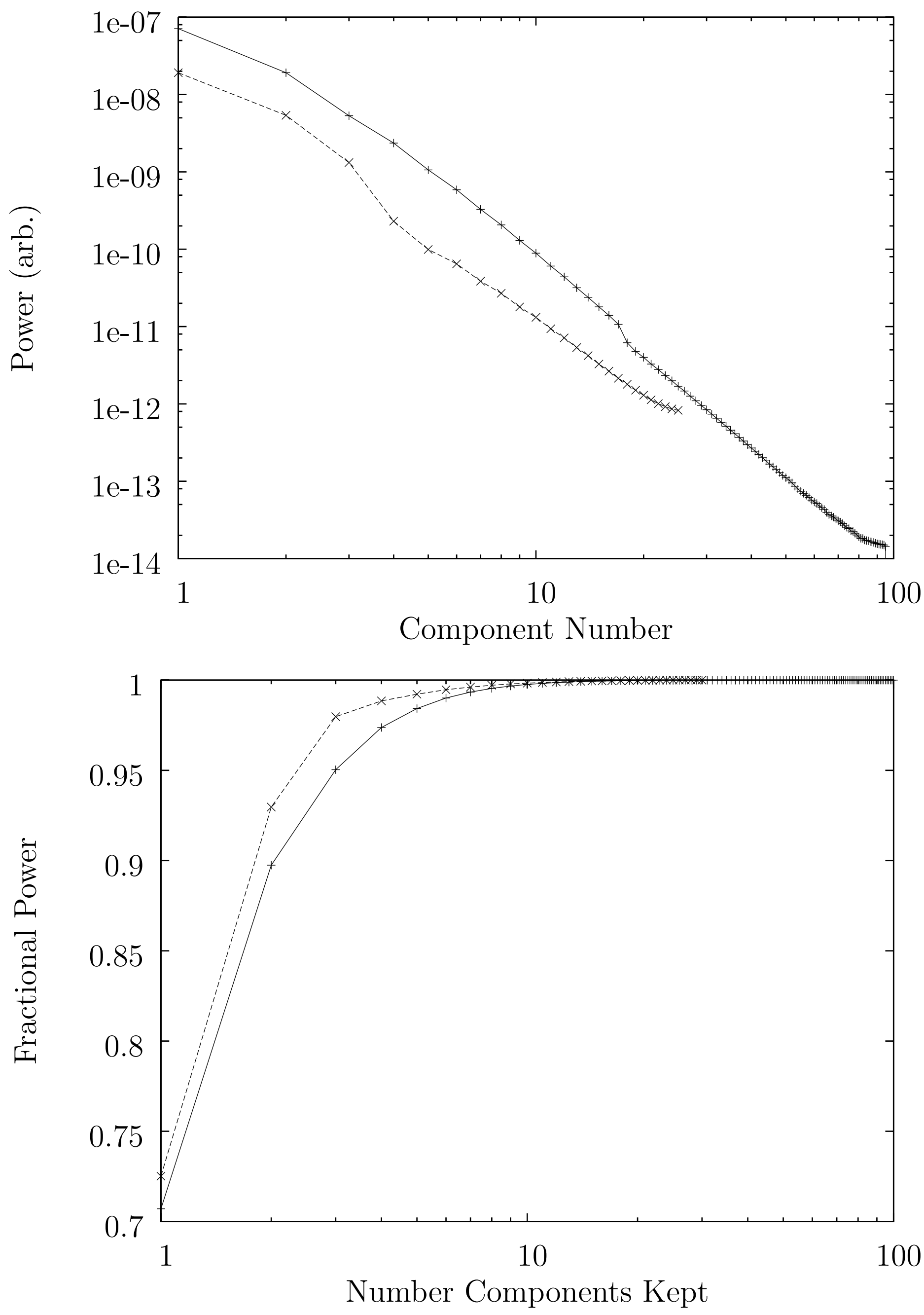

Figure A.2: Principal component eigenvalues for 3 year (dashed) and 10 year (solid) data span (top). Fraction of transmitted power captured by first N components (bottom).

