Gravitational Wave detection
&
data analysis for Pulsar Timing Arrays

Rutger van Haasteren

Gravitational Wave detection
&
data analysis for Pulsar Timing Arrays

Proefschrift

Contents

CONTENTS

1

Introduction

Almost a century ago, Einstein completed his theory of general relativity, thus unifying Newtons law of gravity with special relativity, and radically changing our understanding of space and time. In general relativity, the gravitational force is explained as a manifestation of the curvature of spacetime. Bodies affected by gravity are moving on trajectories referred to as geodesics, the shape of which is determined by the spacetime geometry. This change of paradigm, where space and time are unified into a single manifold that is governed by the Einstein field equations, has far reaching implications for our understanding of the Universe, and has resulted in predictions of several unique phenomena. Some of these predictions were confirmed soon after the completion of general relativity. Others have been confirmed only partially, or remain entirely unconfirmed. This thesis largely focuses on gravitational waves (GWs), one of the partially confirmed predictions of general relativity. Astronomers, by performing precise timing of the Hulse-Taylor pulsar, have been able to convincingly show that GWs exist and are generated by the binary motion in full agreement with the theoretical predictions, resulting in the 1993 Nobel prize in physics (Hulse & Taylor, 1975; Taylor & Weisberg, 1982). However, a direct detection of GWs still has yet to be made.

In the past few decades, scientists have come up with several experimental strategies to detect and measure GWs from astrophysical sources. These strategies have now developed into several large international projects which have made remarkable progress over the past decade. One such class of projects is the Pulsar Timing Array (PTA, Foster & Backer, 1990), which is the focus of this thesis. The basic idea behind the PTA is to detect the GWs by carefully observing pulsars, which are rotating neutron stars that send an electromagnetic pulse towards the

earth every revolution. Optimally extracting a GW signal from PTA data is not a trivial task. Pulsar observations are affected by many systematics, and in order to convincingly show that a GW signal is present in the data, these systematics must be taken into account. In this thesis, a general and robust data analysis method for PTAs is presented. This method is capable of extracting GW signals from PTA observations, correctly taking into account all the systematics that may be influencing the data. The method has been applied to the datasets of the European Pulsar Timing Array (EPTA, Stappers & Kramer, 2011), resulting in the most stringent limit on the stochastic gravitational-wave background (GWB) to date (see chapter 4 of this thesis).

1.1 General Relativity and GR tests

Before the theory of relativity was formulated, the Universe was thought to be static, and space and time were understood to be absolute. Special relativity changed our perspective of space and time, unifying them in one entity, where the notion of space and time is dependent on the reference frame of the observer. General relativity took this one step further, postulating that spacetime is a curved Riemannian manifold, and the curvature is determined by the matter and energy content of the Universe. The spacetime curvature, in turn, has an effect on the trajectories that particles and fields follow. In flat spacetime, a body that experiences no external forces will move along a linear trajectory with constant velocity. In a curved spacetime, the notion of a straight path is substituted with a geodesic, where these geodesics can be thought of as being the shortest path between two points. More precisely, if me measure the "length" of a path by the time passage as shown by a clock attached to the moving particle, then this length between two points has a local maximum for the path that follows a geodesic. In the vicinity of a massive object, bodies that appear to distant observers to be accelerated towards the massive object are residing on the geodesics which are curved towards the world-line of the massive object.

The past century has seen some remarkable experimental triumphs for general relativity. There are a number of observed phenomena that are unexplained within Newtonian gravity, but which general relativity successfully describes, among which:

1) The perihelion advance of Mercury.

2) The bending of light by massive bodies, like the sun.

3) The slow-down of clocks in the gravitational field of the Earth (Pound & Rebka, 1960).

4) The existence of astrophysical black holes (e.g. Schödel et al., 2002).

5) The large body of cosmological data has become very precise in the past two decades, and is successfully described within a paradigm of relativistic cosmology (see e.g. Schrabback et al., 2010; Komatsu et al., 2011, for recent analyses).
6) Shapiro delay, geodetic precession, and the change of orbital period in binary pulsar systems. The latter of these is due to gravitational-wave emission (Taylor & Weisberg, 1989; Stairs et al., 2002; Weisberg & Taylor, 2005; Kramer et al., 2006).

Nevertheless, despite all the observational and experimental tests passed by general relativity, there are still some deep theoretical issues that await resolution. General relativity predicts the existence of singularities, which are regions in space-time where the curvature becomes infinite, and where the physical laws therefore break down. In electromagnetism, the problems related to singularities have been overcome by quantising the electromagnetic field, resulting in quantum electrody-namics, and a similar transition is expected to be found for general relativity. Such a quantum theory of gravity has yet to be formulated, and to this date the presence of singularities in the theory of general relativity remains an unsolved problem. Furthermore, it has recently been established that the expansion of the Universe is accelerating (Riess et al., 1998), which seem in apparent tension with Einstein's field equations without the cosmological constant term. Several explanations have been proposed for this acceleration. One of them involves a yet unknown form of energy with effective negative pressure, usually referred to as "dark energy". An alternative is a modification of general relativity on very large scales (for a re-view, see Peebles & Ratra, 2003). As it stands now, approximately 96% of the mass-energy of the Universe is of a form unknown to us.

As follows from the above discussion, we should expect the theory to break down at very small length scales, and perhaps also at very large length scales. Therefore, new measurable relativistic effects are continuously being sought, and every aspect of general relativity is being directly tested with great accuracy. For example, until recently among the untested predictions of GR were frame-dragging and geodetic precession. The latter is a type of precession which occurs due to the curvature of spacetime around a massive object. A vector, such as the gyroscope axis of rotation, will point in a direction slightly different from the initial one after being transported along a closed path around a massive object. Frame-dragging is the relativistic effect that is caused by the rotation of a massive body, giving rise to the so-called Lense-Thirring precession.

Measurements of both frame-dragging and the geodetic effect were prime tar-gets for the recent NASA mission Gravity Probe B (GP-B, Conklin et al., 2008, and references therein). The GP-B satellite, which was launched in April 2004 and lasted for 16 months, carried a set of four extremely precise gyroscopes. The geodetic and Lense-Thirring precession turn the gyroscopes in different directions, and thus the two can be, in principle, cleanly separated. As of August 2008, the

analysis of the GP-B data has indicated that frame-dragging and the geodetic effect have been confirmed with respectively 15% and 0.1% uncertainty (Conklin et al., 2008). More controversially, the Lunar-Laser Ranging experiment has yielded an indirect confirmation of frame-dragging with 0.1% uncertainty (Murphy et al., 2007).

1.1.1 Black holes

The curvature of spacetime is linked to the stress-energy content of the Universe through Einstein's field equations. Obtaining exact solutions for Einstein's field equations is remarkably difficult, and only a few analytic solutions are known. The first analytic solution was obtained in 1915 by Karl Schwarzschild, and is referred to as the Schwarzschild solution. It describes spacetime in the vicinity of a non-rotating spherically-symmetric object. In the weak-field approximation, far from the massive object, the Schwarzschild solution results in Newton's universal law of gravity. In the strong field limit however, the Schwarzschild solution gives rise to more exotic physics, an extreme case of which is the black hole. A characteristic feature of the Schwarzschild black hole is the existence of a spherical horizon, which causally separates the inside of the hole from the outside (see, e.g. Misner et al., 1973, for a discussion).

The Schwarzschild solution is a special case of a more general class of solutions to Einstein's field equations, only valid for objects that carry no angular momentum. A more general solution is the Kerr solution (Kerr, 1963), and it is this solution that describes effects like frame-dragging. All black holes in the Universe carry angular momentum, and thus all black holes are Kerr black holes. Because the event horizon shields off the interior completely for outside observers, all astrophysical Kerr black holes can be completely described by the following physical parameters:
1) Position.
2) Linear momentum.
3) Angular momentum.
4) Mass.
The statement that these quantities completely describe any black hole in the Universe is referred to as the no-hair theorem (see Misner et al., 1973, for a discussion).

Initially, the scientific community was sceptical about the existence black holes, considering them not much more than a curious mathematical construct. However, over the past 50 years, astronomers have obtained convincing observational evidence that black holes exist. Astrophysical black holes are thought to belong to one of the three classes: the so-called stellar-mass black holes that weigh less than $100M_{\odot}$ (Maeder, 1992), the so-called supermassive black holes

that weigh between 10^6 and $10^{10} M_\odot$ (Kormendy & Richstone, 1995), and the more speculative intermediate-mass black holes with masses in the intermediate range (Portegies Zwart et al., 2004). The stellar-mass black holes are thought to be remnants of very massive, $> 25 M_\odot$ stars, and many of them are seen in x-ray binary systems, where the presence of a black hole is inferred from a combination of the bright x-ray emission and the motion of the companion star. The supermassive black holes reside at the centres of their host galaxies; with the most convincing evidence for a super-massive black hole comes from observations of stellar motion at the dynamical centre of our Galaxy (Schödel et al., 2002). Most galaxies have a spheroidal component (elliptical galaxies, or disk galaxies with a bulge), and current observations with the Hubble Space Telescopes and other instruments indicate that the majority of them host super-massive black hole in their centres (SMBH, Kormendy & Richstone, 1995; Ferrarese & Merritt, 2000). The SMBH at the centre of our Galaxy has a mass of about 4 million solar masses (Ghez et al., 2008; Gillessen et al. A, 2009; Gillessen et al. B, 2009).

1.1.2 Black hole binaries: sources for GW emission

During the evolution the Universe, many galaxies have been formed and are usually clustered together in large dark matter haloes. Mergers of galaxies can be seen in observations, and are expected to have occurred frequently, which suggests that many merger products should exist that host SMBH binaries (Haehnelt, 1994; Volonteri et al., 2003; Menou et al., 2001; Wyithe & Loeb, 2003; Jaffe & Backer, 2003). In this thesis we are interested in black hole binaries because of their expected GW emission; and in particular the SMBH binaries in the centres of galaxies are relevant for the work in this thesis because these binaries are expected to be the main source of the GWs that are detectable by pulsar timing arrays.

Black hole binaries lose energy due to the emission of GWs, which eventually causes the binary to merge. For SMBH binaries, only if the binary separation is smaller than 0.01 of a parsec will the GW emission be efficient enough to shrink the orbit to the point of merger within a Hubble time. Dynamical friction is expected to bring the SMBH components to within several parsecs, but thereafter some other mechanism must dissipate the orbital energy of the SMBH binary in order for the binary to be able to merge. As a general consensus in the scientific community has not yet been reached on what this mechanism is, the question of how SMBH binaries tighten is usually referred to as the "final parsec problem" (Roos, 1981; Milosavljević & Merritt, 2001). However, from recent theoretical work a picture is emerging that both stellar-dynamical and gas-dynamical processes may be efficient in overcoming the final parsec on the binaries' way to merger; see Merritt & Poon (2004); Berczik et al. (2005); Cuadra et al. (2009); Dotti et al. (2007); Escala et al.

(2005); Perets & Alexander (2008); Khan et al. (2011).

If this is indeed the case, then, as we explain below, there should be a sufficient number of merger events in the universe for a GW detection of these systems to become plausible with currently planned GW detection projects.

The GW emission of a black hole binary merger can be divided roughly in three phases: the inspiral, the merger, and the ringdown phase (Flanagan & Hughes, 1998). During the inspiral, the binary radiates GWs with ever increasing frequency while reducing the separation between the two objects up to the point of merger. During the merger phase, the two individual SMBHs become one, in the process converting a significant portion of their rest mass in GWs. During the ringdown, the merger product relaxes asymptotically to a Kerr black hole by radiating the remaining perturbations as gravitational waves.

Detection of GWs of these different phases requires a different kind of GW detector because of the different frequency and duration of the signal. A single black hole binary spends a relatively long time in the inspiral phase compared to the time-scale of the actual merger. Detectors sensitive to high frequency GWs are therefore more likely to observe the black hole binary merger as single "events", whereas detectors sensitive at lower frequencies are more likely to observe continuous signals present throughout the entire dataset.

Pulsar Timing Array projects are sensitive to the ultra-low frequencies of the GW spectrum, typically in the range of several tens of nanoHertz. The frequency range of the PTAs is determined by the interval between the observing runs, which are typically of the order of several weeks and by the durations of the PTA experiment. The current high precision datasets span time-scales of a few years to over a decade. Therefore, PTAs are sensitive to the inspiral phase of SMBH binaries where the orbital period of the binary is of the order of months to years. Although in principle single SMBH binaries can be detected with PTAs, the many SMBH binary inspiral signals will add up to a so-called stochastic background of GWs (Begelman et al., 1980; Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003; Sesana et al., 2008). It is this stochastic gravitational-wave background (GWB) that is thought to be the prime candidate for detection by PTAs.

Another effect resulting from SMBH binary mergers could be of interest for PTA projects. The GWs generated by a single merger consists of a high-frequency ac-part, and a dc-part (Christodoulou, 1991; Thorne, 1992; Favata, 2009, see Figure 1). The ac-part, as mentioned above, is short-period and short-lived, and hence is undetectable by PTAs observing less frequently than daily. The dc-part, referred to as the gravitational-wave memory, grows rapidly during the merger, with the metric change persisting after the gravitational-wave burst has passed. As shown in chapter 3, these gravitational-wave memory bursts may be detectable by PTAs.

1.2 Pulsars and pulsar timing

Since their discovery (Hewish et al., 1968), pulsars have become probes of fundamental physics and astrophysics. Firstly, their extreme compactness allows one to study relativistic matter at supra-nuclear densities (see Haensel et al., 2007, for an overview). Secondly, it was realised that the rotation of a pulsar, and the pulsations it produces in the radio-band, are so stable that pulsars can be used as nearly-perfect Einstein clocks. This has allowed unprecedented tests of general relativity by using pulsar timing as a tool to accurately probe the orbits of pulsars in tight binaries (Taylor & Weisberg, 1989; Stairs et al., 2002; Weisberg & Taylor, 2005; Kramer et al., 2006).

In this section, we present some highlights of the successes of pulsar timing. In Section 1.2.1 we introduce the basics of pulsar timing, after which we briefly review general relativity tests with pulsar timing in Section 1.2.2, and current efforts to detect GWs with pulsar timing in Section 1.3.2.

1.2.1 Pulsar timing

Single pulses vary greatly in shape and intensity, even in the absence of the dispersion due to the interstellar medium (see e.g. Cordes & Shannon, 2010, and references therein). Single pulses are therefore not useful for pulsar timing purposes. The average pulse profile, calculated by stacking a large number of pulses (the typical range is from several thousand up to several million), is very stable, with a specific shape unique to each pulsar and radio frequency band (see Lorimer & Kramer, 2005, for an extensive description of Pulsar Timing). It is this average pulse profile that is very suitable for timing purposes, and it has been utilised with great success for the past few decades. See Figure 1.1 for an example of the average pulse profile and the single pulses it is composed of.

Not all pulsars have similar properties in terms of pulse regularity and profile stability. Typically, pulsars have a relatively stable rotational frequency, and a low negative rotational frequency derivative. During their lifetime, the pulse period increases due to the emission of electromagnetic radiation (Lorimer & Kramer, 2005), a process known as spindown or "quadratic spindown". The lower the magnetic field of the pulsar is, the lower is the spindown rate of the pulsar.

Contrary to what one would guess based on the previous description of the spin properties of pulsars, the fastest spinning pulsars are also the oldest. This special class of pulsars is referred to as the millisecond pulsars (MSPs, Backer et al., 1982). Almost all of these pulsars are in binaries, which ought to be closely related to their fast spin. It is thought that the pulsar companion star had in the past grown to become a red giant, losing mass that subsequently accreted on to the
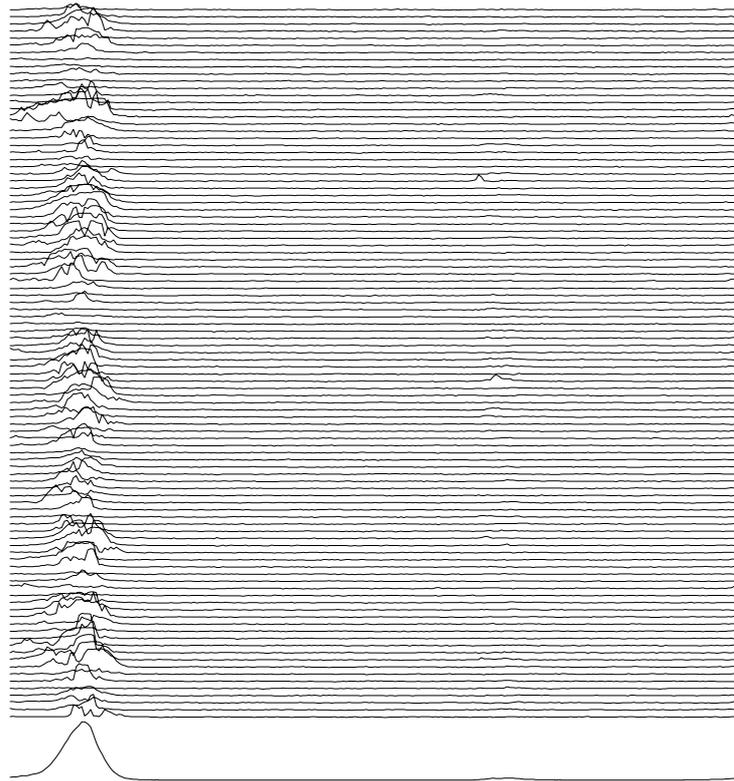
Figure 1.1: Plot of pulses of the 253ms pulsar B0950+08; intensity of the single pulse versus rotational phase. Single pulses of this pulsar are shown at the top, demonstrating the variability in pulse profile shape and intensity, and a 5 minute average pulse profile of the same pulsar is shown at the bottom. The 1200 pulses of this average pulse profile already approach the reproducible standard profile for this pulsar. Observations taken with the Green Bank Telescope. Figure from Stairs (2003), reproduced with permission.

pulsar (Alpar et al., 1982). In such an accretion process, a small part of the orbital angular momentum of the accretion disc is converted into spin angular momentum of the pulsar, thereby greatly increasing the pulsar's spin frequency. Typically, the pulse period is reduced in this fashion to the order of several to several tens of milliseconds. A canonical millisecond pulsar has a very high and stable rotational frequency, a low magnetic field (presumably due to the suppression of the original field by the accretion process), and a low spindown rate, which makes millisecond pulsars very suitable for precision timing experiments.

1.2.2 GR tests with pulsar timing

The precise timing of (millisecond) pulsars allows very accurate modelling of their trajectories. In the case a pulsar is in a binary, the orbital parameters of the binary can be inferred with great precision. Almost 40 years ago, Hulse & Taylor (1975) discovered the first double neutron star system, PSR B1913+16, nowadays often referred to as the Hulse-Taylor binary. One of the two objects of the Hulse-Taylor binary is a radio pulsar, orbiting its companion in only 7.75 hours. This system is therefore in such a tight orbit with its companion that the relativistic effects are well measurable. In 1993 the Nobel prize in physics was awarded to Hulse & Taylor for their discovery of the Hulse-Taylor binary. With observations of that binary, its was possible to convincingly show that the decrease in orbital period was due to the emission of GWs, thereby providing the first proof for the existence of GWs (Taylor et al., 1979; Taylor & Weisberg, 1989). This proof is generally considered indirect evidence that GWs exist, since it only establishes that the energy loss is due to GW emission. A direct detection would have to consist of evidence that GWs are present at some location other than where the GWs were emitted.

In order to test gravitational theories with pulsar timing, the Post-Keplerian (PK, Damour & Deruelle, 1986) formalism is often used, where the PK parameters systematically describe the deviations of motion of binary systems from a classical "Keplerian" orbit as would follow from Newtons laws of motion and gravitation. Five parameters are required to fully describe a Keplerian orbit. Usually the orbital period P_b, the eccentricity e, the projected semimajor axis $x = a \sin i$ (with a the semimajor axis, and i the inclination), the time of periastron T_0, and the longitude of periastron ω are used (Damour & Deruelle, 1986). In any given theory of gravity, the PK parameters can be written as functions of the easily measured Keplerian parameters, the two masses of the orbiting objects, and the two polar angles defining the direction of the pulsar spin axis.

Testing a theory of gravity with the PK parameters is quite straightforward. The PK parameters can both be inferred from observations, and calculated using the respective theory of gravity, which allows a number of independent verifications

depending on how many PK parameters can be determined. In general relativity the PK parameters can be written as a function of the two masses of the orbiting objects, from which it follows that determining two PK parameters is required to find the two masses of the system. When more than two PK parameters can be inferred from the observations, the extra PK parameters each offer an independent test of the theory of gravity, since the now overdetermined system of PK parameters should form a consistent whole. See Figure 1.2 for an example of such a test for general relativity.

For the Hulse-Taylor binary, three post-Keplerian parameters have been measured: the advance of periastron $\dot{\omega}$, the gravitational redshift γ, and the secular change of the orbital period due to gravitational-wave emission $\dot{P}_b$. Because $\dot{P}_b$ could be determined from the other parameters of the system, and observed directly from the data, this system provided the first confirmation of the existence of GWs. Recently, a new very tight binary system has been discovered, where both binary components are pulsars (Burgay et al., 2003). Usually referred to as the "double-pulsar system", this system contains one millisecond pulsar with a pulse period of 23ms, and one normal pulsar with a period of 2.8s. This system has proved to be an excellent laboratory for the testing of gravitational theories, and within a few years of timing five post-Keplerian parameters had been accurately measured (Kramer et al., 2006). The binary pulsar has yielded the most precise tests of general relativity to date, with results for the combination of most stringent PK parameters confirming the validity of general relativity at the 0.05% level (Kramer et al., 2006). See Figure 1.2 for the consistency of the PK parameters of the double-pulsar system with general relativity.

1.3 Current GW experiments, and PTAs

Projects aiming to detect GWs currently receive much attention from the scientific community. Several different approaches are being taken, including earth-based and space-based laser interferometers, resonant mass detectors, and pulsar timing based efforts. The first confirmed detection will be ground breaking, no matter which of these efforts manages to acquire enough evidence to claim this much desired feat. But besides the common goal of verifying the existence of GWs, the different approaches will be complementary because of their coverage of different frequency bands, each capable of exploring different astrophysical environments.

1.3.1 GW detectors

Gravitational waves are propagating perturbations in the curvature of spacetime, and hence any detection method is in some way sensitive to changes of the Rie-

Figure 1.2: The constraints on the masses of the two neutron stars in the double-pulsar system PSR J0737-3039A/B. The Post-Keplerian parameters are given by the different lines:

$\dot{\omega}$: precession of periastron

γ: time dilation gravitational redshift

r: range of Shapiro time delay

s: shape of Shapiro time delay

$\dot{P}_b$: secular change of the orbital period

The white wedge shows the allowed region due to the fact that for the inclination angle: $\sin i \leq 1$, and the solid diagonal line comes from the measurement of the mass ratio R. For a consistent gravitational theory, the Post-Keplerian parameter curves should intersect in one point. The inset shows an expanded view of the region around this intersection. From Kramer et al. (2006), reproduced with permission.

mann curvature tensor. The first class of detectors are the so-called resonant-mass detectors. These devices consist of a large mass of material, optimised in such a way that the vibrational modes are minimally damped, and easily excited by GWs of some specific frequency band. Typically these detectors are sensitive to frequencies in the kHz range, making supernova explosions a prime candidate of sources for these projects (see Levine, 2004, for an overview of early Weber bar experiments).

Another class of detectors are the laser interferometers. These detectors are in principle large Michealson interferometers, relying on the fact that a GW will affect the propagation of the laser beam differently in the different arms of the detector. This results in a detectable relative phase change between the two beams at the point of recombination. Currently, several ground based laser interferometers are or have been operational:

1) The Laser Interferometer Gravitational-Wave Observatory (LIGO), which consists of two observatories in the United States (Abbott et al., 2009). The two observatories are identical in setup, and have arms of 4km.

2) The France-Italy project Virgo, at Cascina, Italy, which consist of two 3km arms (Acernese et al., 2006).

3) Geo600, near Sarstedt, Germany, which consists of two 600m arms (Grote et al., 2008).

4) Tama300, located at the Mitaka campus of the National Astronomical Observatory of Japan (Takahashi et al., 2008), with two 300m arms.

These projects are continuously improving their sensitivity, and are already producing interesting upper-limits on the GW content of the Universe (e.g. Abadie et al., 2010). Typical GW frequencies that these ground based laser interferometry detectors are sensitive to are in the range of tens of Hertz to a few thousand Hertz (Abbott et al., 2009). Another class of laser interferometers are the proposed space based interferometers, like the Laser Interferometer Space Antenna (LISA, see Prince & LISA International Science Team, 2011, for an overview). LISA will work according to the same principles as the ground based interferometers, but by going to space and increasing the arm length to 5 million kilometres many noise sources that are featuring in the ground-based interferometers will be eliminated. The three LISA satellites will be placed in solar orbit at the same distance from the Sun as the Earth, but lagging behind by 20 degrees. Others space based observatories have been proposed over time (see e.g. Crowder & Cornish, 2005), but so far, only for LISA have definite arrangements been made, with the LISA Pathfinder currently planned for launch in 2013 (Racca & McNamara, 2010).

1.3.2 Pulsar Timing Arrays

Pulsar Timing Array projects rely on a principle similar to the laser interferometry experiments: the propagation of the electromagnetic pulses from the pulsar towards the earth is affected by GWs. This results in a phase change on earth, i.e. the arrival time of the pulse is slightly changed by the GWs. Because of the stability of the rotational period of the pulsar, this altered pulse arrival time may be detectable.

An important quantity in pulsar timing array experiments is the timing residual: the deviation of the observed pulse time of arrival from the expected pulse time of arrival based on our knowledge of the motion of the pulsar and the strict periodicity of the pulses. Mathematically, the contribution of a GW to the timing residual can be expressed as follows. Consider a linearly polarised plane GW propagating towards the earth. We denote the start of the experiment by t_0, δt_i^{GW} is the GW-induced timing residual of an observation at time t_i, and $\delta t_i^{\mathrm{other}}$ is the timing residual induced by anything but a GW. The total timing residual can then be written as:

$$\delta t_i = \delta t_i^{\mathrm{GW}} + \delta t_i^{\mathrm{other}}. \tag{1.1}$$

Setting the speed of light $c = 1$, the timing residual δt_i^{GW} is generated according to the following equation:

$$\delta t_i^{\mathrm{GW}} = \int_{t_0}^{t_i} \mathrm{d}t \frac{1}{2}\cos(2\phi)\,(1 - \cos\theta)\,[h(t) - h(t - r - r\cos\theta)] \tag{1.2}$$

where θ is the angle between the direction of the GW propagation and the earth-pulsar direction, ϕ represents the polarisation of the GW in the plane perpendicular to the propagation direction, r is the distance between the earth and the pulsar, $h(t)$ is the metric perturbation at the earth when the pulse was received, and $h(t - r - r\cos\theta)$ is the metric perturbation at the pulsar when the pulse was emitted.

PTA projects rely on observations of astrophysical processes of which not all the details are known. Also, in contrast with the man-made detectors, the propagation of the pulses is influenced by non-GW processes of astrophysical origin, which are impossible to remove by careful isolation of the experimental equipment. These complications therefore require very careful modelling of the data, where all known and unknown noise contribution have to be taken into account when extracting GW information. A rigorous method must be used to distinguish the GW signal from noise sources to convincingly show that the signal was indeed due to a GW, and not due to some unmodelled noise contribution.

When only a single pulsar is observed, it is difficult to convincingly show that a particular contribution to the total timing residual of equation (1.1) is due to GWs, and not because of some other astrophysical process. However, in equation

(1.2), the geometrical pre-factor $\cos(2\phi)\,(1-\cos\theta)$ is a unique feature of general relativity that is only dependent on the position of the pulsar, and the $h(t)$ term is the same for any pulsar that may be observed. Therefore, if one observes a whole array of pulsars, not just one, it is in principle possible demonstrate that a particular contribution to the timing residual is due to GWs (Foster & Backer, 1990; Jenet et al., 2005). In that case, all the timing residuals of all pulsars will contain a contribution which is proportional to $h(t)$, correlated between pulsars with a coefficient unique to general relativity. This feature allows convincing extraction of a GW signal from Pulsar Timing Array observations, and forms the basis of all PTA data analysis techniques.

The prime target of gravitational-wave sources PTAs try to detect is the stochastic background generated by an ensemble of SMBH binaries. This background is thought to be isotropic, which alters the correlations between the residuals of different pulsars. For such an isotropic background, the correlations for any pulsar pair only depend on the angular separation between the two pulsars, as shown in Figure 1.3.

Currently, three independent groups have established PTAs:
1) The Australian-based programme PPTA, the Parkes Pulsar Timing Array, which uses data from the Parkes radio telescope (Hobbs et al., 2009; Verbiest et al., 2010), and archival Arecibo data.
2) The North-American based programme NANOGrav, North-American Nanohertz Observatory for Gravitational waves, which uses both the Green Bank Telescope (GBT), and the Arecibo radio telescope (Jenet, 2009).
3) and the European programme EPTA, European Pulsar Timing Array (Stappers & Kramer, 2011), which uses five different radio telescopes: the Lovell telescope near Manchester, United Kingdom, the Westerbork Synthesis Radio Telescope (WSRT) in the north of the Netherlands, the Effelsberg Telescope (EFF) near Bonn in Germany, the Nançay Radio Telescope (NRT) near Nançay in France, and the Sardinia Radio Telescope (SRT) in Sardinia, Italy, which is expected to become operational in 2011(Tofani et al., 2008).

Besides their independent efforts, these three PTA groups have also started to join their efforts in an umbrella project: IPTA, International Pulsar Timing Array (Hobbs et al., 2010). It is likely that the first detection of GWs will occur as a result of a joint effort of the IPTA.

1.4 Bayesian PTA data analysis

The data of PTA projects contain contributions of many different processes, some of which are well modelled, while others are less well understood. In the duration

Correlation induced by stochastic GW background

GW induced correlation

Pulsar separation [deg]
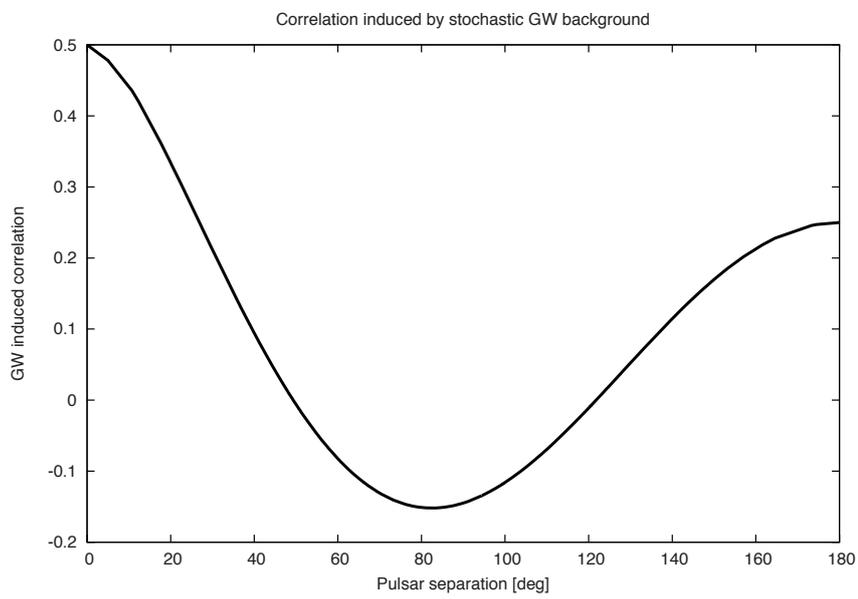
Figure 1.3: The Hellings & Downs curve, which shows the correlations in the timing residuals between two pulsars as induced by a stochastic gravitational-wave background, as a function of their angular separation.

of the experiment, equipment may change at the observatories, and the observations are taken irregularly, sometimes with large gaps in between observations. On top of that, the signal duration is comparable to the duration of the experiment, which requires extra care when performing the data analysis. In this thesis, a data analysis method for Pulsar Timing Array projects is presented that is capable of dealing with all of the above complications.

In this thesis, we develop a Bayesian approach to the data analysis of Pulsar Timing Array projects. The general idea is to (a) first assume that the physical processes which produce the timing can be characterised by several parameters, and then to (b) use Bayes theorem to derive the probability distribution of the parameters of our interest. In our case, (a) enables us to write down the likelihood P(data|parameters), which in essence means that we need to have a generative model for the data. Then, by Bayes theorem, the posterior distribution of the parameters is given by:

$$P(\text{parameters}|\text{data}) \quad = \quad P(\text{data}|\text{parameters}) \times$$
$$\times \frac{P_0(\text{parameters})}{P(\text{data})}. \qquad \boxed{1.3}$$

Here P_0(parameters) is the prior probability which represents all our current knowledge about the unknown parameters, and P(data) is the marginal likelihood, which in a Bayesian approach is a goodness of fit quantity that can be used for model selection. The marginal likelihood can be thought of as a normalisation factor that ensures that P(parameters|data) integrates to unity over the parameter space.

Especially when one has to deal with data containing contributions of numerous different processes, the posterior P(parameters|data) is a function of many parameters, in the case of Pulsar Timing Arrays the number of parameters can go up to several 100s. Properly presenting a distribution of that many parameters is highly impractical, and typically one is only interested in a small subset of the parameters. The usual procedure in Bayesian inference is to integrate the posterior distribution over the "uninteresting" parameters, referred to as "nuisance parameters". This process is referred to as marginalisation, and results in the posterior distribution function of only the interesting parameters. In the case of Pulsar Timing Arrays, there can be several 100s nuisance parameters, and numerically performing the marginalisation process is a serious challenge.

Another serious computational challenge that arises when resorting to Bayesian data analysis, is the evaluation of the above mentioned marginal likelihood. This quantity is essential when one needs to perform model selection, and is notoriously difficult to evaluate using standard techniques. Fortunately, specialised schemes exist to tackle both the evaluation of the marginal likelihood, and to produce the

marginalised posterior distributions. In this thesis, we use in a novel way Markov Chain Monte Carlo methods to produce the marginalised posterior distribution, and in chapter 5 we introduce a new method to calculate the marginal likelihood using the samples of MCMC simulations.

1.4.1 Modelling the PTA data

The pulse times of arrival of realistic data sets are (pre-) processed in several independent steps, a more detailed description of which is presented in chapter 4. For our understanding here, it suffices to note that timing residuals are produced by fitting the pulse times of arrival with the timing model. The timing model is a multi-parameter fit that represents our best knowledge of the many deterministic processes that influence the values of the arrival times of the pulses. Examples of parameters in the timing model are the rotational frequency of the pulsar, the spin-down rate, pulsar position on the sky and its proper motion, and if the pulsar is a member of a binary, the (Post-) Keplerian parameters of its orbit.

Besides the timing model parameters, which describe the deterministic behaviour of the pulse arrival times, we also have to model the stochastic contributions to the timing residuals. Currently, the statistics of this stochastic component is not well understood. On the technical side, the error bars, representing the radiometer noise of the receivers at the observatories, are not fully trusted, and usually a dataset-specific multiplicative coefficient is used to calibrate the uncertainties of the observations (Hobbs et al., 2006). On the astrophysical side, we only have empirical models for the irregularity in the pulsar beam rotation that causes low frequency noise in the timing residuals. We model all these stochastic processes as a random Gaussian process, where we parametrise the power spectra of the different processes individually. Also, the observations of the EPTA are taken with several telescopes. These timing residuals of different telescopes have to be calibrated for offsets with respect to each other. On top of that, hardware changes at each telescope may have taken place, which also requires calibration.

Summarising the treatment of the parameters we use to model the observations of the EPTA (see chapter 4 for details), the likelihood function contains the following contributions that describe the pulse arrival times:

1. The timing model, usually consisting of over 10 parameters for each pulsar.
2. A random calibration offset between observations of the same pulsar, taken with different equipment or at different observatories. These so-called jumps are introduced whenever hardware changes have taken place at the observatories.
3. The error bars on the arrival times of the pulses, including a dataset-specific calibration coefficient, commonly referred to as an "EFAC" factor.
4. A white noise component of unknown amplitude for each dataset.

5. Pulsar specific red timing noise, characterised as a power-law spectrum with unknown amplitude and spectral steepness.

6. The gravitational-wave background, consisting of a power-law spectrum of unknown amplitude and spectral steepness, correlated among all pulsars.

For a typical PTA, consisting of 20 pulsars, the number of parameters we have to take into account in the analysis is of the order of several 100s. Fortunately, an analytical shortcut exists such that we do not have to numerically marginalise over the parameters of the timing model, which reduces number of parameters to about one hundred.

1.4.2 Markov Chain Monte Carlo

Numerically marginalising the full posterior distribution function, shortly the posterior, in one hundred dimensions is a daunting challenge, especially when the values of the posterior are computationally expensive to evaluate. In the case of the observations used in chapter 4, the evaluation of a single value of the posterior takes about a second on a modern workstation. However, the computational time required for a single evaluation of the posterior scales as n^3, with n the total number of observations of the PTA. Full datasets of current and future PTAs will take much more computational resources than the ~ 1000 CPU hours it took to produce the results of chapter 4, since we have only used a fraction of the already available datasets of the EPTA in that analysis.

For the PTA model we have introduced in Section 1.4.1, the posterior is usually a unimodal distribution, and is, especially when the data is informative, only of non-negligible value at a small fraction of the parameter space. In Bayesian inference, it is very common to deal with high-dimensional distributions with very narrow peaks compared to the size of the parameter space. To illustrate the problems this creates in high-dimensional parameter spaces, consider a uniform distribution on a small interval:

$$p\left(\vec{x}\right) = \begin{cases} 10^n & \text{if } \forall i : \frac{-1}{10} < x_i < \frac{1}{10}, \\ 0 & \text{otherwise} , \end{cases} \qquad \boxed{1.4}$$

where the index $i = 0, 1, \ldots n$, with n the number of dimensions, and $x_i \in [-1, 1]$. In one dimension, this distribution is non-zero at a fraction of $1/10$ of the entire parameter space, but already at 9 dimensions, this fraction is reduced to one billionth. Therefore, picking a sample at random from the whole parameter space, for $n = 9$ we have virtually no chance of hitting a point with non-zero $p(\vec{x})$. This property of high dimensional parameter spaces is referred to as the curse of dimensionality.

The curse of dimensionality has dramatic effect on direct integration techniques. Performing an integration of the distribution of equation (1.4) on a regular

grid would require at least a billion samples in 9 dimensions, only one of which is will on average effectively add to the value of the integral. In this case the precision of the estimated value of the integral is going to be quite low, and we have virtually no hope of evaluating this integral in much higher dimensions with this technique.

A common class of techniques to numerically evaluate these kinds of integrals make use of random sampling of the parameter space, usually referred to as Monte Carlo methods. The above mentioned curse of dimensionality can be overcome by changing the distribution from which we sample from a uniform distribution over the entire parameter space, to a distribution that has a higher probability to yield samples in the region of parameter space where equation (1.4) is non-zero. By doing that, more samples are effectively adding towards the integral, thereby increasing the precision of the evaluation. In this thesis we rely on Markov Chain Monte Carlo methods, more specifically Metropolis Hastings, to generate samples with that property. This way, we can evaluate integrals over the posterior in a parameters space that includes all the parameters mentioned in Section 1.4.1.

1.5 Thesis summary

In chapter 2, the basic Bayesian data analysis method is introduced. First, the theoretical framework for Bayesian data analysis is explained, covering the concepts of parameter estimation and marginalisation. Then, the theory behind the GW-induced timing residuals of PTAs is dealt with, focusing on signals coming from a stochastic GWB. All this is finally combined to form the basis of the method by constructing the likelihood function.

We motivate our Bayesian method by noting several of its strengths.
(1) It analyses the data without any loss of information. Extracting a signal of the complexity of the GWB is a non-trivial task, and by using the Bayesian framework, by construction we are ensured optimality.
(2) It trivially removes systematic contribution to the pulse times of arrival of known functional form, including quadratic pulsar spin-down, annual modulations and jumps due to a change of equipment. Serious systematics are continuously dealt with when producing pulsar timing data, all of which affect the form and strength of the GWB signal in some way. Marginalisation over all of the systematics enables us to correctly deal with these effects.
(3) It measures simultaneously both the amplitude and the slope of the GWB spectrum. To date, no other data analysis method exists that can reliably extract both these signal parameters from the data.
(4) It can deal with unevenly sampled data and coloured pulsar noise spectra. By taking the unknown stochastic timing noise into account simultaneously with the

GWB signal extraction, we are ensured not to confuse the two, mistakenly identifying timing noise as a GWB or vice versa.

We extensively test our approach on mock PTA datasets, and we show that the experiments signal-to-noise (S/N) ratio strongly decreases with the redness of the pulsar timing noise, and strongly increases with the duration of the PTA experiment. In order to illuminate some key aspects important for formulating observing strategies for PTA experiments, we also carry out some parameter studies where we explore the dependence of the S/N ratio on the duration of the experiment, number of monitored pulsars, and the magnitude of the timing noise.

In chapter 3, we show that, next to the stochastic GWB, PTAs are also sensitive to individual physical mergers of SMBHs. While the mergers occur on a time-scale too short to be resolvable by PTAs, they generate a permanent signal referred to as the gravitational-wave memory effect. This is a change of metric which persists for the duration of the experiment, and could potentially be detectable. We extend the theory presented in chapter 2 to single source detection in general, and then apply the analysis method to the gravitational-wave memory effect in particular. Our analytical estimates show that individual mergers of of $10^8 M_\odot$ black holes are 2-σ-detectable (in a direction, polarization, and time-dependent way) out to co-moving distances of $\sim$ 1 billion light years. We test the analysis method on mock data, in a manner similar to what we have done for the GWB in chapter 2, and find that the analytical estimates are applicable even when taking into account all the systematic contributions of known functional form like quadratic spindown.

The Bayesian inference method for GWB detection is applied to real European Pulsar Timing Array (EPTA) data in chapter 4. In this chapter, we show how to use the method of chapter 2 when working with real observations, introducing a robust prescription on how to calculate an upper limit on the GWB in such case. The data sets of the EPTA have varying duration, regularity, and quality, taken with multiple telescopes. Our approach to the data analysis can serve as a useful template for future intercontinental PTA collaborations.

Parametrising the GWB as a power-law of the form $h_c(f) = A(f/\mathrm{yr}^{-1})^\alpha$, where f is the GW frequency, the current upper limit on the GWB, calculated with EPTA data, is $h_c \leq 6 \times 10^{-15}$ in the case of $\alpha = -2/3$, as predicted for a GWB created by an ensemble of supermassive BH binaries (Maggiore, 2000; Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003; Sesana et al., 2008). This is 1.8 times lower than the 95% confidence GWB limit obtained by the Parkes PTA in 2006 (Jenet et al., 2006). More generally, the analysis has resulted in a marginalised posterior as a function of the parameters of the GWB: the GWB amplitude and the spectral index.

Finally, in chapter 5 we introduce a numerical method to evaluate the marginal likelihood, using the results of standard MCMC methods like Metropolis Hastings.