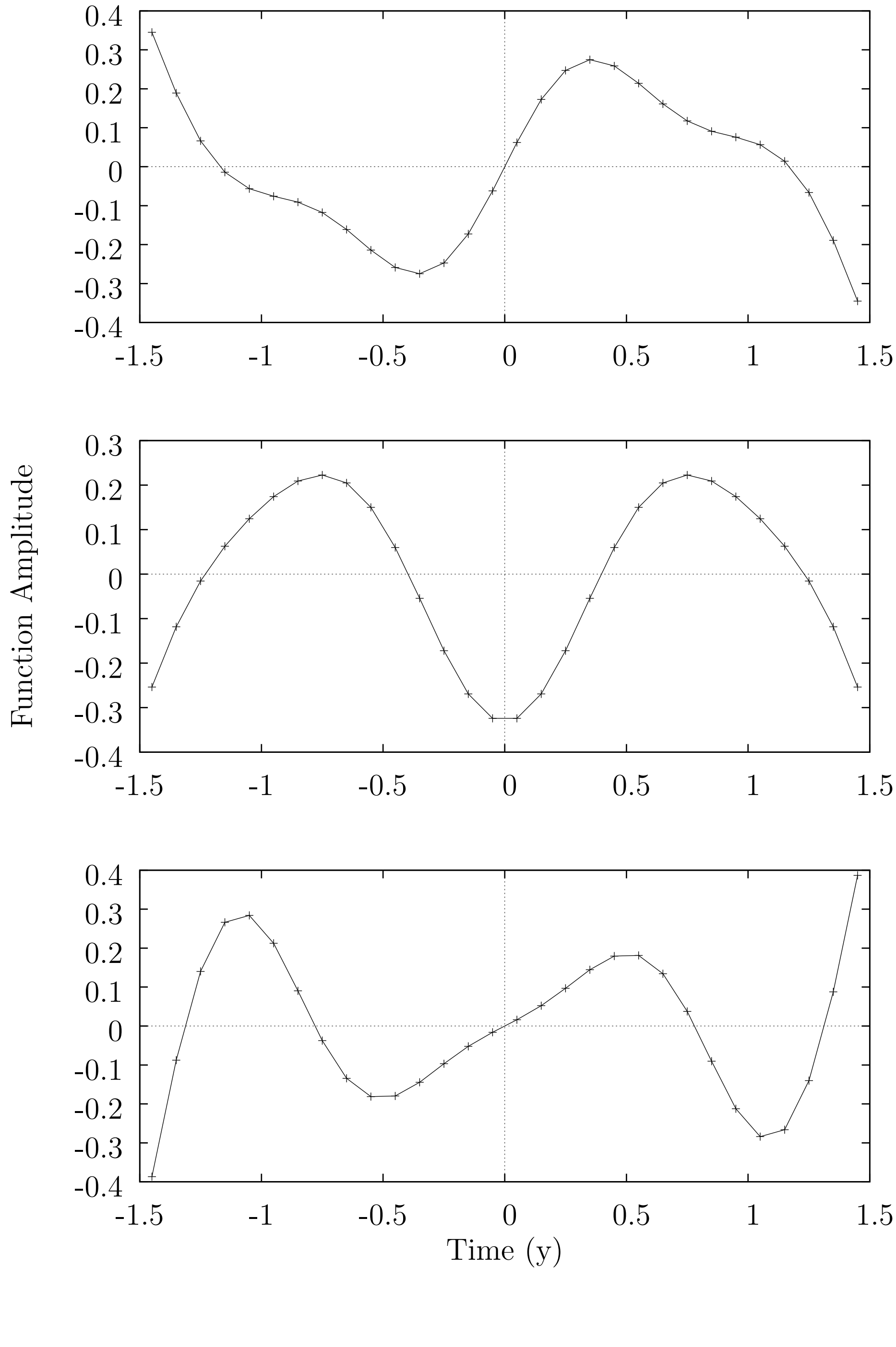

Figure A.3: First few optimal basis functions for a 3 year data span.