In Bayesian data analysis, the calculation of the marginal likelihood has proved to be a difficult quantity to evaluate using standard techniques. Some specialised numerical methods do exist (including Newton & Raftery, 1994; Earl & Deem, 2005; Skilling, 2004; Feroz et al., 2009), but it has not been shown that any one of these is general and efficient enough to be the best choice in all cases. We show that it is possible to calculate the marginal likelihood by utilising only high probability density (HPD) regions of the posterior samples of standard MCMC methods. Especially in cases where MCMC methods sample the likelihood function well, this new method produces very little computational overhead, provided that the values of the likelihood times the prior have been saved together with the values of the parameters at each point in the MCMC chain. We test the new method on several toy problems, and we compare the results to other methods. We conclude that the new method could be of great value, provided that the problem is well-suited for MCMC methods. However, in its current form the method has some drawbacks as well. We demonstrate that it suffers when correlated MCMC samples are used; such correlated MCMC samples are produced for instance by Metropolis-Hästings. We show that the use of correlated MCMC samples significantly increases the uncertainty of the marginal likelihood estimator compared to when a chain of statistically completely independent samples is used. We have also found that the new estimator for the marginal likelihood is slightly biased, where the bias is problem dependent. Additional tests to assess this bias are required before this new method is suitable for scientific applications.
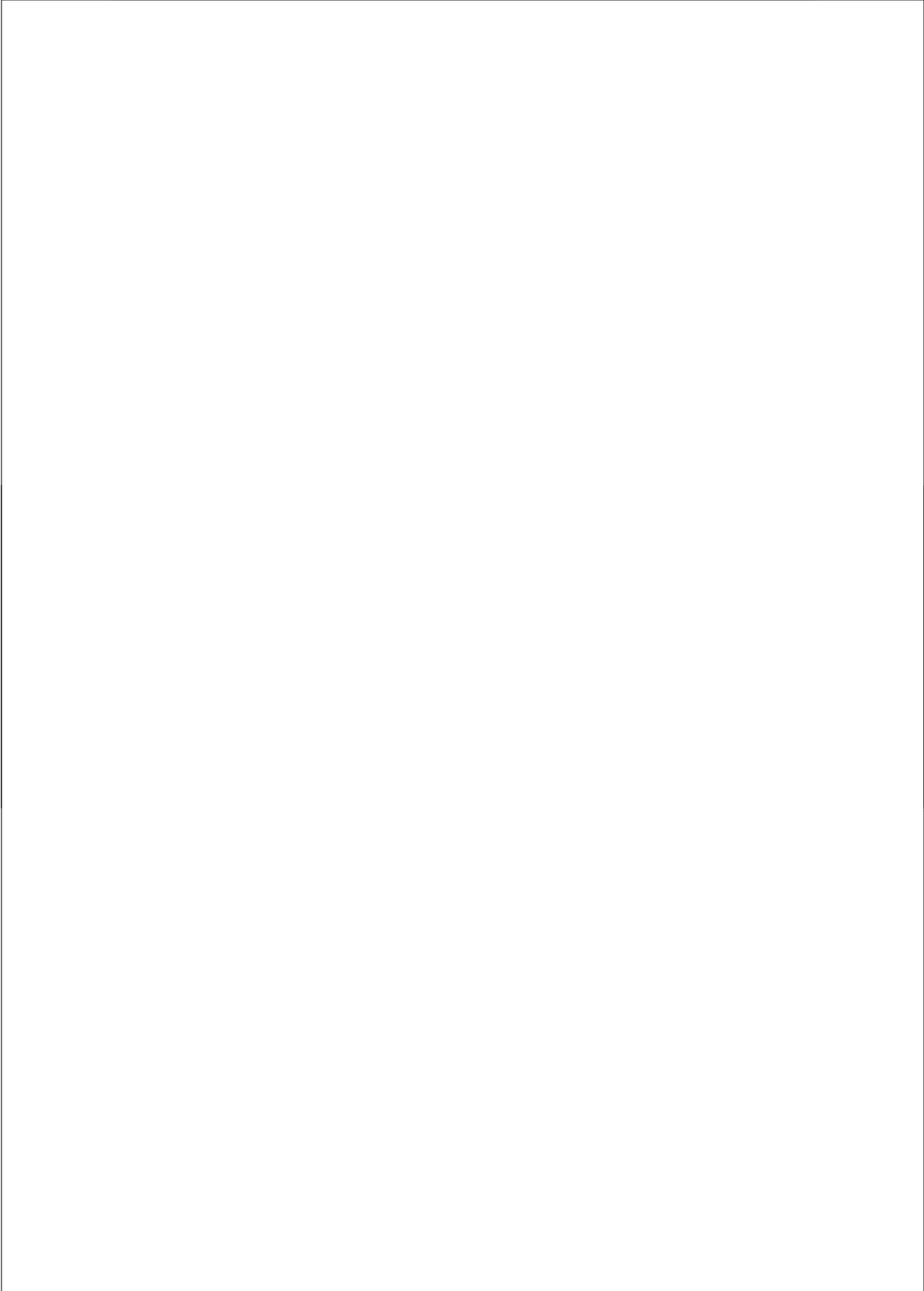
2

On measuring the gravitational-wave background using Pulsar Timing Arrays

Science may be described as the art of systematic over-simplification.

Karl Popper

Abstract

Long-term precise timing of Galactic millisecond pulsars holds great promise for measuring the long-period (months-to-years) astrophysical gravitational waves. Several gravitational-wave observational programs, called Pulsar Timing Arrays (PTA), are being pursued around the world.

Here we develop a Bayesian inference method for measuring the stochastic gravitational-wave background (GWB) from the PTA data. Our method has several strengths: (1) It analyses the data without any loss of information, (2) It trivially removes systematic errors of known functional form, including quadratic pulsar spindown, annual modulations and jumps due to a change of equipment, (3) It measures simultaneously both the amplitude and the slope of the GWB spectrum, (4) It can deal with unevenly sampled data and coloured pulsar noise spectra. We sample the likelihood function using Markov Chain Monte Carlo (MCMC) simulations. We extensively test our approach on mock PTA datasets, and find that the Bayesian inference method has significant benefits over currently proposed counterparts. We show the importance of characterising all red noise components in pulsar timing noise by demonstrating that the presence of a red component would significantly hinder a detection of the GWB

Lastly, we explore the dependence of the signal-to-noise ratio on the duration of the experiment, number of monitored pulsars, and the magnitude of the pulsar timing noise. These parameter studies will help formulate observing strategies for the PTA experiments.

2.1 Introduction

At the time of this writing several large projects are being pursued in order to directly detect astrophysical gravitational waves. This chapter concerns a program to detect gravitational waves using pulsars as nearly-perfect Einstein clocks. The practical idea is to time a set of millisecond pulsars (called the "Pulsar Timing Array", or PTA) over a number of years (Foster & Backer, 1990). Some of the millisecond pulsars create pulse trains of exceptional regularity. By perturbing the space-time between a pulsar and the Earth, the gravitational waves (GWs) will cause extra deviations from the periodicity in the pulse arrival times (Estabrook & Wahlquist, 1975; Sazhin, 1978; Detweiler, 1979). Thus from the measurements of these deviations (called "timing residuals", or TR), one may measure the gravitational waves. Currently, several PTA project are operating around the globe. Firstly, at the Arecibo Radio Telescope in North-America several millisecond pulsars have been timed for a number of years. These observations have already been used to place interesting upper limits on the intensity of gravitational waves which are passing through the Galaxy (Kaspi et al., 1994; Lommen, 2001). Together with the Green Bank Telescope, the Arecibo Radio Telescope will be used as an instrument of NANOGrav, the North American PTA. Secondly, the European PTA is being set up as an international collaboration between Great Britain, France, Netherlands, Germany, and Italy, and will use 5 European radio telescopes to monitor about 20 millisecond pulsars (Stappers et al., 2006). Finally, the Parkes PTA in Australia has been using the Parkes multi-beam radio-telescope to monitor 20 millisecond pulsars (Manchester, 2006). Some of the Parkes and Arecibo data have also been used to place the most stringent limits on the GWB to date (Jenet et al., 2006).

One of the main astrophysical targets of the PTAs is the stochastic background of the gravitational waves (GWB). This GWB is thought to be generated by a large number of black-hole binaries which are thought to be located at the centres of galaxies (Begelman et al., 1980; Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003; Sesana et al., 2008), by relic gravitational waves (Grishchuk, 2005), or, more speculatively, by cusps in the cosmic-string loops (Damour & Vilenkin, 2005). This chapter develops an inference method for the optimal PTA measurement of such a GWB.

The main difficulty of such a measurement is that not only Gravitational Waves create the pulsar timing residuals. Irregularities of the pulsar-beam rotation (called the "timing noise"), the receiver noise, the imprecision of local clocks, the polarisation calibration of the telescope (Britton, 2000), and the variation in the refractive index of the interstellar medium all contribute significantly to the timing residuals, making it a challenge to separate these noise sources from the gravitational-wave

signal. However, the GWB is expected to induce correlations between the timing residuals of different pulsars. These correlations are of a specific functional form [given by Equation (2.9) below], which is different from those introduced by other noise sources (Hellings & Downs, 1983). Jenet et al. (2005, hereafter J05) have invented a clever method which uses the uniqueness of the GWB-induced correlations to separate the GWB from other noise sources, and thus to measure the magnitude of the GWB. Their idea was to measure the timing residual correlations for all pairs of the PTA pulsars, and check how these correlations depend on the sky-angles between the pulsar pairs. J05 have derived a statistic which is sensitive to the functional form of the GWB-induced correlation; by measuring the value of this statistic one can infer the strength of the GWB. While J05 method appears robust, we believe that in its current form it does have some drawbacks, in particular:

(1) The statistic used by J05 is non-linear and non-quadratic in the pulsar-timing residuals, which makes its statistical properties non-transparent.

(2) The pulsar pairs with the high and low intrinsic timing noise make equal contributions to the J05 statistic, which is clearly not optimal.

(3) The J05 statistic assumes that the timing residuals of all the PTA pulsars are measured during each observing run, which is generally not the case.

(4) The J05 signal-to-noise analysis relies on the prior knowledge of the intrinsic timing noise, and there is no clean way to separate this timing noise from the GWB.

(5) The prior spectral information on GWB is used for whitening the signal; however, there is no proof that this is an optimal procedure. The spectral slope of the GWB is not measured.

In this chapter we develop an inference method which addresses all of the problems outlined above. Our method is based on essentially the same idea as that of J05: we use the unique character of the GWB-induced correlations to measure the intensity of the GWB. The method we develop below is Bayesian, and by construction uses optimally all of the available information. Moreover, it deals correctly and efficiently with all systematic contributions to the timing residuals which have a known functional form, i.e. the quadratic pulsar spindowns, annual variations, one-time discontinuities (jumps) due to equipment change, etc. Many parameters of the timing model (the model popular pulsar timing packages use to generate TRs from pulsar arrival times) fall in this category.

The plan of this chapter is as follows. In the next section we review the theory of the GWB-generated timing residuals and introduce our model for other contributions to the timing residuals. In Secion 2.3 we explain the principle of Bayesian analysis for GWB-measurement with a PTA, and we evaluate the Bayesian likelihood function. There we also show how to analytically marginalise over the contributions of known functional form but unknown amplitude (i.e., annual variations, quadratic residuals due to pulsar spindown, etc.). The details of this calculation are

laid out in Appendix A of this chapter. Section 2.4 discusses the numerical integration technique which we use in our likelihood analysis: the Markov Chain Monte Carlo (MCMC). In Section 2.5 we show the analyses of mock PTA datasets. For each mock dataset, we construct the probability distribution for the intensity of the GWB, and demonstrate its consistency with the input mock data parameters. We study the sensitivity of our inference method for different PTA configurations, and investigate the dependence of the signal-to-noise ratio on the duration of the experiment, on redness and magnitude of the pulsar timing noise, and on the number of clocked pulsars. In Section 2.6 we summarise our results.

2.2 The Theory of GW-generated timing residuals

2.2.1 Timing residual correlation

The measured millisecond-pulsar timing residuals contain contributions from several stochastic and deterministic processes. The latter include the gradual deceleration of the pulsar spin, resulting in a pulsar rotational period derivative which induces timing residuals varying quadratically with time (hereafter referred to as "quadratic spindown"), the annual variations due to the imperfect knowledge of the pulsar positions on the sky, the ephemeris variations caused by the known planets in the solar system, and the jumps due to equipment change (Manchester 2006). The stochastic component of the timing residuals will be caused by the receiver noise, clock noise, intrinsic timing noise, the refractive index fluctuations in the interstellar medium, and, most importantly for us, the GWB. For the purposes of this chapter we restrict ourselves to considering the quadratic spindowns, intrinsic timing noise, and the GWB; other components can be similarly included, but we omit them for mathematical simplicity. In this case, the ith timing residual of the ath pulsar can be written as

$$\delta t_{ai} = \delta t_{ai}^{\text{GW}} + \delta t_{ai}^{\text{PN}} + Q(t_{ai}), \qquad (2.1)$$

where $\delta t_{ai}^{\text{GW}}$ and $\delta t_{ai}^{\text{PN}}$ are caused by the GWB and the pulsar timing noise, respectively, and

$$Q_a(t_{ai}) = A_{a1} + A_{a2}t_{ai} + A_{a3}t_{ai}^2 \qquad (2.2)$$

represent the quadratic spindown. One expects the timing noise from different pulsars to be uncorrelated, while the GWB will cause correlations in the timing residuals between different pulsars. Therefore, the information about GWB can be extracted by correlating the timing residual data between the different pulsars (J05). If one assumes that both GWB-generated residuals and the intrinsic timing

noise are stochastic Gaussian processes, then we can represent them by the $(n \times n)$ coherence matrices:

$$
\begin{aligned}
\langle \delta t_{ai}^{\mathrm{GW}} \delta t_{bj}^{\mathrm{GW}} \rangle &= C_{(ai)(bj)}^{\mathrm{GW}} \\
\langle \delta t_{ai}^{\mathrm{PN}} \delta t_{bj}^{\mathrm{PN}} \rangle &= C_{(ai)(bj)}^{\mathrm{PN}},
\end{aligned}
\tag{2.3}
$$

with the total coherence matrix given by

$$
C_{(ai)(bj)} = C_{(ai)(bj)}^{\mathrm{GW}} + C_{(ai)(bj)}^{PN}.
\tag{2.4}
$$

The timing residuals are then distributed as a multidimensional Gaussian:

$$
\begin{aligned}
P\left(\vec{\delta t}\right) &= \frac{1}{\sqrt{(2\pi)^n \det C}} \exp \left[-\frac{1}{2} \sum_{(ai)(bj)} (\vec{\delta t}_{(ai)} - Q_a(t_{ai})) \right. \\
&\left. C_{(ai)(bj)}^{-1} (\vec{\delta t}_{(bj)} - Q_b(t_{bj})) \right],
\end{aligned}
\tag{2.5}
$$

where P denotes the probability distribution of the timing residuals. To be able to use Equation (2.5) we must

(1) be able to evaluate the GWB-induced coherence matrix from the theory, as a function of variables that parametrise the GWB spectrum, and

(2) introduce well-motivated parametrisation of the pulsar timing noise. In this chapter, the spectral density of the stochastic GW background is taken to be a power law (Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003; Maggiore, 2000)

$$
S_h = A^2 \left(\frac{f}{\mathrm{yr}^{-1}} \right)^{-\gamma},
\tag{2.6}
$$

where S_h represents the spectral density, A is the GW amplitude, f is the GW frequency, and γ is an exponent characterising the GWB spectrum. If the GWB is dominated by the supermassive black hole binaries, then $\gamma = 7/3$ (Phinney 2001). This definition is equivalent to the use of the characteristic strain as defined in Jenet et al. (2006):

$$
h_c = A \left(\frac{f}{\mathrm{yr}^{-1}} \right)^{\alpha},
\tag{2.7}
$$

with $\gamma = 1 - 2\alpha$. The GWB-induced coherence matrix is then given by

$$
\begin{aligned}
C^{\text{GW}}_{(ai)(bj)} \quad = \quad & \frac{A^2 \alpha_{ab}}{(2\pi)^2 f_L^{1+\gamma}} \left\{ \Gamma(-1 - \gamma) \sin\left(\frac{-\pi\gamma}{2} \right)(f_L \tau)^{\gamma+1} \right. \\
& \left. - \sum_{n=0}^{\infty} (-1)^n \frac{(f_L \tau)^{2n}}{(2n)!\,(2n - 1 - \gamma)} \right\}.
\end{aligned}
\tag{2.8}
$$

Here α_{ab} is the geometric factor given by

$$
\alpha_{ab} = \frac{3}{2} \frac{1 - \cos\theta_{ab}}{2} \ln\left(\frac{1 - \cos\theta_{ab}}{2} \right) - \frac{1}{4} \frac{1 - \cos\theta_{ab}}{2} + \frac{1}{2} + \frac{1}{2}\delta_{ab},
\tag{2.9}
$$

where θ_{ab} is the angle between pulsar a and pulsar b (Hellings & Downs, 1983), $\tau = 2\pi\left(t_{ai} - t_{bj}\right)$, Γ is the gamma function, and f_L is the low cut-off frequency, chosen so that $1/f_L$ is much greater than the duration of the PTA operation. Introducing f_L is a mathematical necessity, since otherwise the GWB-induced correlation function would diverge. However, we show below that the low-frequency part of the GWB is indistinguishable from an extra spindown of all pulsars which we already correct for, and that our results do not depend on the choice of f_L provided that $f_L \tau \ll 1$.

The pulsar timing noise is assumed to be Gaussian, with a certain functional form of the power spectrum. The true profile of the millisecond pulsar timing noise spectrum is not well-known at present time. The timing residuals of the most precisely observed pulsars indicate that pulsar timing noise has a white and poorly-constrained red component (J. Verbiest and G. Hobbs, private communications).

For the purposes of this chapter we will always choose the spectra to be of the same functional form for all pulsars, but this is not an inherent limitation of the inference method. We consider 3 cases of pulsar timing noise spectra:
(1) White (flat) spectra
(2) Lorentzian spectra
(3) Power-law spectra
Obviously, one could also consider a timing noise which is a superposition of these components; we do not do this at this exploratory stage. If we choose the pulsar timing noise spectrum to be white, with an amplitude N_a, the resulting correlation matrix becomes:

$$
C^{\text{PN-white}}_{(ai)(bj)} \quad = \quad N_a^2 \delta_{ab} \delta_{ij}.
\tag{2.10}
$$

The Lorentzian spectrum is a red spectrum with a typical frequency that deter-

mines the redness of the timing noise:

$$S_a(f) = \frac{N_a^2}{f_0\left(1 + \left(\frac{f}{f_0}\right)^2\right)}, \tag{2.11}$$

which yields the following correlation matrix:

$$C_{(ai)(bj)}^{\text{PN-lor}} = N_a^2 \delta_{ab} \exp\left(-f_0 \tau\right), \tag{2.12}$$

where f_0 is a typical frequency and N_a is the amplitude.

By using a power law spectral density with amplitude N_a and spectral index γ_a, one gets a timing-noise coherence matrix analogous to the one in Equation (2.8):

$$
\begin{aligned}
C_{(ai)(bj)}^{\text{PN-pl}} = \; & \frac{N_a^2 \delta_{ab}}{f_L^{\gamma_a - 1}} \left\{ \Gamma(1 - \gamma_a) \sin\left(\frac{\pi \gamma_a}{2}\right) (f_L \tau)^{\gamma_a - 1} \right. \\
& \left. - \sum_{n=0}^{\infty} (-1)^n \frac{(f_L \tau)^{2n}}{(2n)! \, (2n + 1 - \gamma_a)} \right\}.
\end{aligned}
\tag{2.13}
$$

2.3 Bayesian approach

2.3.1 Basic ideas

The method described in this report is based upon a Bayesian approach to the parameter inference. The general idea of the method is to (a) assume that the physical processes which produce the timing residuals can be characterised by several parameters, and (b) use the Bayes theorem to derive from the measured data the probability distribution of the parameters of our interest. In our case, we assume that the timing residuals are created by

(1) the GWB; we parametrise it by its amplitude A and slope γ, as in Equation (2.6).

(2) the intrinsic timing noise of the 20 monitored millisecond pulsars. We assume that the timing noise of each of the pulsars is the random Gaussian noise, with a variety of possible spectra described in the previous section. We shall refer to the variables parametrizing the timing noise spectral shape as TN_a.

(3). The quadratic spindowns, parametrised for each of the pulsars by A_{a1}, A_{a2}, and A_{a3}, cf. Eq. (2.2).

With these assumptions, we shall write down below the expression for the probability distribution $P(\text{data}|\text{parameters})$ of the data, as a function of the parameters.

By Bayes theorem, we can then compute the posterior distribution function; the probability distribution of the parameters given a certain dataset:

$$P(\text{parameters}|\text{data}) = P(\text{data}|\text{parameters}) \times \qquad \boxed{2.14}$$
$$\times \frac{P_0(\text{parameters})}{P(\text{data})}.$$

Here $P_0(\text{parameters})$ is the prior probability of the unknown parameters, which represents all our current knowledge about these parameters, and $P(\text{data})$ is the marginal likelihood, which we will use here as a normalisation factor to ensure that $P(A, \gamma, TN_a, A_{a1}, A_{a2}, A_{a3}|\text{data})$ integrates to unity over the parameter space. We note here that the marginal likelihood is in essence a goodness of fit measure that can be used for model selection. However, we will ignore the marginal likelihood in this chapter and postpone the model selection part of the data analysis to future work. For our purposes, we are only interested in A and γ, which means that we have to integrate $P(A, \gamma, TN_a, A_{a1}, A_{a2}, A_{a3}|\text{data})$ over all of the other parameters. Luckily, as we show below, for a uniform prior the integration over A_{a1}, A_{a2}, and A_{a3} can be performed analytically. This amounts to the *removal* of the quadratic spindown component to the pulsar data. We emphasise that this removal technique is quite general, and can be readily applied to unwanted signal of any known functional form (i.e., annual modulations, jumps, etc.—see Section 2.3.2), even if those parameters have already been fit for while calculating the timing residuals. The integration over TN_a must be performed numerically.

In this chapter we shall use MCMC simulation as a multi dimensional integration technique. Besides flat priors for most of the parameters, we will use slightly peaked priors for parameters which have non-normalisable likelihood functions. This ensures that the Markov Chain can converge.

In the rest of the chapter, we detail the implementation and tests of our inference method.

2.3.2 Removal of quadratic spindown and other systematic signals of known functional form

While this subsection is written with the PTA in mind, it may well be useful for other applications in pulsar astronomy. We thus begin with a fairly general discussion, and then make it more specific for the PTA case.

Consider a random Gaussian process δx_i^{G} with a coherence matrix $C(\sigma)$, which is contaminated by several systematic signals with known functional forms $f_p(t_i)$ but a-priori unknown amplitudes ξ_p. Here σ is a set of interesting parameters which

we want to determine from the data δx. The resulting signal is given by

$$\delta x_i = \delta x_i^{\text{G}} + \sum_p \xi_p f_p(t_i), \qquad (2.15)$$

or, in the vector form, by

$$\vec{\delta x} = \vec{\delta x}^{\text{G}} + F\vec{\xi}. \qquad (2.16)$$

Here the components of the vectors $\vec{\delta x}$, $\vec{\delta x}^{\text{G}}$, and $\vec{\xi}$ are given by δx_i, δx_i^{G}, and ξ_p, respectively, and F is the non-square matrix with the elements $F_{ip} = f_p(t_i)$. Note that the dimensions of $\vec{\delta x}$ and $\vec{\xi}$ are different. The Bayesian probability distribution for the parameters is given by

$$\begin{aligned} P(\sigma, \vec{\xi}|\vec{\delta x}) &= \frac{M}{\sqrt{\det C}} \exp\left[-\frac{1}{2}(\vec{\delta x} - F\vec{\xi})C^{-1}(\vec{\delta x} - F\vec{\xi})\right] \\ &\times P_0(\sigma, \vec{\xi}), \end{aligned} \qquad (2.17)$$

where P_0 is the prior probability and M is the normalisation. Since we are only interested in σ, we can integrate $P(\sigma, \vec{\xi}|\vec{\delta x})$ over the variables $\vec{\xi}$. This process is referred to as marginalisation; it can be done analytically if we assume a flat prior for $\vec{\xi}$ [i.e., if $P_0(\sigma, \vec{\xi})$ is $\vec{\xi}$-independent], since ξ_p enter at most quadratically into the exponential above. After some straightforward mathematics which we have detailed in Appendix A of this chapter, we get

$$\begin{aligned} P(\sigma|\vec{\delta x}) &= \frac{M'}{\sqrt{\det(C)\det(F^{\text{T}}C^{-1}F)}} \\ &\times \exp\left[-\frac{1}{2}\vec{\delta x} \cdot C'\vec{\delta x}\right], \end{aligned} \qquad (2.18)$$

where M' is the normalisation, and

$$C' = C^{-1} - C^{-1}F(F^{\text{T}}C^{-1}F)^{-1}F^{\text{T}}C^{-1}, \qquad (2.19)$$

and the T-superscript stands for the transposed matrix. Equation (2.18) is one of the main equations of this work, since it provides a statistically rigorous way to remove (i.e., marginalise over) the unwanted systematic signals from random Gaussian processes. One can check directly that the above expression for $P(\sigma|\vec{\delta x})$ is insensitive to the values ξ_p of the amplitudes of the systematic signals in the Eq. (2.15).

We now apply this formalism to account for the quadratic spindowns in the PTA. As in Section 2.2, it will be convenient to use the 2-index notation for the

timing residuals, δt_{ai} measured at the time t_{ai}, where a is the pulsar index, and i is the number of the timing residual measurement for pulsar a. The space of the spindown parameters A_{aj}, $j = 1, 2, 3$ has $3N$ dimensions, where N is the number of pulsars in the array. In the component language, we write

$$\delta t_{ai} = \delta t_{ai}^{G} + \sum_{b,j} F_{(ai)(bj)} A_{bj}, \qquad (2.20)$$

where

$$F_{(ai)(bj)} = \delta_{ab} t_{ai}^{j-1}, \qquad (2.21)$$

δt^{G} is the part of the timing residual due to a random Gaussian process (i.e., GWB, timing noise, etc.), and $j = 1, 2, 3$. The quantities $F_{(ai)(bj)}$ are components of the matrix operator which acts on the $3N$-dimensional vector in the parameter space and produces a vector in the timing residual space. For example, for 20 pulsars, each with 250 timing residual observations, the matrix $F_{(ai)(bj)}$ has $20 \times 250 = 5000$ rows, each marked by 2 indices $a = 1, ..., 20$, $i = 1, ..., 250$, and $20 \times 3 = 60$ columns, each marked by 2 indices $b = 1, ..., 20$, $j = 1, 2, 3$. Thus in the vector form, one can write Eq. (2.20) as

$$\vec{\delta t} = \vec{\delta t}^{G} + F \vec{A}, \qquad (2.22)$$

which is identical to the Eq. (2.16). We thus can use Eq. (2.18) to remove the quadratic spindown contribution from the PTA data.

Although we only demonstrate this technique for quadratic spindown, this removal technique will be useful for treating other noise sources in the PTA. All sources of which the functional form is known (and therefore can be fit for, as most popular pulsar timing packages do) can be dealt with, i.e.

(1) Annual variation of the timing residuals due to the imprecise knowledge of the pulsar position on the sky. The annual variation in each of the pulsars will be a predictable function of the associated 2 small angular errors (latitude and longitude). Thus our parameter space will expand by 2N, but this will still keep the F matrix manageable.

(2) Changes of equipment will introduce extra jumps, and must be taken into account. This is trivial to deal with using the techniques described above.

(3) Some of the millisecond pulsars are in binaries, and their orbital motion must be subtracted. The errors one makes in these subtractions will affect the timing residuals. They can be parametrised and dealt with using the techniques of this section (we thank Jason Hessels for pointing this out).

2.3.3 Low-frequency cut-off

All predictions for GWB spectrum show a steep power law $\propto f^{-\gamma}$, where for black-hole binaries $\gamma = 7/3$ (Phinney, 2001). Physically, there is a low-frequency cut-off to the spectrum, due to the fact that black-hole binaries with periods greater than 1000 years shrink mostly due to the external friction (i.e., scattering of circum-binary stars or excitation of density waves in a circum-binary gas disc), and not to gravitational radiation. However, while the exact value of the low-frequency cut-off is poorly constrained, the PTA should not be sensitive to it since the duration of the currently planned experiments is much shorter than 1000 years. In this subsection, we show this formally by explicitly introducing the low-frequency cut-off and by demonstrating that our Bayesian probabilities are insensitive to its value.

Consider the expression in Equation (2.8) for the GWB-generated correlation matrix for the timing residuals. This expression contains an integral of the form

$$I = \int_{f_L}^{\infty} \cos(f\tau) f^{-(\gamma+2)} df, \qquad (2.23)$$

where $\tau = 2\pi(t_i - t_j)$. When the low-frequency cut-off is much smaller than the inverse of the experiment duration, i.e. when $f_l \tau \ll 1$, the integral above can be expanded as

$$I = B\tau^{\gamma+1} + \frac{1}{f_L^{\gamma+1}} \left\{ \frac{1}{(\gamma+1)} - \frac{(f_L\tau)^2}{2(\gamma-1)} + O\left[(f_L\tau)^4\right] \right\}, \qquad (2.24)$$

where

$$B = \Gamma(-1-\gamma) \sin\left(\frac{-\pi\gamma}{2}\right) \tau^{\gamma+1}. \qquad (2.25)$$

In the expansion above we have assumed $1 < \gamma < 3$. The terms which contain f_L diverge when f_L goes to zero, and scale as τ^0 or τ^2 with respect to the time interval. We now show that these divergent terms get absorbed in the process of elimination of the quadratic spindowns.

Suppose that we add to the timing residuals of a pulsar a quadratic spindown term, $A_1 + A_2 t + A_3 t^2$. The spindown-removal procedure described in the previous section makes our results completely insensitive to this addition: A's could be arbitrarily large but the measured GWB would still be the same. Clearly, the same is true if one treats A_1, A_2, A_3 not as fixed numbers, but as random variables drawn from some Gaussian distribution. The correlation introduced into the timing

residuals by adding a *random* quadratic spindown is given by

$$
\begin{aligned}
\langle \delta t_i \delta t_j \rangle &= \langle A_1^2 \rangle + \langle A_1 A_2 \rangle (t_i + t_j) \\
&+ 2\langle A_2^2 \rangle t_i t_j + \langle A_1 A_3 \rangle (t_i^2 + t_j^2) \\
&+ \langle A_2 A_3 \rangle t_i t_j (t_i + t_j) + \langle A_3^2 \rangle t_i^2 t_j^2 .
\end{aligned}
\tag{2.26}
$$

The f_L-dependent part of Eq. (2.24) contains terms which scale as $t_i^2 + t_j^2$, $t_i t_j$, and *const*, and thus have the same functional t_i, t_j dependence as some of the terms in Eq. (2.26). Since the terms in Eq. (2.26) can be made arbitrarily large, it is clear that the terms corresponding to the low-frequency cutoff could be absorbed into the correlation function corresponding to the quadratic spindown with the stochastic coefficients. We have made this argument for the timing residuals from a single pulsar, but it is trivial to extend it to the case of multiple pulsars. Thus our results are not sensitive to the actual choice of the f_L so long as $f_l \tau \ll 1$; this is confirmed by direct numerical tests.

2.4 Numerical integration techniques

2.4.1 Metropolis Monte Carlo

The Bayesian probability distribution for the PTA is computed in multi dimensional parameter space, where all of the parameters except two characterise intrinsic pulsar timing noise and other potential interferences. To obtain meaningful information about the GWB, we need to integrate the probability function over all of the unwanted parameters. This is a challenging numerical task: a direct numerical integration over more than several parameters is prohibitively computationally expensive. Fortunately, numerical shortcuts do exist, and the most common among them is the Markov Chain Monte Carlo (MCMC) simulation. In a typical MCMC, a set of semi-random walkers sample the parameter space in a clever way, each generating a large number of sequential locations called a *chain* (Newman & Barkema, 1999). After a sufficient number of steps, the density of points of the chain becomes proportional to the Bayesian probability distribution. We found that for our purposes, the number of steps required for the chain convergence scales linearly with the number of dimensions of the parameter space; typically few$\times 10^4$ steps are required for reliable convergence. In this work we use the Metropolis (Newman & Barkema, 1999) computational scheme for generating the chain, which can be used with an arbitrary distribution, the proposal distribution, for generating new locations of the chain. We use a Gaussian proposal distribution, centred at the current location in the parameter space. During an initial period, the burn-in pe-

riod, the width of the proposal distribution in all dimensional directions is set to yield the asymptotically optimal acceptance rate of 23.4% for Metropolis (Roberts et al., 1997). At the end of the MCMC simulation we check the convergence of the chain using a batched mean method (Efron, 1979). We also calculate the global maximum likelihood value for all parameters using a conjugate directions search (Brent, 1973). By carefully inspecting the marginalised posterior distributions we found that the posterior distribution only has a single peak for all the parameters and parametrisations we have considered so far. We therefore assume that we are not hindered by chains getting stuck in local maxima when assessing convergence.

2.4.2 Current MCMC computational cost

The greatest computational challenge in constructing the chain is the fast evaluation of the matrix C^{-1} in Eqs. (2.18)&(2.19). If 250 timing residuals are measured for each of the pulsars (50 weeks for 5 years), the size of the matrix C becomes (5000×5000). We find it takes about 20 seconds to invert C and thus about 1.5 times as much to arrive to the next point in the chain. Therefore, for the required 10^5 chain points to get the convergent distribution, we need of order 1 month of the single-processor computational time. On a cluster this can be done in a couple of days. We emphasise that this is an order n^3 process. For matrices of (2000×2000) the calculation can be done overnight on a single modern workstation, but for $(10^4 \times 10^4)$ the calculation is already a serious challenge.

For the currently projected size of the datasets (Manchester, 2006), the amount of timing residuals will most likely not exceed the 250 (Hobbs, private communications). Thus, the brute-force method presented here is not computationally expensive for the projected data volume over the next 5 years.

2.4.3 Choosing a suitable prior distribution

For some models (e.g. the power law spectral density for pulsar timing noise) the likelihood function proves to be not normalisable. This would pose a serious problem in combination with uniform priors as the nuisance parameters then cannot be marginalised and the posterior cannot represent a probability distribution. Although this is a sign that our model is incorrect (zero evidence/normalisation), this can be easily solved by choosing a more realistic prior. The prior distribution reflects our a priori knowledge of the model parameters. Because timing noise is currently not well enough understood to properly parametrise it, the natural choice of prior would be an uninformative one. Studies by Verbiest et al. (2009, see also chapter 4 of this thesis) have suggested that the timing noise that is dominant in MSPs at current sensitivity levels is less "red" than the GWB. We therefore pro-

pose to use a proper prior distribution in the form of a Lorentzian shaped profile, much broader than the likelihood distribution in the spectral index parameter:

$$P_0(\gamma_i) = \frac{\Delta_i}{\pi\left(\Delta_i^2 + \gamma_i^2\right)},$$

(2.27)

where γ_i is the parameter for which we are construction a prior, and Δ_i is some typical width/value for this parameter.

As an example we show the likelihood function and the prior for the pulsar timing noise spectral index parameter of Equation (2.13) in Fig. 2.1. The likelihood function seems to drop to zero for high γ_i, but it actually has a non-negligible value for all γ_i greater than the maximum likelihood value. The broadness of the prior is chosen such that it does not change the representation of the significant part of the likelihood in the posterior, but it does make sure that the posterior is normalisable.

2.4.4 Generating mock data

In order to generate mock data, we produce a realisation of the multi dimensional Gaussian process of Equation (2.5), as follows. We rewrite Equation (2.5) is a basis in which C is diagonal:

$$P\left(\vec{\delta t}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \varphi\left(\frac{y_i}{\sqrt{\lambda_i}}\right),$$

(2.28)

where,

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

(2.29)

Here λ_i are the eigenvalues of C, and

$$\vec{y} = T^{-1}\vec{\delta t},$$

(2.30)

where T is the transformation matrix which diagonalises C:

$$(T^{-1}CT)_{ij} = \lambda_i\delta_{ij}.$$

(2.31)

Thus we follow the following steps:
(1) Diagonalise matrix C, find T and λ_i.
(2) Choose y_i from random Gaussian distributions of widths $\sqrt{\lambda_i}$.
(3) Compute the timing residuals via Eq. (2.30).

It is then trivial to add deterministic processes, like quadratic spindowns, to the
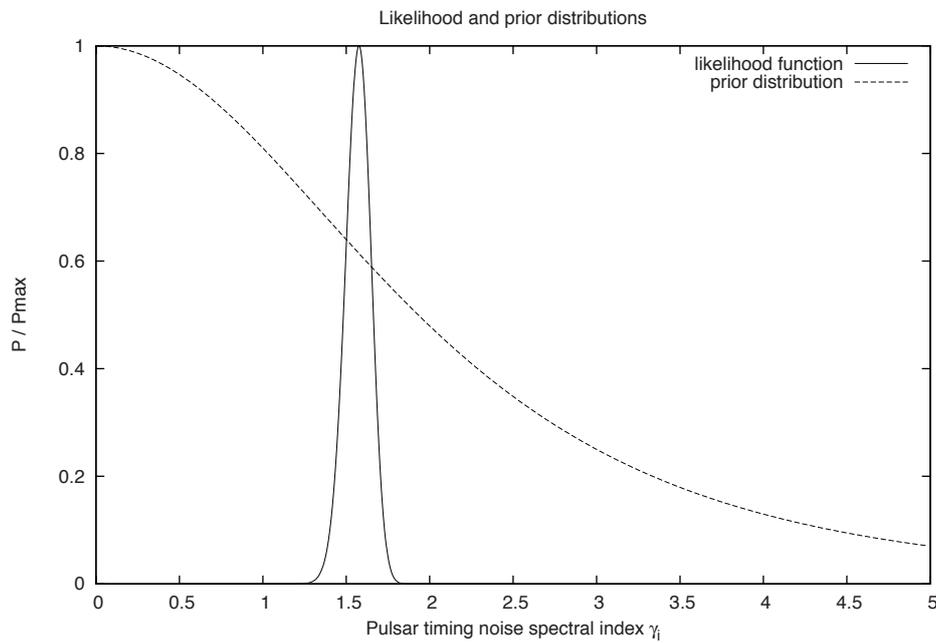
Figure 2.1: The likelihood and prior distribution for a pulsar timing noise spectral index parameters γ_i. The solid line represents the likelihood function. It is sharply peaked and it looks as if it drops to zero for high γ_i. However, for high γ_i it will have a constant non-negligible value. The dashed line represents our chosen prior distribution. The prior is normalisable, and its application makes the posterior distribution normalisable as well.

simulated timing residuals.

2.5 Tests and parameter studies

We test our inference method by generating mock timing residuals for a number of millisecond pulsars which are positioned randomly in the sky. We found it convenient to parametrise the GWB spectrum by [cf.Equation (2.6)]

$$S_h(f) = A^2 \left(\frac{f}{\mathrm{yr}^{-1}} \right)^{-\gamma}.$$

(2.32)

Our mock timing residuals are a single realisation of GWB for some values of A and γ and the pulsar timing noise. Random quadratic spindown terms are added. We then perform several separate investigations as follows:

2.5.1 Single dataset tests

Our inference method is tested on several datasets in the following way:
The mock datasets were generated with parameters resembling an experiment of 20 pulsars, with observations approximately every 5 weeks for 5 years. The pulsar timing noise was set to an optimistic level of 100 ns each (rms timing residuals). In all cases the level of GWB has been set to $A = 10^{-15}\mathrm{yr}^{1/2}$, with $\gamma = 7/3$. This level of GWB is an order of magnitude smaller than the most recent upper limits of this type(Jenet et al., 2006). We then analyse this mock data using the MCMC method. In Figs 2.2—2.4 we see examples of the joint A—γ probability distribution, obtained by these analyses. For each dataset we also calculate the maximum likelihood value of all parameters using a conjugate directions search. The Bayesian inference method gives results consistent with the input parameters (i.e., they recover the amplitude and the slope of the GWB within measurement errors). This was observed in all our tests.

For all datasets we also calculated the Fisher information matrix, a matrix consisting of second-order derivatives to all parameters, at the maximum likelihood points. We can use this matrix to approximate the posterior by a multi dimensional Gaussian, since for some particular models this approximation is quite good. The Fisher information matrix can be calculated in a fraction of the time needed to perform a full MCMC analysis. For all datasets we have plotted the 1σ contour of the multi dimensional Gaussian approximation.

As an extra test, we have also used datasets generated by the popular pulsar timing package tempo2 (Hobbs et al., 2006) with a suitable GWB simulation plug-

Joint A-γ distribution, white noise
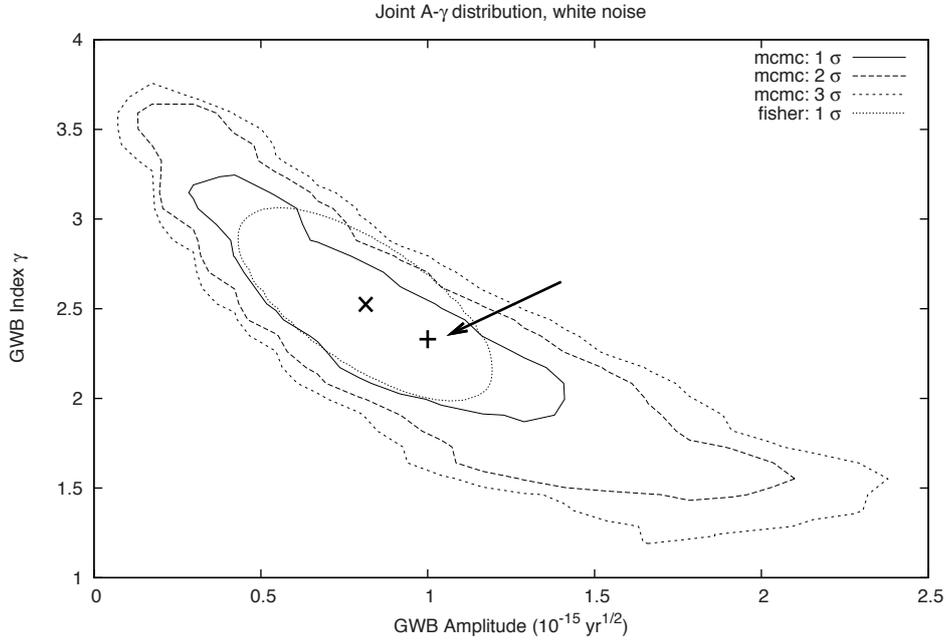
Figure 2.2: The GW likelihood function (GW amplitude, GW slope vs prob. density contours), determined with the MCMC method for a set of mock data with 20 pulsars, and 100 data points per pulsar approximately evenly distributed over 5 years. Each pulsar has a white timing noise of 100ns. The true GW amplitude and slope are shown as a "+" with an arrow, and the maximum likelihood values are shown as "x". The contours are in steps of σ, with the inner one at 1σ. The 1σ contour of the Gaussian approximation is also shown.

in (Hobbs et al., 2009). We were able to generate datasets with exactly the same parameters as with our own inference method, provided that the timing noise was white. We have confirmed that those datasets yield similar results when analysed with our inference method.

An important point is that that the spectral form of the timing noise has a large impact on the detectability of the GWB. For a red Lorentzian pulsar timing noise there is far greater degeneracy between the spectral slope and amplitude in the timing residual data for the GWB than for white pulsar timing noise, and thus the overall signal-to-noise ratio is significantly reduced by the red component of the timing noise.
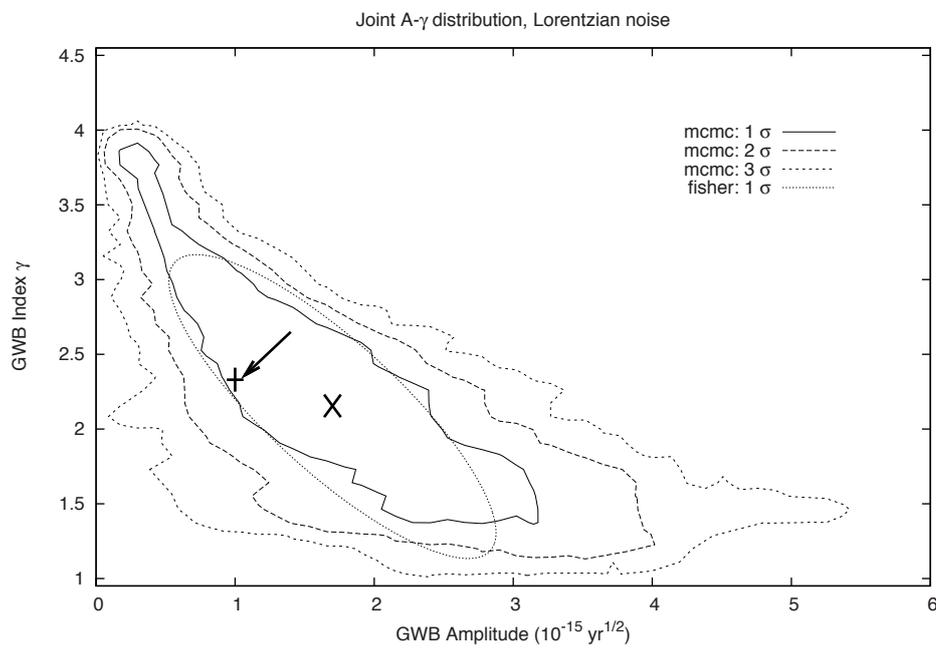
Figure 2.3: Same as in Fig. 2.2, but the mock data is generated and analysed using Lorentzian timing noise. Overall timing noise amplitude and characteristic frequency f_0 are taken to be 100ns and $1\mathrm{yr}^{-1}$ for each pulsar.
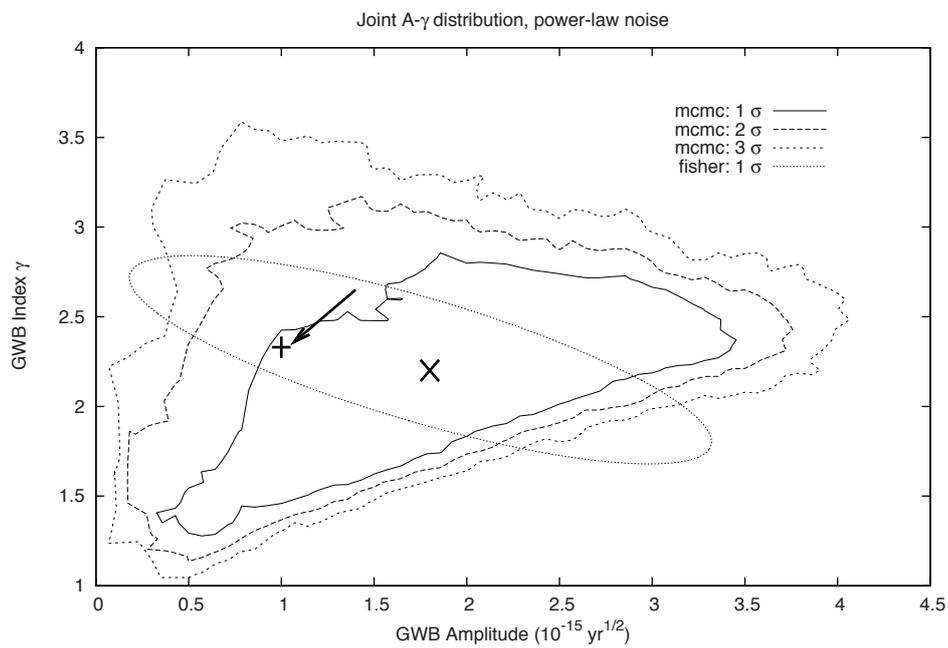
Figure 2.4: Same as in Fig. 2.2, but the mock data is generated and analysed using power law timing noise. Overall timing noise amplitude and spectral index γ_i are taken to be 100ns and 1.5 for each pulsar. For all γ_i, a prior distribution according to Equation 2.27.