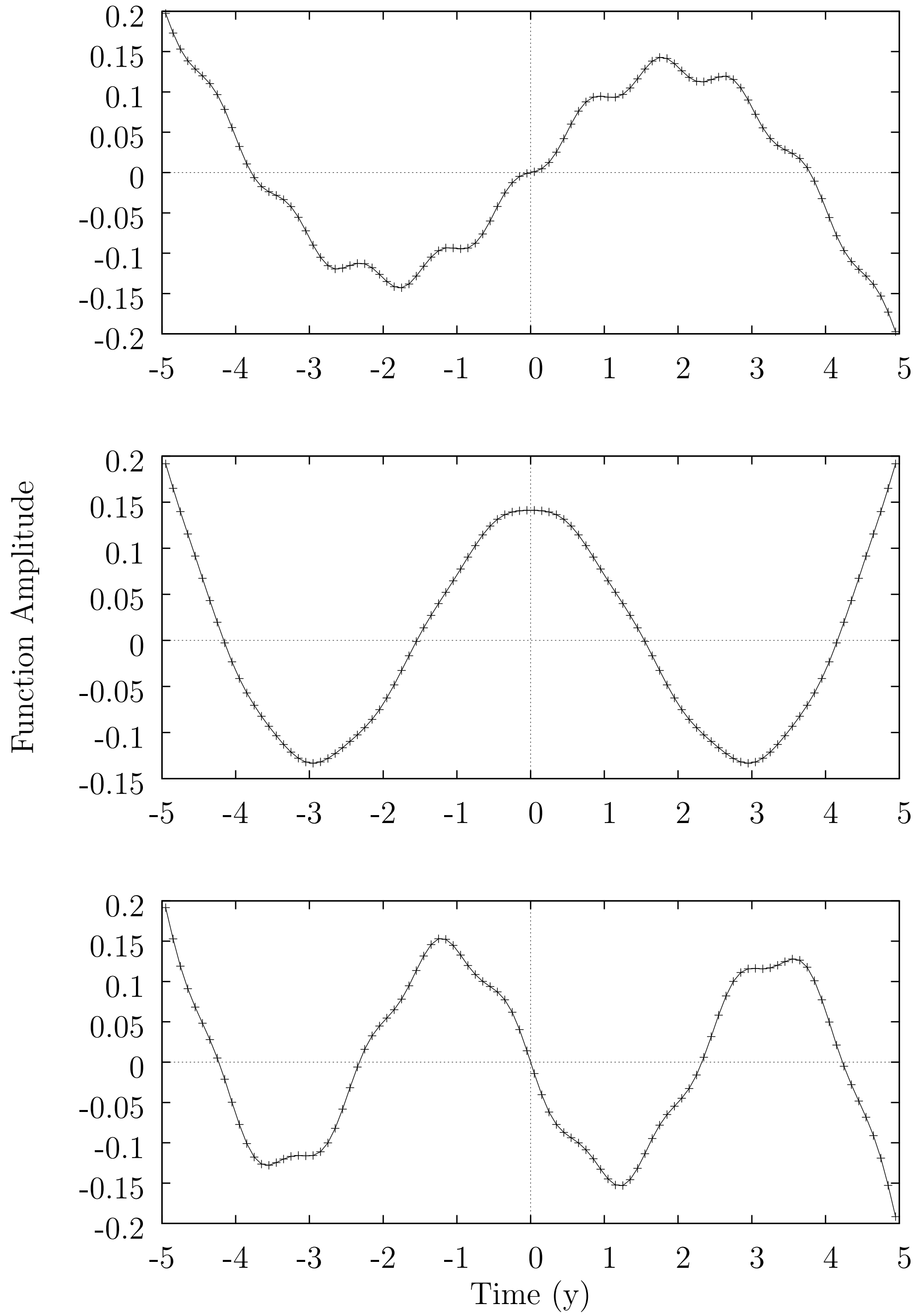

Figure A.4: First few optimal basis functions for a 10 year data span.

rather than total power, is constant. If the form of the spectrum is a power-law as given in Equation A.9, and we know its value $S(\omega_1)$ at some reference frequency ω_1, then $C(0)$ is given by

$$C(0) = S(\omega_1)\frac{\omega_0}{\alpha - 1}\left(\frac{\omega_1}{\omega_0}\right)^{\alpha} \tag{A.11}$$

and Equation A.9 can be scaled accordingly.

A.4 Cross-correlation

A similar situation arises when trying to detect gravitational radiation using the cross-correlation method described in Chapter 5. In this case, we would like to know how the timing fit affects the correlation between two sets of residuals. This can be answered using the framework we have developed here. In this case, we have two input signals, $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$, each of which has gone through a independent fit, giving us two R-matrices, $\mathbf{R}^{(1,2)}$. Assuming we know the expected cross-correlation $\mathbf{\Sigma_y}^{(12)}$ between the input signals, $\Sigma_{yij}^{(12)} \equiv E\left\{y_i^{(1)} y_j^{(2)}\right\}$, the expected cross-correlation of the residuals is then:

$$\mathbf{\Sigma_r}^{(12)} = \mathbf{R}^{(1)}\mathbf{\Sigma_y}^{(12)}\mathbf{R}^{(2)T} \tag{A.12}$$