2.5.2 Multiple datasets, same input parameters

To estimate the robustness of our inference method, we also perform a maximum likelihood search on many datasets:

(a) We generate a multitude of mock timing residual data for the same PTA configurations as in Section 2.5.1, with white timing noise.

(b) For every one of these datasets we calculate the maximum likelihood parameters using the conjugate directions search. The ensemble of maximum likelihood estimators for (A, γ) should be close to the true values used to generate the timing residuals.

The results of maximum likelihood search on many datasets is demonstrated in Fig. 2.5. The points are the maximum likelihood values for individual datasets. It can be seen that the points are distributed in a shape similar, but not identical, to Fig. 2.2: some points are quite far off from the input parameters. In order to test the validity of the results, we calculate the Fisher information matrix at the maximum likelihood points, and show the 1σ contour of the multidimensional Gaussian approximation based on the Fisher information matrix for three points. We wee that the error contours do not exclude the true values at high confidence, even though the Fisher matrix does not yield a perfect representation of the error contours (the true posterior is not perfectly Gaussian), and we have a posteriori selected outliers for 2 of the 3 cases.

2.5.3 Parameter studies

To test the accuracy of the Bayesian inference method, and to provide suggestions for optimal PTA configurations, we conduct some parameter studies on simplified sets of mock timing residuals:

(a) We generate many sets of mock timing residuals for the simplified case of white pulsar timing noise spectra, all with the same white noise amplitude. The datasets are timing residuals for some number of millisecond pulsars which are positioned randomly in the sky. We parametrise the GWB by Equation (2.32). We then generate many sets of timing residuals, varying several parameters [i.e., timing noise amplitude (assumed the same for all pulsars), duration of the experiment, and number of pulsars].

(b) For each of the mock datasets we approximate the likelihood function by a Gaussian in the GWB amplitude A, with all other parameters fixed to their real value. We use A as a free parameter since it represents the strength of the GWB, and therefore the accuracy of A is a measure of the detectability. All other parameters are fixed to keep the computational time low, but this does result in a higher signal
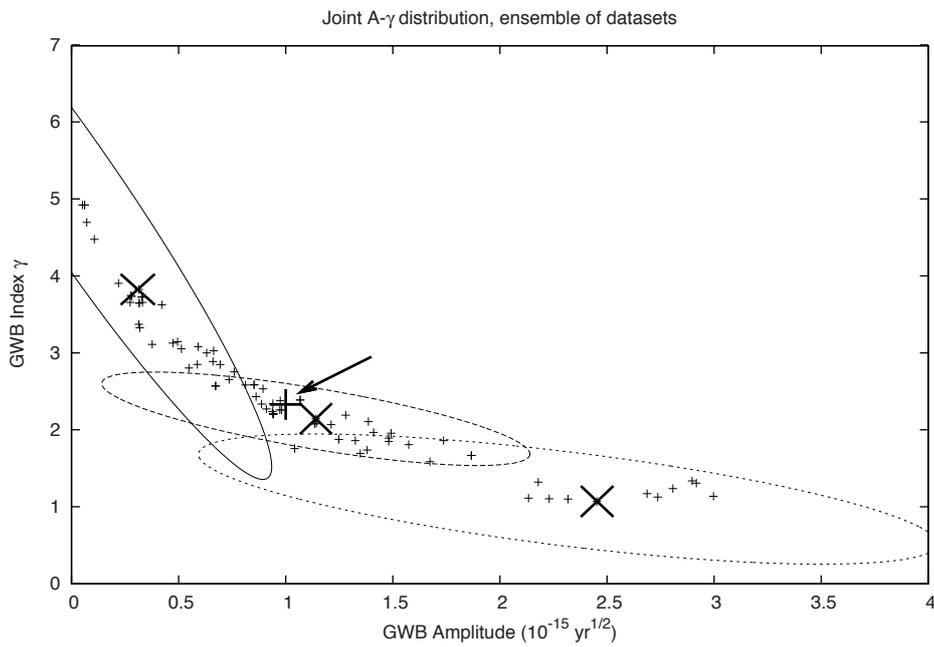
Figure 2.5: The maximum likelihood values for an ensemble of realisations of mock datasets, all with the same model parameters: 100 ns white noise, 20 pulsars, and 100 data points per pulsar approximately evenly distributed over 5 years. The contours are confidence contours based on Fisher information matrix approximations of the likelihood function.
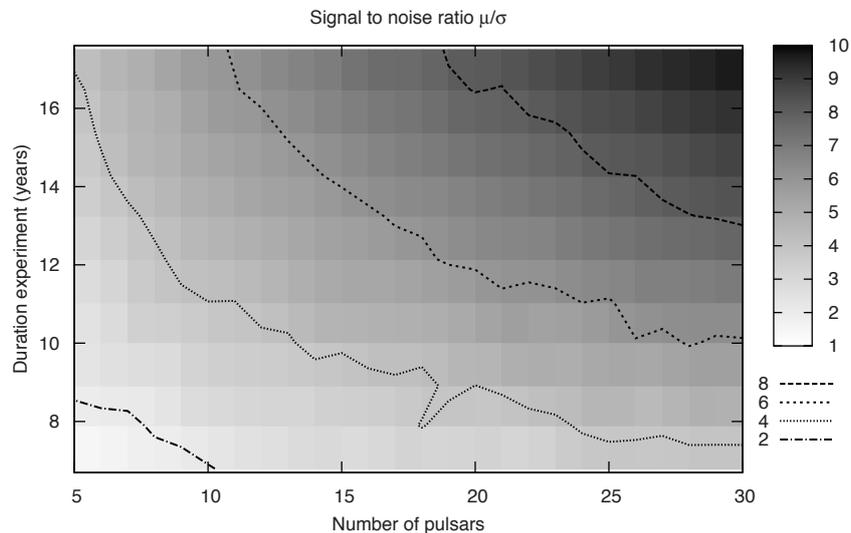
Figure 2.6: Density plot of the signal to noise ratio μ/σ for different realisations of timing residuals. We have assumed monthly observations of pulsars with white timing noise of 100 ns each. The GWB amplitude has been set to $10^{-15}\text{yr}^{1/2}$.

to noise ratio than is obtainable with a full MCMC analysis.

(c) For this Gaussian approximation, we calculate the ratio μ/σ as an estimate for the signal to noise ratio, where μ is the value of A at which the likelihood function maximises, and σ is the value of the standard deviation of the Gaussian approximation. Our results, represented as signal-to-noise contour plots for pairs of the input parameters, can be seen in Fig. 2.6—Fig. 2.12.

2.5.4 Comparison to other work

More then a decade ago, McHugh et al. (1996) used a Bayesian technique to produce upper limits on the GWB using pulsar timing. We found the presentation of this work rather difficult to follow. Nonetheless, it is clear that the analysis presented here is more general than that of McHugh et al.: we treat the whole pulsar array, and not just a single pulsar; we take into account the extreme redness of the noise and develop the formalism to treat the systematic errors like quadratic spindown.
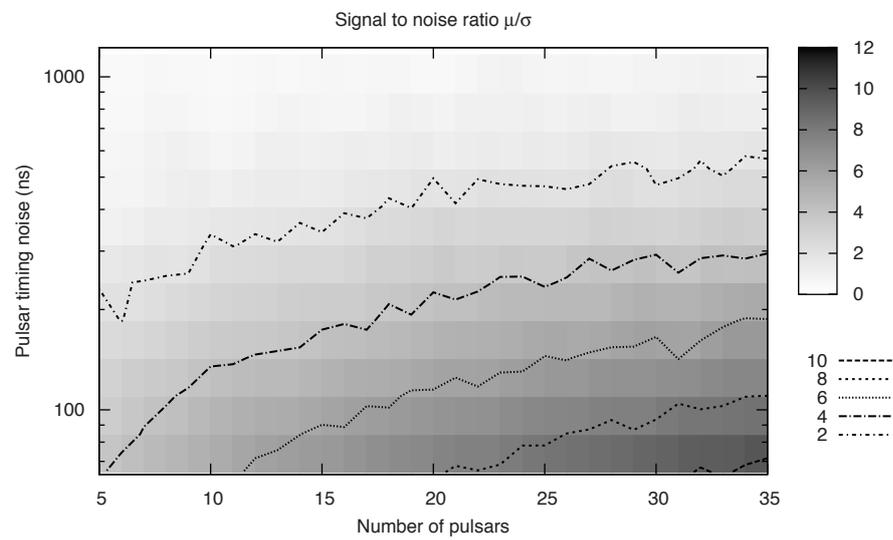
Figure 2.7: Density plot of the signal to noise ratio μ/σ for different realisations of timing residuals. We have assumed 100 data points per pulsars, approximately evenly distributed over a period of 7.5 years. The GWB amplitude has been set to $10^{-15} \mathrm{yr}^{1/2}$.
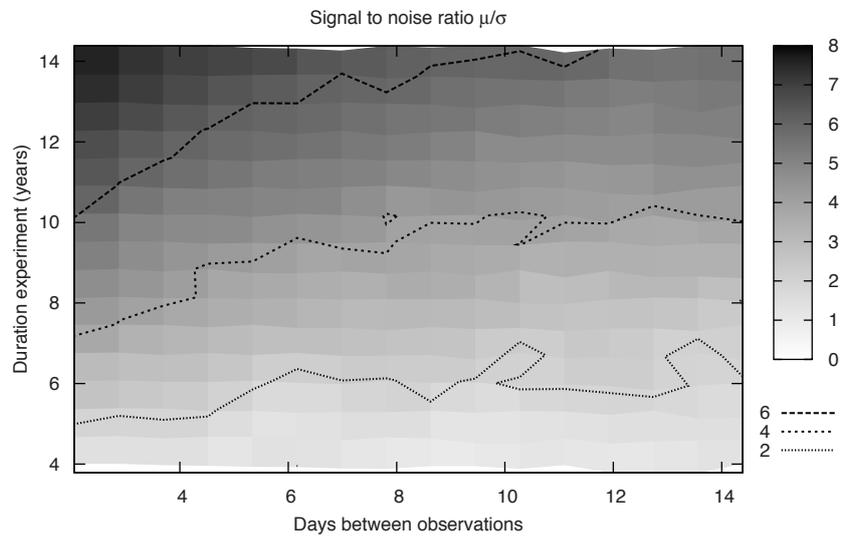
Figure 2.8: Density plot of the signal to noise ratio μ/σ for different realisations of timing residuals. We have used a constant GWB amplitude of $10^{-15}\mathrm{yr}^{1/2}$ and 20 pulsars.
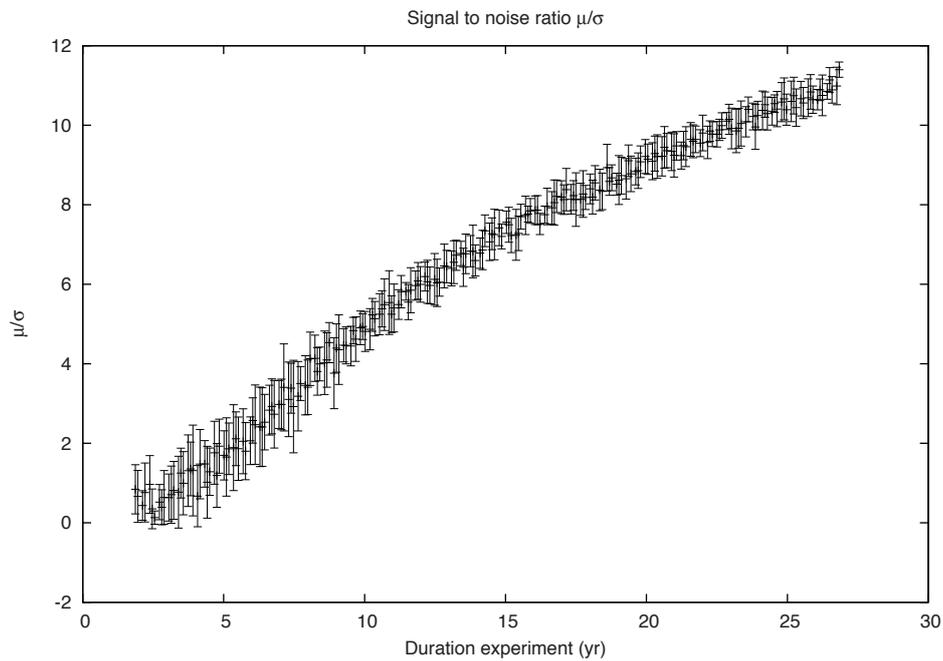
Figure 2.9: Density plot of the signal to noise ratio $\frac{\mu}{\sigma}$ for different realisations of timing residuals. We have used 20 pulsars with white pulsar timing noise levels of 100 ns each, with monthly observations. The GWB amplitude has been set to $10^{-15}\text{yr}^{1/2}$. The points and error bars are the mean and standard deviation of 10 realisations.
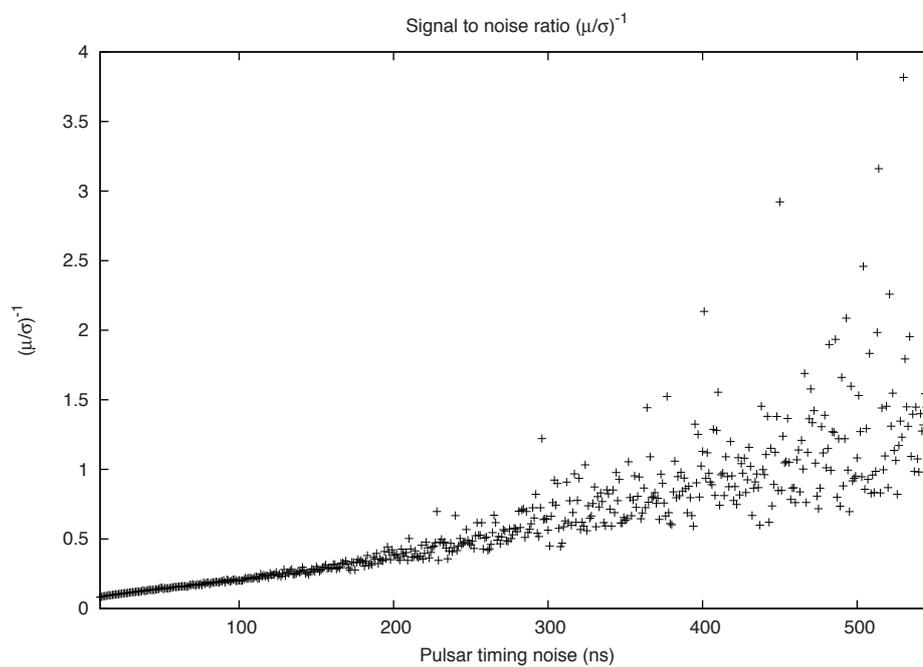
Figure 2.10: Plot of one over the signal to noise ratio $(\mu/\sigma)^{-1}$ with respect to the pulsar timing noise for an experiment of 5 years, 20 pulsars and monthly observations. The GWB amplitude has been set to $10^{-15}\text{yr}^{1/2}$.
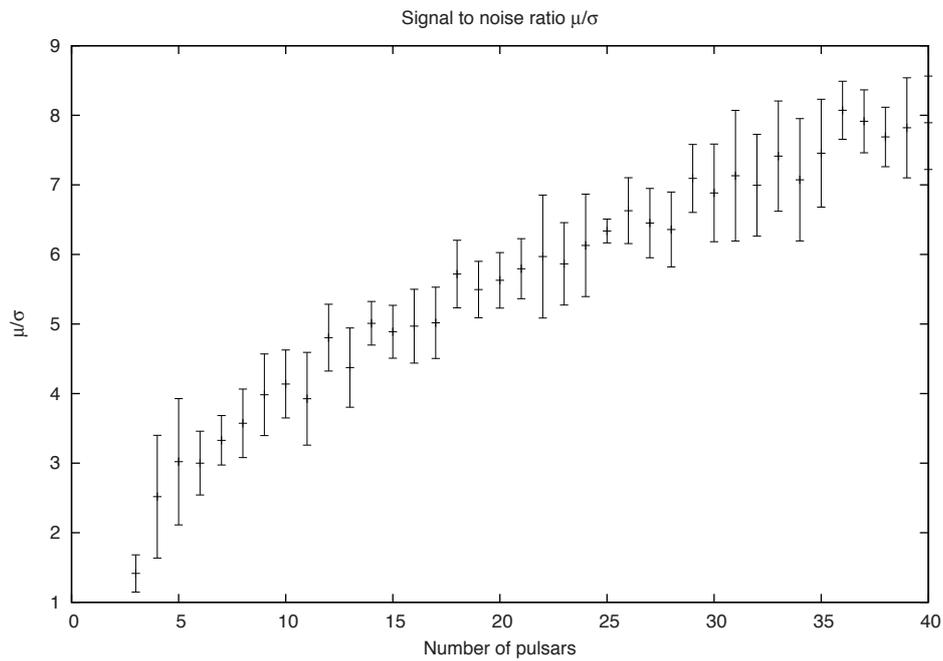
Figure 2.11: Plot of the signal to noise ratio $\frac{\mu}{\sigma}$ with respect to the number of observed pulsars. The white timing noise of each pulsar has been set to 100 ns and the observations were taking every 2 months for a period of 7.5 years. The GWB amplitude has been set to $10^{-15}\mathrm{yr}^{1/2}$. The points and error bars are the mean and standard deviation of 10 realisations.
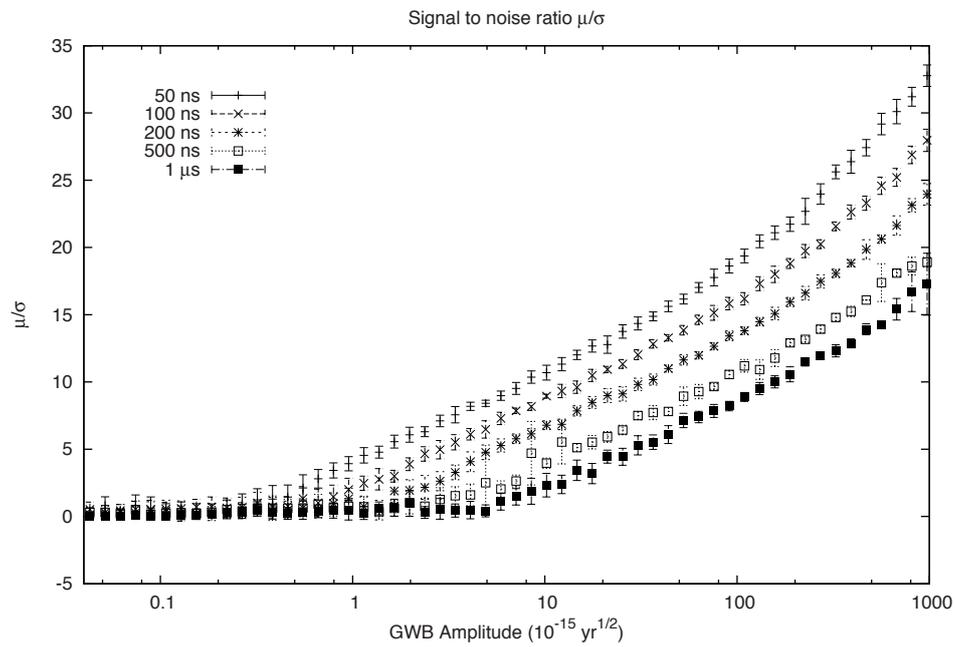
Figure 2.12: Several plots of the signal to noise ratio $\frac{\mu}{\sigma}$ with respect to the level of the GWB amplitude. The number of pulsars was set at 20, with bi-weekly observations for a period of 5 years. The pulsar noise levels were set at 50, 100, 200, 500, 1000 ns for the different plots. The points and error bars are the mean and standard deviation of 10 realisations.

Simultaneously with our work, a paper by Anholm et al. (2009) has appeared on the arxiv preprint service. Their approach was to construct a quadratic estimator (written explicitly in the frequency domain), which aims to be optimally sensitive to the GWB. This improves on the original non-quadratic estimator of J05. However, a number of issues important for the pulsar timing experiment remained unaddressed, the most important among them the extreme redness of the GWB and the need to subtract consistently the quadratic spindown.

2.6 Conclusion

In this chapter we have introduced a practical Bayesian inference method for measuring the GWB using Pulsar Timing Arrays. Several attractive features of the inference method should make it useful to the PTA community:

(1) the ability to simultaneously measure the amplitude and slope of GWB,

(2) its ability to deal with unevenly sampled datasets, and

(3) its ability to treat systematic contributions of known functional form. From the theoretical point of view, the inference method is guaranteed to extract information without loss of information, provided that our parametrisation of the timing noise is correct.

Test runs of our inference method have shown that the experiments signal-to-noise (S/N) ratio strongly decreases with the redness of the pulsar timing noise, and strongly increases with the duration of the PTA experiment. We have also charted the S/N dependence on the number of well-clocked pulsars and the level of their timing noise. These charts should be helpful in the design of the optimal strategy for future PTA observations.

Appendix A

In this Appendix we show explicitly how to perform marginalisation over the nuisance parameters $\vec{\xi}$ in Eq. (2.16), rewritten here for convenience:

$$
\begin{aligned}
P(\sigma, \vec{\xi}|\vec{\delta x}) &= \frac{M}{\sqrt{\det C}} \exp\left[-\frac{1}{2}(\vec{\delta x} - F\vec{\xi})C^{-1}(\vec{\delta x} - F\vec{\xi})\right] \\
&\times P_0(\sigma, \vec{\xi}),
\end{aligned}
\tag{2.33}
$$

From here on we assume that P_0 is independent of $\vec{\xi}$ (a flat prior). All values are therefore equally likely for all elements of $\vec{\xi}$ prior to the observations. This assumption is also implicitly made in the frequentist approach when fitting for these kinds of parameters as is done in popular pulsar timing packages. We now perform the marginalisation:

$$
P(\sigma|\vec{\delta x}) = \int P(\sigma, \vec{\xi}|\vec{\delta x}) \mathrm{d}^m\xi,
\tag{2.34}
$$

where m is the dimensionality of of $\vec{\xi}$. The idea now is to rewrite the the exponent E of Equation (2.33) in such a way that we can perform a Gaussian integral with respect to $\vec{\xi}$ (we have to get rid of the F in front of $\vec{\xi}$). Therefore, we will expand E and complete the square with respect to ξ:

$$
\begin{aligned}
E &= \left(\vec{\delta x} - F\vec{\xi}\right)^T C^{-1} \left(\vec{\delta x} - F\vec{\xi}\right) \\
&= \vec{\delta x}^T C^{-1}\vec{\delta x} - 2\vec{\xi}^T F^T C^{-1}\vec{\delta x} + \vec{\xi}^T F^T C^{-1} F\vec{\xi} \\
&= \vec{\delta x}^T C^{-1}\vec{\delta x} + \left(\vec{\xi} - \vec{\chi}\right)^T F^T C^{-1} F \left(\vec{\xi} - \vec{\chi}\right) \\
&\quad - \vec{\chi}^T F^T C^{-1} F\vec{\chi},
\end{aligned}
\tag{2.35}
$$

where we have used the substitution:

$$
\vec{\chi} = \left(F^T C^{-1} F\right)^{-1} F^T C^{-1}\vec{\delta x}.
\tag{2.36}
$$

Using this, we can write the $\vec{\xi}$ dependent part of the integral of Equation (2.34) as a multi dimensional Gaussian integral:

$$
\begin{aligned}
I &= \int \exp\left(\frac{-1}{2}\left(\vec{\xi} - \vec{\chi}\right)^T F^T C^{-1} F \left(\vec{\xi} - \vec{\chi}\right)\right) \mathrm{d}^m\xi \\
&= (2\pi)^m \det\left(F^T C^{-1} F\right)^{-1}.
\end{aligned}
\tag{2.37}
$$

From this it follows that:

$$P(\sigma|\vec{\delta x}) \quad = \quad \frac{M'}{\sqrt{\det(C)\,\det(F^{\mathrm{T}}C^{-1}F)}}$$
$$\times \exp\left[-\frac{1}{2}\vec{\delta x} \cdot C'\vec{\delta x}\right], \tag{2.38}$$

where we have absorbed all constant terms in the normalisation constant M', and where we have used:

$$C' = C^{-1} - C^{-1}F(F^{\mathrm{T}}C^{-1}F)^{-1}F^{\mathrm{T}}C^{-1}. \tag{2.39}$$
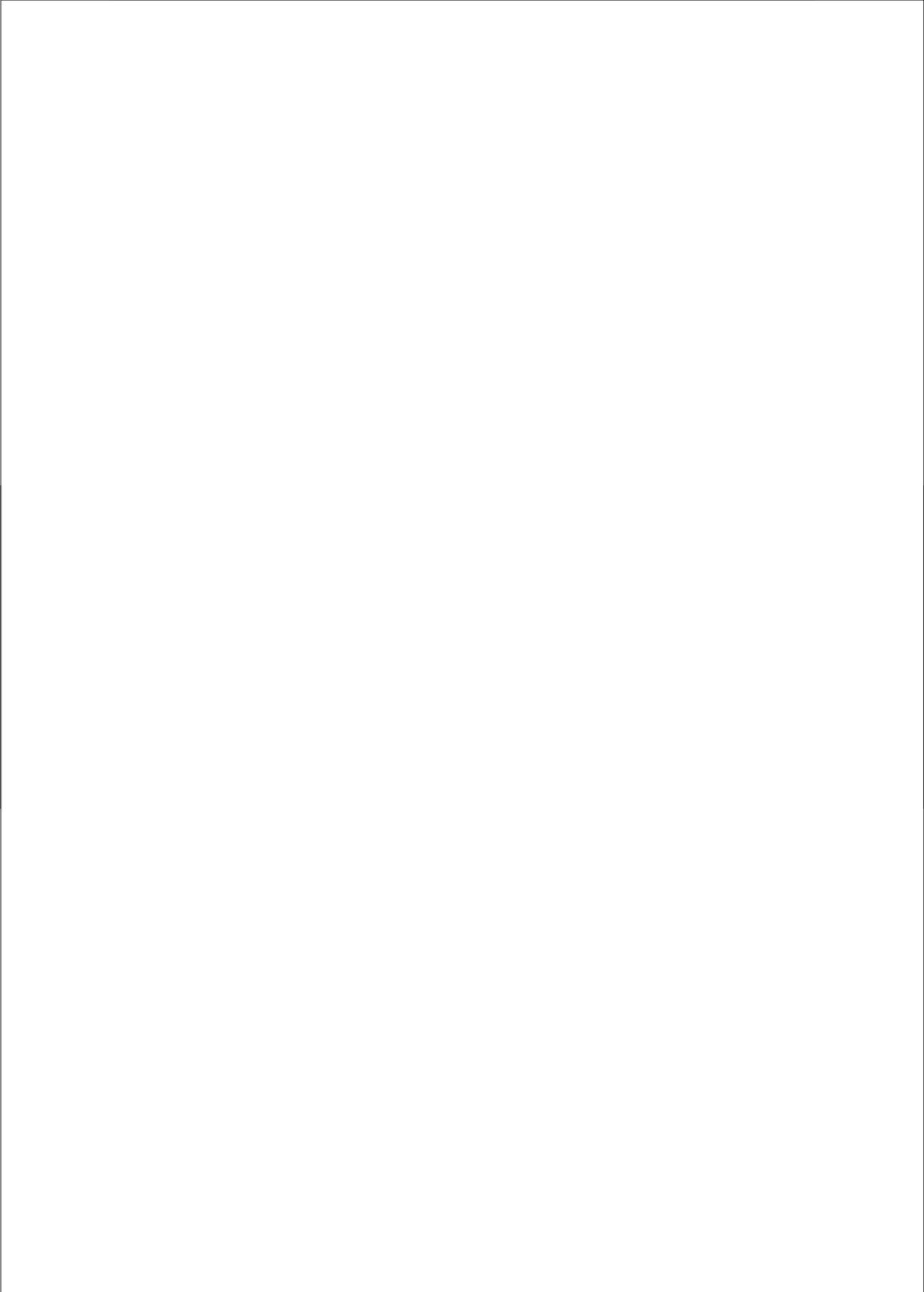
3

Gravitational-wave memory and Pulsar Timing Arrays

The world is full of obvious things which nobody by any chance ever observes.

Sherlock Holmes (Sir Arthur Conan Doyle)

Abstract

Pulsar timing arrays (PTAs) are designed to detect gravitational waves with periods from several months to several years, e.g. those produced by wide supermassive black-hole binaries in the centers of distant galaxies. Here we show that PTAs are also sensitive to mergers of supermassive black holes. While these mergers occur on a timescale too short to be resolvable by a PTA, they generate a change of metric due to non-linear gravitational-wave memory which persists for the duration of the experiment and could be detected. We develop the theory of the single-source detection by PTAs, and derive the sensitivity of PTAs to the gravitational-wave memory jumps. We show that mergers of $10^8 M_\odot$ black holes are $2-\sigma$-detectable (in a direction, polarisation, and time-dependent way) out to co-moving distances of ~ 1 billion light years. Modern prediction for black-hole merger rates imply marginal to modest chance of an individual jump detection by currently developed PTAs. The sensitivity is expected to be somewhat higher for futuristic PTA experiments with the Square Kilometre Array (SKA).

3.1 Introduction

Bursts of gravitational waves leave a permanent imprint on spacetime by causing
a small permanent change of the metric, as computed in the transverse traceless
gauge (Payne, 1983; Christodoulou, 1991; Blanchet & Damour, 1992; Thorne,
1992). These gravitational-wave "memory jumps" are particularly significant in
the case of merger of a binary black hole, as was recently pointed out by Favata
(2009, hereafter F09). Favata has shown (see Figure 1 of F09) that for the case of
an equal-mass binary, a metric memory jump δh was of the order of ~ 5 percent
of M/R, where M is the mass of the binary component and R is the co-moving
distance to the binary measured at redshift 0 (hereafter M is expressed in the geo-
metric units, i.e. $M = GM/c^2$). Furthermore, Favata has argued that the memory
jumps were potentially detectable by LISA with high signal-to-noise ratio. Fa-
vata's memory calculations make use of an approximate analytical treatment of the
mergers, and need to be followed up with more definitive numerical calculations.
Nevertheless, a number of analytical models explored in F09 show that the effect
is clearly of high importance, and thus further investigations of detectability of the
memory jumps are warranted.

Recently, there has been a renewed effort to measure gravitational waves from
widely separated supermassive black-hole (SMBH) binaries by using precise tim-
ing of galactic millisecond pulsars (Jenet et al., 2005; Manchester, 2006). In this
chapter we investigate whether pulsar timing arrays (PTAs) could be sensitive to
the memory jumps from physical mergers of the SMBHs at the end of the binary's
life. We demonstrate that modern PTAs (Manchester, 2006), after 10 years of oper-
ation, will be sensitive to mergers of $10^8 M_\odot$ black holes out to ~billion light years;
however the chances of actual detection are small. Futuristic PTA experiments,
like those performed on the Square Kilometer Array (Cordes et al., 2005), offer
a somewhat better prospect for the direct detection of gravitational-wave memory
jumps.

3.2 The signal

The gravitational waveform from a merger of SMBH pair consists of an *ac*-part
and a *dc*-part; see Figure 1 of F09. The *ac*-part is short-period and short-lived,
and hence is undetectable by a PTA. The *dc*-part is the gravitational-wave mem-
ory; it grows rapidly during the merger, on the timescale of $\sim 10M(1 + z) \simeq
10^4(M/10^8 M_\odot)(1 + z)$s, where M is the mass of the SMBHs (assumed equal) and
z is the redshift of the merger. After the burst passes, the change in metric persists,
and as we explain below, it is this durable change in the metric that makes the main

impact on the timing residuals. Realistic PTA programs are designed to clock each of the pulsars with $\sim$ 2-week intervals (Manchester, 2006, Bailes, private communications). Therefore, for $M = 10^8 M_\odot$ SMBHs the growth of the memory-related metric change is not time-resolved by the timing measurements. Moreover, even for $M = 10^{10} M_\odot$ SMBHs this growth occurs on the timescale much shorter than the duration on the experiment. We are therefore warranted to treat the dc-part of the gravitational wave as a discontinuous jump propagating through space,

$$ h(\vec{r}, t) = h_0 \times \Theta \left[(t - t_0) - \vec{n} \cdot \vec{r} \right], \qquad (3.1) $$

where h_0 is the amplitude of the jump, of the order of $0.05 M/R$, $\Theta(t)$ is the Heavy-side function, t_0 is the moment of time when the gravitational-wave burst passes an observer, $\vec{r}$ is the location in space relative to the observer, and $\vec{n}$ is the unit vector pointed in the direction of the wave propagation. Here and below we set $c = 1$. We have used the plane-wave approximation, which is justified for treating extragalactic gravitational waves as they propagate through the Galaxy.

For a single pulsar, the frequency of the pulse-arrival ν responds to a linearly polarised plane gravitational wave[1] according to the following equation (Estabrook & Wahlquist, 1975; Hellings & Downs, 1983):

$$ \frac{\delta \nu(t)}{\nu} = B(\theta, \phi) \times \left[h(t) - h(t - r - r \cos \theta) \right], \qquad (3.2) $$

where

$$ B(\theta, \phi) = \frac{1}{2} \cos(2\phi) \left(1 - \cos \theta \right). \qquad (3.3) $$

Here r is the Earth-pulsar distance at an angle θ to the direction of the wave propagation, ϕ is the angle between the wave's principle polarisation and the projection of the pulsar onto the plane perpendicular to the propagation direction, and $h(t)$ is the gravitational-wave strain at the observer's location. Substituting Equation (3.1) into the above equation, we obtain the mathematical form of the signal:

$$ \frac{\delta \nu(t)}{\nu} = h_0 B(\theta, \phi) \times \left[\Theta(t - t_0) - \Theta(t - t_1) \right], \qquad (3.4) $$

where $t_1 = t_0 + r(1 + \cos \theta)$. Thus the memory jump would cause a pair of pulse frequency jumps of equal magnitude and the opposite sign, separated by the time interval $r(1 + \cos \theta)$. Since typical PTA pulsars are at least $\sim 10^3$ light years away, a single merger could generate at most one of the frequency jumps as seen during the $\sim$ 10 years of a PTA experiment. The timing residuals from a single jump at

---

[1]It is always possible to represent the signal this way

$t = t_0$ are given by

$$m(t) = B(\theta, \phi)h_0 \times \Theta(t - t_0) \times (t - t_0). \qquad \boxed{3.5}$$

For a single pulsar the frequency jump is indistinguishable from a fast glitch, and therefore single-pulsar data can only be used for placing upper limits on gravitational-wave memory jumps. The situation would be different for an array of pulsars, where simultaneous pulse frequency jumps would occur in all of them at the time $t = t_0$ when the gravitational-wave burst would reach the Earth[2]. Therefore a PTA could in principle be used to detect memory jumps.

3.3 Single-source detection by PTAs.

In this section we develop a mathematical framework for single-source detection by PTAs. Our formalism is essentially Bayesian and follows closely the spirit of van Haasteren et al. (2009, see also chapter 2), although we will make a connection with the frequentist Wiener-filter estimator. We will then apply our general formalism to the memory jumps. The reader uninterested in mathematical details should skip the following subsection and go straight to the results in Section 3.5.

There is a large body of literature on the single-source detection in the gravitational-wave community (Finn, 1992; Owen, 1996; Brady et al., 1998). The techniques which have been developed so far are designed specifically for the interferometric gravitational-wave detectors like LIGO and LISA. There are several important modifications which need to be considered when applying these techniques to PTAs, among them

1. **Discreteness of the data set.** A single timing residual per observed pulsar is obtained during the observing run; these runs are separated by at least several weeks. This is in contrast to the continuous (for all practical purposes) data stream in LIGO and LISA.

2. **Subtraction of the systematic corrections.** The most essential of these is the quadratic component of the timing residuals due to pulsar spindown, but there may be others, e.g. jumps of the zero point due to equipment change, annual modulations, etc.

3. **Duration of the signal may be comparable to the duration of the experiment.** This is the case for both cosmological stochastic background considered in chapter 2, and for the memory jumps considered here. Thus frequency domain

---

[2]For an array of 20 pulsars, there is a significant chance that a pulsar would have a strong alignment with the burst source, in which case 2 oppositely directed frequency jumps could be observed for this pulsar. However, since this would be for one pulsar at most, and since one of the two jumps is then indistinguishable from a glitch, we can safely ignore this fact.

methods are not optimal, and time-domain formalism should be developed instead.

The Bayesian time-domain approach developed in chapter 2 and in this subsection is designed to tackle these three complications.

Consider a collection of N timing residuals δt_p obtained from clocking a number of pulsars. Here p is the composite index meant to indicate both the pulsar and the observing run together. Mathematically, we represent the residuals as follows:

$$\delta t_p = A \times s(t_p) + \delta t_p^n + Q(t_p). \tag{3.6}$$

Here $s(t_p)$ and A are the known functional form and unknown amplitude of a gravitational-wave signal from a single source, δt_p^n is the stochastic contribution from a combination of the timing and receiver noises, and

$$Q(t_p) = \Sigma_m \xi_m f_m(t_p) \tag{3.7}$$

is the contribution from systematic errors of known functional forms $f_m(t_p)$ but a-priory unknown magnitudes ξ_m. Below we shall specify $Q(t_p)$ to be the unsubtracted part of the quadratic spindown, however for now we prefer to keep the discussion as general as possible. We follow chapter 2 and rewrite Equation (3.6) in a vector form:

$$\vec{\delta t} = A\vec{s} + \vec{\delta t}^n + F\vec{\xi}. \tag{3.8}$$

Here the components of the column vectors $\vec{\delta t}$, $\vec{\delta t}^n$, $\vec{s}$, and $\vec{\xi}$ are given by δt_p, δt_p^n, $s(t_p)$, and ξ_m, and F is a non-square matrix with the elements $F_{pm} = f_m(t_p)$. Henceforth we assume that δt_p^n is the random Gaussian process, with the symmetric positive-definite coherence matrix C:

$$C_{pq} = < \delta t_p^n \delta t_q^n > . \tag{3.9}$$

We can now write down the joint probability distribution for A and ξ_m:

$$
\begin{aligned}
P(A, \xi_m | \vec{\delta t}) \quad = \quad & (1/M) P_0(A, \xi_m) \times \\
& \exp\left[-\frac{1}{2}(\vec{\delta t} - A\vec{s} - F\vec{\xi})^T \times C^{-1} \times, \right. \\
& \left. (\vec{\delta t} - A\vec{s} - F\vec{\xi}) \right].
\end{aligned} \tag{3.10}
$$

Here $P_0(A, \xi_m)$ is the prior probability distribution, and M is the overall normalisation factor. We now assume a flat prior $P_0(A, L_m) = const$, and marginalise over $\vec{\xi}$ in precisely the same way as shown in the Appendix of chapter 2. As a result, we

get the following Gaussian probability distribution for A:

$$P(A|\vec{\delta t}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(A - \bar{A})^2}{2\sigma^2}\right]. \tag{3.11}$$

Here, the mean value $\bar{A}$ and the standard deviation σ are given by

$$\bar{A} = \frac{\vec{s}^T C' \vec{\delta t}}{\vec{s}^T C' \vec{s}}, \tag{3.12}$$

and

$$\sigma = \left(\vec{s}^T C' \vec{s}\right)^{-1/2}, \tag{3.13}$$

where

$$C' = C^{-1} - C^{-1}F\left(F^T C^{-1}F\right)^{-1} F^T C^{-1}. \tag{3.14}$$

It is instructive and useful to re-write the above equations by introducing an inner product $\langle \vec{x}, \vec{y} \rangle$ defined as

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T C^{-1} \vec{y}. \tag{3.15}$$

Let us choose an orthonormal basis[3] $\hat{f}_i$ in the subspace spanned by $\vec{f}_m$, so that $\langle \hat{f}_i, \hat{f}_j \rangle = \delta_{ij}$. We also introduce a projection operator

$$R = 1 - \Sigma_m \left| \hat{f}_m \rangle\langle \hat{f}_m \right|, \tag{3.16}$$

so that $R\vec{x} = \vec{x} - \Sigma_m \langle \hat{f}_m, \vec{x} \rangle \hat{f}_m$. All the usual identities for projection operators are satisfied, i.e. $R^2 = R$ and $\langle R\vec{x}, R\vec{y} \rangle = \langle \vec{x}, R\vec{y} \rangle$. We can then write

$$\bar{A} = \frac{\langle \vec{s}, R\vec{\delta t} \rangle}{\langle \vec{s}, R\vec{s} \rangle}, \tag{3.17}$$

and

$$\sigma = \langle \vec{s}, R\vec{s} \rangle^{-1/2}. \tag{3.18}$$

If there are no systematic errors that need to be removed, then $R = 1$ and the Equations (3.17) and (3.18) represent the time-domain version of the Wiener-filter estimator.

3.3.1 Other parameters

So far we have assumed that the gravitational-wave signal has a known functional form but unknown amplitude, and have explained how to measure or constrain this

---

[3]This is always possible by e.g. the Gramm-Schmidt procedure.

amplitude. In reality, the waveform $\vec{s}(\vec{\eta}, t_P)$ will depend on a number of a-priori unknown parameters $\vec{\eta}$, such as the starting time of the gravitational-wave burst and the direction from which this burst has come. These parameters enter into the probability distribution function through $\vec{s}$ in Equation (3.10), and generally their distribution functions have to be estimated numerically. The estimates can be done via matched filtering (Owen, 1996) or by performing Markov-Chain Monte-Carlo (MCMC) simulations. In Section 3.5, we will demonstrate an example of an MCMC simulation for the memory jump. In this section, we show how to estimate an *average* statistical error on $\vec{\eta}$ for signals with high signal-to-noise ratios.

Let us begin with a joint likelihood function for the amplitude A and other parameters η:

$$L(A, \vec{\eta}) = -(1/2)\langle A\vec{s}(\vec{\eta}) - \vec{\delta t}, R(A\vec{s}(\vec{\eta}) - \vec{\delta t})\rangle + Const. \qquad (3.19)$$

We now fix A to its maximum-likelihood value $\langle \vec{s}(\vec{\eta}), R\vec{\delta t}\rangle / \langle \vec{s}(\vec{\eta}), R\vec{s}(\vec{\eta})\rangle$, and average over a large number of statistical realisations of the noise $\vec{\delta t}^n$. The so-averaged likelihood function is given by

$$
\begin{aligned}
L_{\mathrm{av}}(\vec{\eta}) \;=\; & -\frac{(1/2)A_t^2}{\langle \vec{s}(\vec{\eta}), R\vec{s}(\vec{\eta})\rangle} \Big[\langle \vec{s}(\vec{\eta}_t), R\vec{s}(\vec{\eta}_t)\rangle \langle \vec{s}(\vec{\eta}), R\vec{s}(\vec{\eta})\rangle - \\
& \langle \vec{s}(\vec{\eta}_t), R\vec{s}(\vec{\eta})\rangle^2 \Big],
\end{aligned}
\qquad (3.20)
$$

where $A_t, \vec{\eta}_t$ are the true values for the signal present in all data realisations. We have omitted the additive constant.

The expression in the square bracket is positive-definite, and L_{av} is quadratic in $\vec{\eta} - \vec{\eta}_t$ for the values of $\vec{\eta}$ close to the true values,

$$L_{\mathrm{av}}(\vec{\eta}) \simeq -(1/2)(\vec{\eta} - \vec{\eta}_t)G(\vec{\eta} - \vec{\eta}_t), \qquad (3.21)$$

where G is the positive-definite Fisher information matrix. Its elements can be expressed as

$$
\begin{aligned}
G_{ij} \;=\; & A_t^2/\langle \vec{s}, R\vec{s}\rangle \Big[\langle \vec{s}, R\vec{s}\rangle \langle \frac{\partial \vec{s}}{\partial \eta_i}, R\frac{\partial \vec{s}}{\partial \eta_j}\rangle - \\
& \langle \vec{s}, R\frac{\partial \vec{s}}{\partial \eta_i}\rangle \langle \vec{s}, R\frac{\partial \vec{s}}{\partial \eta_j}\rangle \Big],
\end{aligned}
\qquad (3.22)
$$

evaluated at $\eta = \eta_t$. The inverse of G specifies the average error with which parameters $\vec{\eta}$ can be estimated from the data.

3.4 Detectability of memory jumps

We now make an analytical estimate for detectability of the memory jumps. For
simplicity, we assume that all of the pulsar observations are performed regularly
so that the timing-residual measurements are separated by a fixed time Δt, and that
the whole experiment lasts over the time interval $[-T, T]$ (expressed in this way
for mathematical convenience). Furthermore, we assume that the timing/receiver
noise is white, i.e. that for a pulsar a

$$\langle \delta t_i^n \delta t_j^n \rangle = \sigma_a^2 \delta_{ij}. \tag{3.23}$$

This assumption is probably not valid for some of the millisecond pulsars (Verbiest
et al., 2009; van Haasteren et al., 2011, see also chapter 4). We postpone discussion
of non-white timing noise to future work.

   To keep our exposition transparent, we consider the case when the array con-
sists of a single pulsar a; generalisation to several pulsars is straightforward and is
shown later this section. Finally, we assume that the systematic error $Q(t_i)$ com-
prises only an unsubtracted component of the quadratic spindown,

$$Q(t_i) = A_0 + A_1 t_i + A_2 t_i^2, \tag{3.24}$$

where A_0, A_1, and A_2 are a-priori unknown parameters.

   We now come back to the formalism developed in the previous section. The
inner product defined in Equation (3.15) takes a simple form:

$$\begin{aligned} \langle \vec{x}, \vec{y} \rangle &= \frac{1}{\sigma_a^2} \sum_i x(t_i) y(t_i) \\ &\simeq \frac{1}{\sigma_a^2 \Delta t} \int_{-T}^{T} x(t) y(t) dt, \end{aligned} \tag{3.25}$$

where we have assumed $\Delta t \ll T$ and have substituted the sum with the integral in
the last equation. We now choose orthonormal basis vectors $\hat{f}_{1,2,3}(t)$ which span
the linear space of quadratic functions:

$$\begin{aligned} \hat{f}_1(t) &= \sigma_a \sqrt{\frac{\Delta t}{T}} \frac{1}{\sqrt{2}} \\ \hat{f}_2(t) &= \sigma_a \sqrt{\frac{\Delta t}{T}} \sqrt{\frac{3}{2}} \frac{t}{T} \\ \hat{f}_3(t) &= \sigma_a \sqrt{\frac{\Delta t}{T}} \sqrt{\frac{45}{8}} \left[\left(\frac{t}{T} \right)^2 - \frac{1}{3} \right]. \end{aligned} \tag{3.26}$$

From Equation (3.5) the gravitational-wave induced timing residuals are given by $\delta t(t) = h_0 s(t)$, where

$$s(t) = B(\theta, \phi) \times \Theta(t - t_0) \times (t - t_0).$$ (3.27)

The expected measurement error of the jump amplitude h_0 is given by Equation (3.18):

$$\sigma_{h_0} = \left[\langle \vec{s}, \vec{s} \rangle^2 - \sum_{i=1,2,3} \langle \vec{s}, \hat{f}_i \rangle^2 \right]^{-1/2}.$$ (3.28)

Substituting Equations (3.25), (3.26), and (3.27) into Equation (3.28), one gets after some algebra:

$$\sigma_{h_0} = \frac{1}{B(\theta, \phi)} \frac{\sigma_a}{T} \sqrt{\frac{48}{N p^3 \left(1 - \frac{15}{16} p\right)}}.$$ (3.29)

Here $N = 2T/\Delta t$ is the number of measurements, and

$$p = 1 - (t_0/T)^2.$$ (3.30)

For an array consisting of multiple pulsars, and with the assumption that the timing residuals are obtained for all of them during each of the N observing runs, the above expression for σ_{h_0} is modified as follows:

$$\sigma_{h_0} = \frac{\sigma_{\text{eff}}}{T} \sqrt{\frac{48}{N p^3 \left(1 - \frac{15}{16} p\right)}},$$ (3.31)

where

$$\sigma_{\text{eff}} = \left[\sum_a \left(B^2(\theta_a, \phi_a)/\sigma_a^2 \right) \right]^{-1/2}.$$ (3.32)

Several remarks are in order:

1. The error σ_{h_0} diverges when $p = 0$, i.e. when $t_0 = \pm T$. This is as expected: when the memory jump arrives at the beginning or at the end of the timing-array experiment, it gets entirely fitted out when the pulsar spin frequency is determined, and is thus undetectable.

2. Naively, one may expect the optimal sensitivity when the jump arrives exactly in the middle of the experiment's time interval, i.e. when $t_0 = 0$. This is not

so; the optimal sensitivity is achieved for $t_0/T = \pm 1/\sqrt{5}$ when the error equals

$$\sigma_{h_0} = \frac{\sigma_{\text{eff}}}{T}\sqrt{\frac{375}{N}}. \tag{3.33}$$

3. The sky-average value for $B^2(\theta,\phi)$ is $1/6$. Therefore, for an array consisting of a large number of pulsars N_p which are distributed in the sky isotropically and which have the same amplitude of timing/receiver noise $\sigma_a = \sigma$, the σ_{eff} in Equation (3.31) is given by

$$\sigma_{\text{eff}} = \sigma\sqrt{6/N_p}. \tag{3.34}$$

4. While the timing precision of future timing arrays is somewhat uncertain, it is instructive to consider a numerical example. Lets assume $T = 5\text{yr}$ (i.e., the 10-year duration of the experiment), $N = 250$ (i.e., roughly bi-weekly timing-residual measurements), $N_p = 20$ isotropically-distributed pulsars (this is the current number of clocked millisecond pulsars), and $\sigma_a = 100\text{ns}$ (this sensitivity is currently achieved for only several pulsars). Then for optimal arrival time $t_0 = \pm T/\sqrt{5}$, the array sensitivity is

$$\sigma_{h_0} = 4.5 \times 10^{-16}. \tag{3.35}$$

For a binary consisting of two black holes of the mass M, the memory jump is estimated in F09 to be

$$h_0 = \eta\frac{M}{R} \simeq 8 \times 10^{-16}\frac{\eta}{0.05}\left(\frac{M}{10^8 M_\odot}\right)\left(\frac{10^9\text{light-years}}{R}\right), \tag{3.36}$$

where $\eta \sim 0.05$ is the direction-dependent numerical parameter. In this example, the pulsar-timing array is sensitive to the memory jumps from black-hole mergers at redshifts $z < z_0$, where $z_0 \sim 0.1$ for $M = 10^8 M_\odot$, and $z_0 \sim 1$ for $M = 10^9 M_\odot$.

3.4.1   Sensitivity to the arrival time

It is possible to estimate the array's sensitivity to the memory-jump arrival time, t_0. We use Equations (3.22) and (3.27), and after some algebra[4] get

$$\sigma_{t_0} = T\left(\frac{h_0 T}{\sigma_{\text{eff}}}\right)^{-1}\sqrt{2/N}\chi(p), \tag{3.38}$$

---

[4]A useful identity:

$$\frac{\partial[(t - t_0)\Theta(t - t_0)]}{\partial t_0} = -\Theta(t - t_0). \tag{3.37}$$

where

$$\frac{1}{\chi^2(p)} = \frac{1}{2}p\left(1 + \frac{5}{4}p^2 - 2p\right) - \frac{3(1-p)p\left[1 - (5/8)p\right]^2}{2[1 - (15/16)p]}.$$ (3.39)

3.4.2 Source position

The array's sensitivity gravitational-wave memory is dependent on source position since the number and the position of the pulsars in current PTAs is not sufficient to justify the assumption of isotropy made in Equation (3.34). We will therefore calculate the value of $\left[\sum_a B^2 (\theta_a, \phi_a)\right]^{-1/2}$ for current PTAs. Since the polarisation of the gravitational-wave memory signal is an unknown independent parameter, we average over the polarisation and obtain for the angular sensitivity:

$$\sigma_{h_0} (\phi_s, \theta_s) \quad \sim \quad \left[\sum_a B^2 (\phi_s, \theta_s, \theta_a, \phi_a)\right]^{-1/2}$$ (3.40)

$$= \quad \left[\sum_a \frac{1}{8} (1 + \cos \theta_a (\phi_s, \theta_s))^2\right]^{-1/2}.$$ (3.41)

Here we have assumed that all pulsars have equal timing precision. ϕ_s and θ_s are the position angles of the gravitational-wave memory source, and θ_a is the polar angle of pulsar a in a coordinate system with (ϕ_s, θ_s) at the north-pole. In figure 3.1 and 3.2 the sensitivity to different gravitational-wave memory source positions is shown for respectively the European Pulsar Timing Array and the Parkes Pulsar Timing Array projects.

3.5 Tests using mock data

We test the array's sensitivity to gravitational-wave memory signals using mock timing residuals for a number of millisecond pulsars. In this whole section, all the mock timing residuals were generated in two steps:

1) A set of timing residuals was generated using the pulsar timing package tempo2 (Hobbs et al., 2006). We assume that the observations are taken tri-weekly over a time-span of 10 years. The pulsar timing noise was set to 100 ns white noise.

2) A gravitational-wave memory signal was added according to Equation (3.5), with a memory-jump arrival time set to be optimal for sensitivity: $\frac{t_0}{T} = \frac{1}{\sqrt{5}}$. The direction and polarisation of the gravitational-wave memory signal were chosen randomly - the coordinates happened to have declination 90°.

In the following subsections we describe tests which have fixed parameters for step
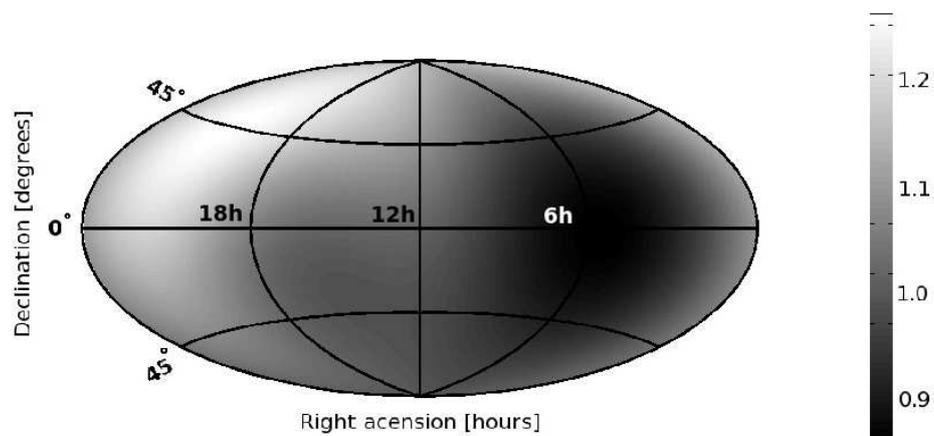
Figure 3.1: The relative sensitivity σ_{eff} for the pulsars of the European Pulsar Timing Array. The scaling has been chosen such that a value of 1 indicates that the same sensitivity for that source position would have been achieved with a perfect isotropic PTA (i.e. $B^2 = \frac{1}{6}$)
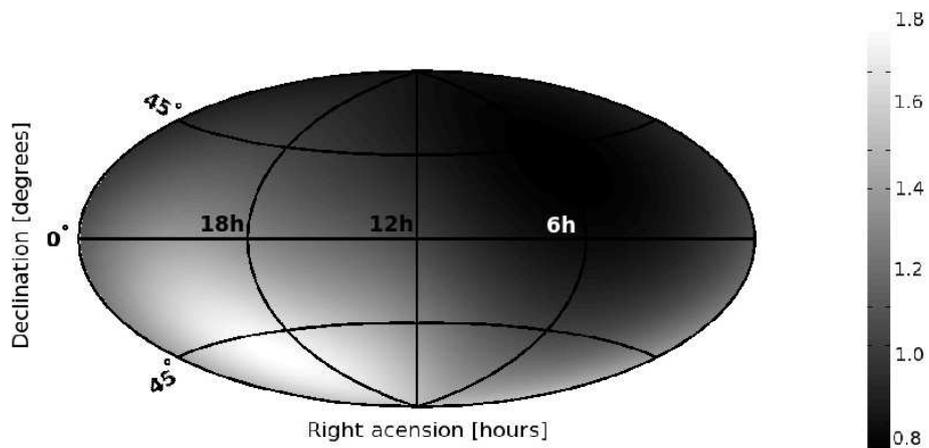
Figure 3.2: The relative sensitivity σ_{eff} for the pulsars of the Parkes Pulsar Timing Array. The scaling has been chosen such that a value of 1 indicates that the same sensitivity for that source position would have been achieved with a perfect isotropic PTA (i.e. $B^2 = \frac{1}{6}$)

1, but systematically varied amplitude for step 2, and we use these tests to study the sensitivity of the array.

3.5.1 Used models

In principle, we would like to realistically extrapolate the results we obtain here for mock datasets to future real datasets from PTA projects. Several practical notes are in order to justify the models we use here to analyse the mock datasets:

1) From equation (3.24) onward, we assume that the systematic-error contributions to the timing residuals consist only of the quadratic spindown. In reality, pulsar observers must fit many model parameters to the data, and have developed appropriate fitting routines within timing packages like tempo2. Similar to the quadratic spindown discussed in this chapter, all the parameters of the timing model are linear or linearised in Tempo2, and therefore those parameters are of known functional form. Since the subtraction of quadratic spindown decreases the sensitivity of the PTA to gravitational-wave memory signals, we would expect the same thing to be true for the rest of the timing model.

2) The error-bars on the pulse arrival time obtained from correlating the measured pulse-profile with the template of the pulse-profile are generally not completely trusted. Many pulsar astronomers invoke an extra "fudge" factor that adjusts the error-bars on the timing-residuals to make sure that the errors one gets on the parameters of the timing-model are not underestimated. Usually the "fudge" factor, which is known as an EFAC value is set to the value which makes the reduced χ^2 of the timing solution to be equal to 1.

In order to check the significance of both limitations 1 & 2, we perform the following test. We take a realistic set of pulsars with realistic timing models: the pulsar positions and timing models of the PPTA pulsars. We then simulate white timing-residuals and a gravitational-wave memory signal with amplitude $h_0 = 10^{-15}$, and we produce the posterior distribution of Equation (3.11) in three different ways:

a) We marginalise over only the quadratic functions of Equation (3.26), which should yield the result of Equation (3.31).

b) We marginalise over the all timing model parameters included in the Tempo2 analysis when producing the timing-residuals.

c) We marginalise over all the timing model parameters, and we also marginalise over the EFAC values using the numerical techniques of chapter 2. By estimating the EFAC value simultaneously with the gravitational-wave memory signal, we are able to completely separate the two effects. Note that this procedure will not destroy information about the relative size of the error-bars for timing-residuals of the same pulsar.

We present the result of this analysis in Figure 3.3. Based on the 185 observations per pulsar in the dataset and the direction of the gravitational-wave memory signal, we can calculate the theoretical sensitivity of the array using Equation (3.33) and Equation (3.34). This yields a value of:

$$\sigma_{h_0} = 6.4 \times 10^{-16}. \qquad \boxed{3.42}$$

We can also calculate this value for the three graphs in Figure 3.3. The three graphs lie close enough on top of each other to conclude that one value applies to all three of them:

$$\sigma_{h_0} = 6.5 \times 10^{-16}, \qquad \boxed{3.43}$$

which is in good agreement with the theoretical value. It appears that both note 1 and 2 mentioned above are not of great influence to the sensitivity of PTAs to gravitational-wave memory detection; the theoretical calculations of this work are a good representation of the models mentioned in this section.

3.5.2 Upper-limits and detecting the signal

When there is no detectable gravitational-wave memory signal present in the data, we can set some upper-limit on the signal amplitude using the method presented in this chapter. Here we will analyse datasets with no or no fully detectable gravitational-wave memory signal in it, and a dataset with a well-detectable signal using the MCMC method of chapter 2. We will calculate the marginalised posterior distributions for the 5 parameters of the gravitational-wave memory signal. The interesting parameters in the case of an upper-limit are the amplitude and the arrival time of the jump. A marginalised posterior for those two parameters are then presented as two-dimensional posterior plots. Note that the difference with the analysis in Section 3.5.1 is that we vary all gravitational-wave memory parameters, instead of only the amplitude. Note that we do marginalise over all the EFAC values as discussed in Section 3.5.1, unless stated otherwise.

In Figure 3.4 we show the result of an analysis of a dataset where we have not added any gravitational-wave memory signal to the timing-residuals. The $3 - \sigma$ contour is drawn, which serves as an upper-limit to the memory amplitude. We see that we can exclude a gravitational-wave memory signal at $\frac{t_0}{T} = \frac{1}{\sqrt{5}}$ of amplitude 3×10^{-15} and higher. We see that this value is over a factor of 4 higher than what is predicted by Equation (3.42). This is to be expected, since:
1) We give a $3 - \sigma$ limit here, instead of the $1 - \sigma$ sensitivity.
2) We also marginalise over the arrival time and other parameters of the memory signal, reducing the sensitivity.
Because of these reasons, we argue that the minimal upper-limit one can set on the
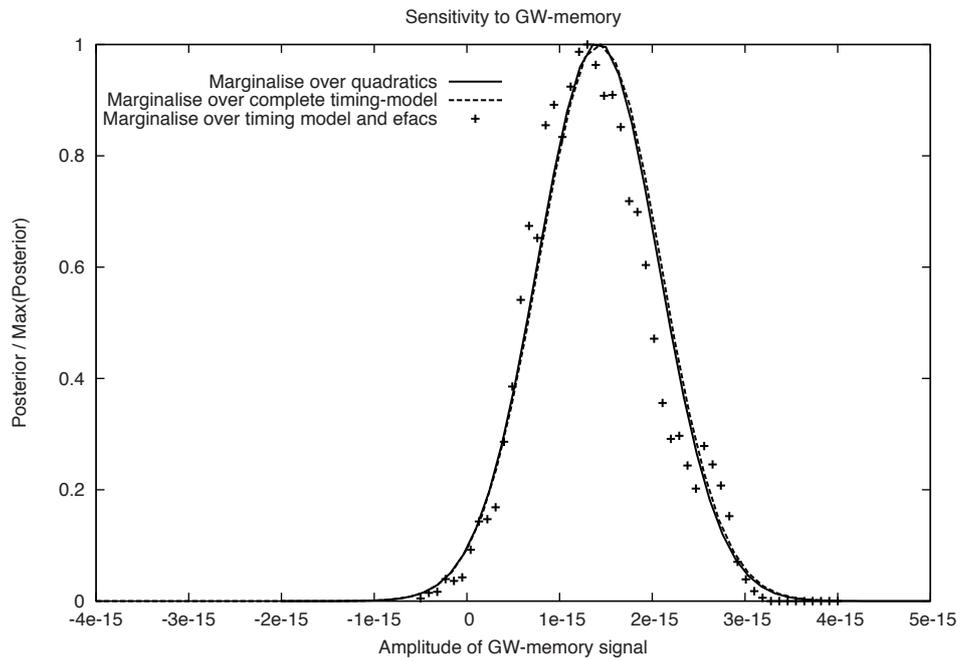
Figure 3.3: The posterior distribution of the gravitational-wave memory ampli-
tude. The two solid lines are the result of an analysis where we only analytically
marginalise over the full timing model or just quadratic spindown. The points
are the result of a marginalisation over the full timing model and the EFAC val-
ues as well. From the Gaussians, the sensitivity can be reliably estimated at:
$\sigma_{h_0} = \frac{FWHM}{2\sqrt{2\ln 2}} = 6.5 \times 10^{-16}$
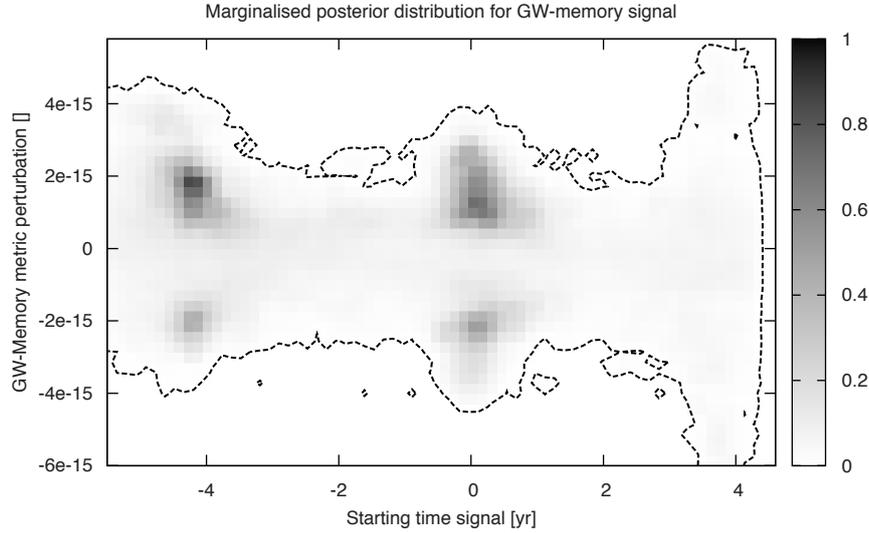
Figure 3.4: The marginalised posterior distribution for the gravitational-wave memory signal amplitude and arrival time of the jump. In this case a dataset was analysed that did not contain any gravitational-wave memory signal.

gravitational-wave memory signal using a specific PTA is the sensitivity calculated using Equation (3.42) multiplied by 4.

Next we produce a set of timing-residuals with a memory signal of amplitude $h_0 = 10^{-15}$. According the result mentioned above, the memory signal should not be resolvable with this timing precision. The result is shown in Figure 3.5. We see that we can indeed merely set an upper-limit again. In order to check the effect of marginalising over the EFAC values as mentioned in Section 3.5.1, we also perform an analysis where we pretend we do know the EFAC values prior to the analysis. The result is shown in Figure 3.6. We see no significant difference between the two models.

Finally, we also analyse a dataset with a gravitational-wave memory signal with an amplitude larger than the 3×10^{-15} upper-limit of the white set mentioned above. Here we have added a memory signal with an amplitude of 10^{-14}. In Figure 3.7 we see that we have a definite detection of the signal: if we consider the $3 - \sigma$ contours, we see that we can restrict the gravitational-wave memory amplitude between $\left[6.6 \times 10^{-15}, 1.35 \times 10^{-14}\right]$. Again, this value is higher than the value
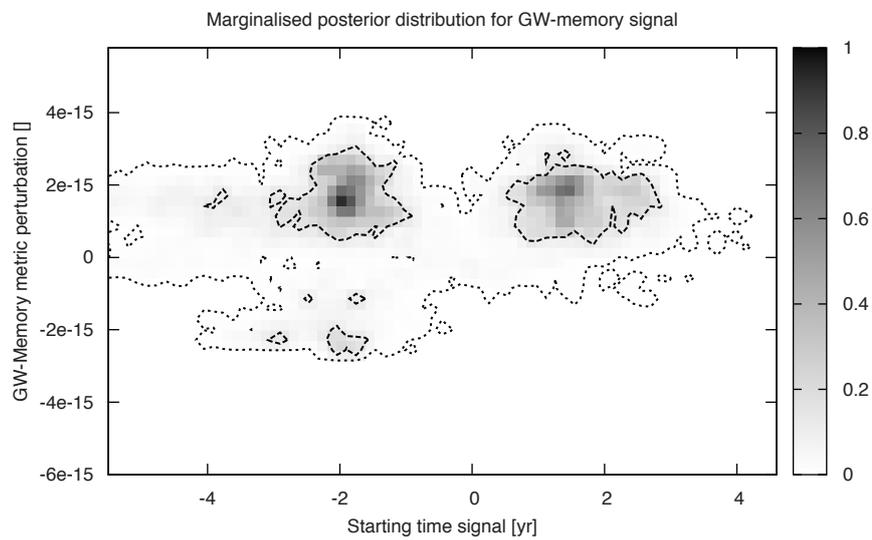
Figure 3.5: The marginalised posterior distribution for the gravitational-wave memory signal amplitude and arrival time of the jump. Here a gravitational-wave signal with an amplitude of 10^{-15} was added to the white residuals. The contour drawn is the $3 - \sigma$ contour.
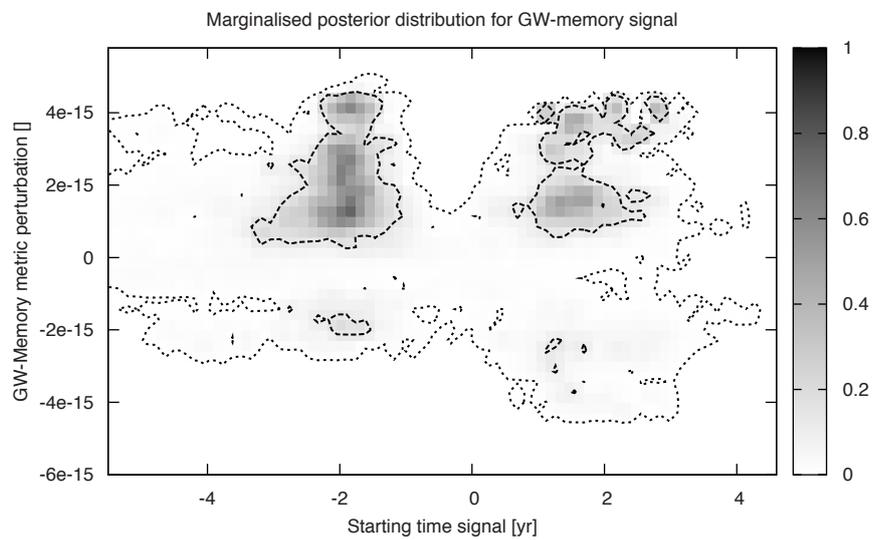
Figure 3.6: The marginalised posterior distribution for the gravitational-wave memory signal amplitude and arrival time of the jump. Here a gravitational-wave signal with an amplitude of 10^{-15} was added to the white residuals. This analysis has been done without marginalising over the EFAC values. The contour drawn is the $3 - \sigma$ contour.
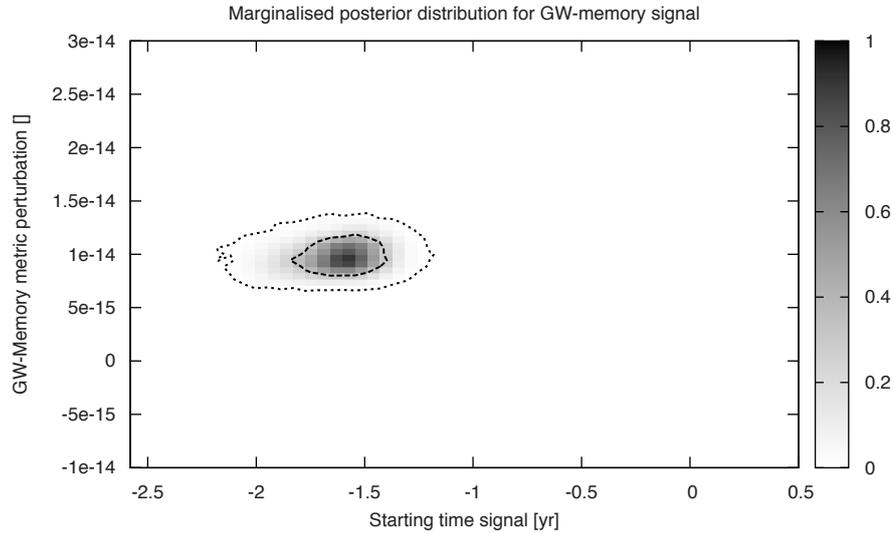
Figure 3.7: The marginalised posterior distribution for the gravitational-wave memory signal amplitude and arrival time of the jump. Here a gravitational-wave signal with an amplitude of 10^{-14} was added to the white residuals, indicated with a '+' in the figure. The contours drawn are the $1 - \sigma$ and $3 - \sigma$ contours.

predicted by Equation (3.42) due to us including more parameters in the model than just the memory amplitude. In Figure 3.8 we see that we can also reliably resolve the position of the source in this case.

3.6 Discussion

In this chapter, we have shown that gravitational-wave memory signals from SMBH binary mergers are in principle detectable by PTAs, and that $2 - \sigma$ constraints are possible on $M = 10^8 M_\odot$ mergers out to redshift of ~ 0.1 (while those with $M = 10^{10} M_\odot$ should be detectable throughout the Universe). How frequently do these mergers occur during the PTA lifetime? Recent calculations of Sesana et al. (2007, SVH) are not too encouraging. SVH compute, for several models of SMBH merger trees, the rate of SMBH mergers as seen on Earth, as a function of mass (their figure 1d), as well as a multidute of other parameters for these mergers.
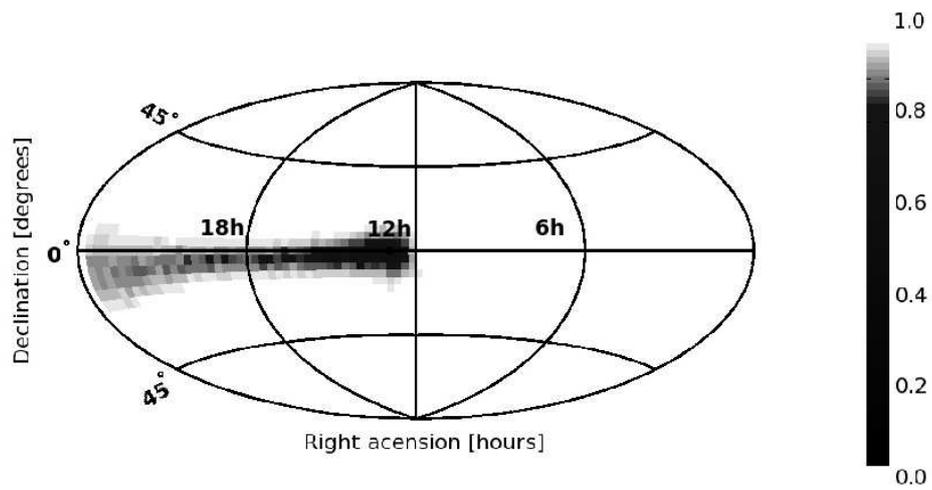
Figure 3.8: The marginalised posterior distribution for the sky location of the gravitational-wave memory signal. We can see here that we can marginally determine the direction of the source. The source positions used to generate the residuals were: (declination, right ascension) = $(0^0, 12.4\text{hr})$.

From their plots one infers few$\times 10^{-2} - 10^{-3}$ PTA-observable mergers per year,
which converts to at most $0.1 - 0.01$ detected mergers during the PTA lifetime of
~ 10 years (NB: during the PTA existence, only a fraction of time will spent near
the arrival times with optimal sensitivity). It is conceivable that SVH estimates are
on the conservative side, since the mergers of heavy black holes may be stalled
(due to the "last parsec" problem) and may occur at a significantly later time than
the mergers of their host halos. In this case, some fraction of high-redshift mergers
may be pushed towards lower redshifts and become PTA-detectable. Detailed cal-
culations are needed to find out whether this process could substantially increase
the rate of PTA-detectable mergers. It is also worth pointing out that a futuristic
PTA experiment based on a Square Kilometre Array may attain up to an order of
magnitude higher sensitivity than the currently developed PTAs.

The methods presented in this chapter are useful beyond the particular applica-
tion that we discuss. The method presented here is suitable for any single-source
detection in general when the gravitational waveform has known functional form.
Further applications will be presented elsewhere.

3.6.1   Comparison with other work

When this work was already finished, a preprint by Pshirkov et al. (2009, PBP) has
appeared on the arxiv which has carried out a similar analysis to the one presented
here. Our expressions for the signal-to-noise ratio for the memory jump agree for
the case of the white pulsar noise. PBPs treatment of cosmology is more detailed
than ours, while the moderately pessimistic predicted detection rates are broadly
consistent between the 2 analyses. Our method for signal extraction is more gen-
erally applicable than PBS's since it is optimised for any spectral type of pulsar
noise, takes into consideration not just the signal magnitude but also other signal
parameters, and is tested on mock data.

Simultaneously with this work, an analysis by Seto (2009) also appeared on
the arxiv preprint service. Seto has also noticed that the memory effect from black-
hole mergers is of interest for PTAs; however, the details of his analysis are very
different from ours. His signal-to-noise value is estimated very approximately, as
compared to both PBP's and our treatment. However, Seto made an interesting
observation that the gravitational-wave background from wide black-hole binaries
may be an important source of noise for the memory-jump detection. We plan to
address this issue in our future work.

4

Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data

The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.

Albert Einstein

Abstract

Direct detection of low-frequency gravitational waves ($10^{-9} - 10^{-8}$ Hz) is the main goal of pulsar timing array (PTA) projects. One of the main targets for the PTAs is to measure the stochastic background of gravitational waves (GWB) whose characteristic strain is expected to approximately follow a power-law of the form $h_c(f) = A(f/\mathrm{yr}^{-1})^\alpha$, where f is the gravitational-wave frequency. In this chapter we use the current data from the European PTA to determine an upper limit on the GWB amplitude A as a function of the unknown spectral slope α with a Bayesian inference method, by modelling the GWB as a random Gaussian process. For the case $\alpha = -2/3$, which is expected if the GWB is produced by supermassive black-hole binaries, we obtain a 95% confidence upper limit on A of 6×10^{-15}, which is 1.8 times lower than the 95% confidence GWB limit obtained by the Parkes PTA in 2006. Our approach to the data analysis incorporates the multi-telescope nature of the European PTA and thus can serve as a useful template for future intercontinental PTA collaborations.

4.1   Introduction

The first direct detection of gravitational waves (GWs) would be of great importance to astrophysics and fundamental physics: it would confirm some key predictions of general relativity, and lay the foundation for observational gravitational-wave astronomy. Pulsar Timing Array (PTA) projects are collaborations which aim to detect low-frequency (10^{-9}—10^{-8}Hz) extragalactic gravitational waves directly, by using a set of Galactic millisecond pulsars as nearly-perfect Einstein clocks (Foster & Backer, 1990). The basic idea is to exploit the fact that millisecond pulsars create pulse trains of exceptional regularity. GWs perturb space-time between the pulsars and the Earth, and this creates detectable deviations from the strict periodicity in the arrival times of the pulses (TOAs) (Estabrook & Wahlquist, 1975; Sazhin, 1978; Detweiler, 1979).

One of the main astrophysical targets of the PTAs is to measure the stochastic background of gravitational waves (GWB). This GWB is expected to be generated by a large number of black-hole binaries located at the centres of galaxies (Begelman et al., 1980; Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003; Sesana et al., 2008), by relic gravitational waves (Grishchuk, 2005), or, more speculatively, by oscillating cosmic-string loops (Damour & Vilenkin, 2005; Ölmez et al., 2010).

Currently, there are three independent PTA groups:
(i) the Australian-based programme PPTA, the Parkes Pulsar Timing Array, which uses data from the Parkes telescope (Hobbs et al., 2009; Verbiest et al., 2010), and archival Arecibo data.
(ii) the North-American based programme NANOGrav, North-American Nanohertz Observatory for Gravitational waves, which uses both the Green Bank Telescope (GBT), and the Arecibo radio telescope (Jenet, 2009).
(iii) and the European programme EPTA, European Pulsar Timing Array, which uses five different radio telescopes: the Lovell telescope near Manchester, United Kingdom, the Westerbork Synthesis Radio Telescope (WSRT) in the north of the Netherlands, the Effelsberg Telescope (EFF) near Bonn in Germany, the Nançay Radio Telescope (NRT) near Nançay in France, and the Sardinia Radio Telescope (SRT) in Sardinia, Italy[1].
It is likely that the first detection of GWs by a PTA will occur as a result of a joint effort of all current PTA projects: an International Pulsar Timing Array (IPTA; Hobbs et al., 2010). This will involve the combination of data from several different telescopes, each of them with its own specific hardware elements and software analysis tools. Combining data of different observatories is a challenging task,

---

[1]The SRT is expected to become operational in 2011 (Tofani et al., 2008)

which requires extra care when dealing with the high quality data of modern observatories (Janssen, 2009).

In this chapter, we present a methodology on how to combine the data from several radio telescopes and use it in an optimal way to obtain the information on extragalactic gravitational waves. We use the data from three different radio telescopes located on the European continent, to place a new upper limit on the amplitude of the GWB. As part of our analysis, we obtain detailed information about the statistical properties of the individual pulse time series.

The calculation of upper limits on the GWB, based on pulsar timing, go as far back as the early 1990's (Stinebring et al., 1990; Kaspi et al., 1994; McHugh et al., 1996; Lommen, 2002). These analyses have been based on high quality datasets for single millisecond pulsars. The most stringent upper limits have been obtained recently by Jenet et al. (2006), who have used PPTA data and archival Arecibo data for several millisecond pulsars. Our dataset is different from that used by Jenet et al. (2006) since it includes only the pulse times of arrival measured by the EPTA telescopes, even though some of the pulsars are being timed by multiple PTA groups. The Bayesian inference method we use to obtain an upper limit on the GWB is also different from the methods used by all of the previous studies. Its potential advantages include the use of cross correlations between TOAs of different pulsars, and the simultaneous constraint on both the amplitude and spectral index of the GWB.

The outline of the chapter is as follows. In Section 4.2 we give a brief general overview of pulsar timing observations. In Section 4.3 we detail the observations from all of the EPTA telescopes which were used for our analysis. In Section 4.4 we compare the different data analysis methods that have thus far been developed for PTAs, and we motivate our choice for the Bayesian inference for placing a limit on the GWB. We outline the data analysis procedure in Section 4.5, after which, in Section 4.6, we present the upper limits on the amplitude of the GWB, and also the spectral analysis of the individual pulsar noises. Finally, in Section 4.7 we discuss the astrophysical implications of our results, and provide an outlook for the future.

4.2 EPTA data analysis

In this section we present a brief overview of the observations, instrumentation and data analysis used at the different EPTA observatories for transforming a series of measured pulses to a TOA.

The complete data reduction process that converts the incoming data stream from a radio telescope into one single TOA per observation, called "the pipeline", is optimised by hand with much care and is observatory specific. The process can

be described in five general steps, shown in Figure 4.1:
1) The incoming radio waves are received by the telescope.
2) The signal is converted from analog to digital, at a Nyquist sampled rate.
3) Data is (coherently) de-dispersed and, if possible, Stokes parameters are formed.
4) The de-dispersed timeseries are folded at the pulsar period, resulting in averaged pulse profiles. Typically a timespan containing several 10^5 pulses is used for each TOA.
5) A cross-correlation with a template pulse profile yields a TOA and associated uncertainty (Taylor, 1992).

Individual pulse amplitudes and pulse shapes are highly irregular, and pulse phases vary significantly from pulse to pulse (Cordes & Shannon, 2010). Therefore careful averaging *(folding)* has to be performed to obtain a single TOA. Furthermore, the interstellar medium (ISM) results in significant delays of the arrival time of the pulses over the receiver bandwidth. As a large bandwidth is required to reliably detect a pulse, accounting for the ISM is key for precision timing.

Differences in templates used, e.g. the use of integrated profiles versus analytic templates, all based on single–observatory data, and the difference in definition of the reference point in a template will result in offsets between data sets generated by different observatories. All extra offsets in our data will lead to information loss of other signals like the GWB. Therefore, using a common template for each pulsar at all observatories is desirable, and will be implemented in the near future.

The realisation of the five steps and therefore their output (the resulting TOA) might differ among observatories. Understanding and accounting for those differences is essential for the correct analysis and optimal combining of the EPTA data. A more detailed study on this subject is in preparation (Janssen et al. 2011).

The cross-correlation between the folded profile and the template yields an uncertainty of the TOA (Taylor, 1992). One would like this uncertainty to be solely due to the radiometer noise, i.e. the noise intrinsic to the measurement, but in practice the errors sometimes appear to have been systematically over- or underestimated. It is a common practice, which we follow here, to allow for an extra parameter to multiply these uncertainties for each pulsar-observatory-backend combination (Hobbs & Edwards, 2006). This extra multiplicative factor allows the TOA uncertainties to statistically account for the TOA scatter: the deviations of the strict periodicity of the pulses. This is clearly unsatisfactory, and in future timing experiments the origin of the predicted and measured TOA scatter will have to be thoroughly investigated.

Input

Output
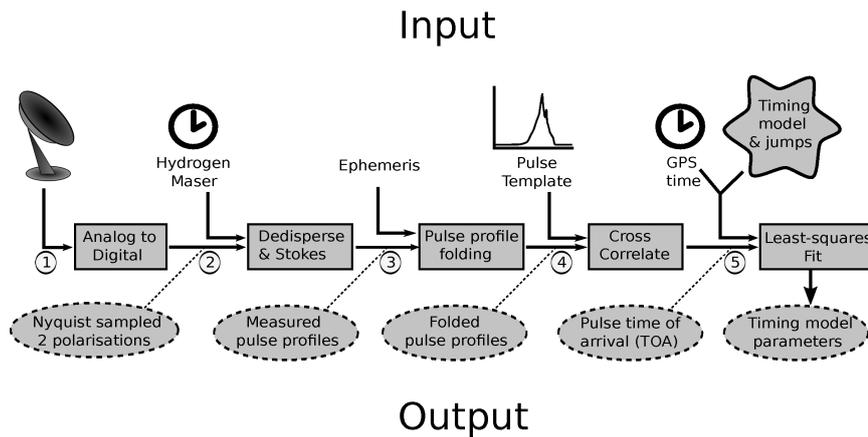
Figure 4.1: The processing pipeline for pulsar timing, step by step

Telescope	WSRT	NRT	EFF
Equivalent dish size (m)	93.5	94.4	100
Centre observing frequencies (MHz)	1380	1398, 2048	1400
Observing bandwidth (MHz)	80	64/128	28-112
Obs. time per month per pulsar	1x30 min	4-6x60min	1x30 min
Pulsar backend	PuMaI	BON	EBPP
Dedispersion	incoherent	coherent	coherent
Used templates	integrated profiles	integrated profiles	analytic

Table 4.1: Details of the different EPTA observatories relevant for this chapter. The NRT observing bandwidth has doubled to 128 MHz in July 2009.

4.3 EPTA observations

4.3.1 Overview of the observatories

We have used pulsar timing observations of five radio pulsars, observed with three
of the EPTA telescopes, to set a limit on the GWB. See Table 4.1, Fig. 4.2 and Ap-
pendix B of this chapter for an overview of the data sets used and the properties of
each telescope. Each pulsar was observed on average once every month for 30 min-
utes at each telescope. Although additional observing frequencies are commonly
used at WSRT and EFF, their respective 1380 and 1400 MHz observing bands have
the best sensitivity and result in the highest precision TOAs. Therefore we have
only used observations taken at those frequencies at WSRT and EFF for the anal-
ysis presented in this chapter. The data were either coherently de-dispersed (NRT
and EFF) or incoherently de-dispersed (WSRT). The observations were folded and
cross-correlated with an analytic template (EFF), or a high S/N, observatory spe-
cific, template (WSRT & NRT), to calculate one time-of-arrival (TOA) per obser-
vation. See e.g. Lazaridis et al. (2009) for a more complete description of the
observing procedures and data analysis at the different observatories.

As discussed, any change to the pipeline or to the input of the pipeline can
result in a difference in the calculated TOAs. We emphasise that it is essential to
correctly identify these systematic effects and include them in the modelling of the
TOAs. In our analysis, we have done this by introducing jumps between TOAs of
the same pulsar anywhere the pipeline differs in some way.

Once the complete set of data for each pulsar is obtained, and corrected for
global drifts by comparing to UTC, it is fit with the timing model. The timing
model is a multi-parameter fit that represents our best knowledge of the many de-
terministic processes that influence the values of the TOAs. The timing residuals
are then produced by subtracting the timing model, which is subsequently opti-
mised by minimising these residuals through a least–squares fit. This was done
using the pulsar timing package Tempo2 (Hobbs et al., 2006).

4.3.2 Selection of data sets

The European observatories have been timing millisecond pulsars for many years,
and potentially all of that data could be used in the calculation of an upper limit on
the GWB. However, like Jenet et al. (2006) we choose to use only the data from the
pulsars which perform best as ideal clocks, e.g. those with the highest precision
TOAs and the most straight-forward noise characteristics.

TOA precision is not the only factor that determines the sensitivity to the GWB;
other factors like the total timing baseline and the number of observations (i.e.

TOAs) affect this sensitivity as well. A great advantage of the EPTA data is that several pulsars have been monitored for a relatively long time: over 10 years. To determine which timing residuals (i.e. pulsar-observatory combinations) are most useful for GWB detection, we analyse each dataset separately. By doing this we can determine the sensitivity to the GWB of a set of TOAs: the lower the 3-σ upper limit $h_c^{\max}(1\mathrm{yr})$ we get using only a particular set of TOAs, the more sensitive that set of TOAs is to the GWB.

The timing residuals of the selected pulsars are shown in Figure 4.2. These five pulsars significantly outperform the other pulsars being timed by the EPTA in terms of how well they can limit the GWB amplitude: these five pulsars can each individually limit the GWB well below $h_c(1\mathrm{yr}) = 10^{-13}$ for $\alpha = -2/3$, whereas other current EPTA datasets typically perform worse by a factor of several. Since there is such a difference between this set of five pulsars, and the other pulsars that have been observed by the EPTA, we do not expect to gain any significant sensitivity by including more pulsars that cannot meet this constraint. We therefore choose $h_c^{\max}(1\mathrm{yr}) \leq 10^{-13}$ with $\alpha = -2/3$ as a constraint for including a dataset in our calculation.

In addition to this constraint, we also demand that datasets that just barely satisfy $h_c^{\max}(1\mathrm{yr}) \leq 10^{-13}$ do not show prominent low-frequency ("red") timing noise. Our criterion for presence of the latter is a peak in the posterior distribution which is inconsistent with zero amplitude for $\alpha \leq 0$.

4.4 Overview of data analysis methods

Over the years, different data analysis methods for PTAs have been developed. In this section we briefly discuss the differences between these methods, covering the merits and the weaknesses of each, and we motivate our choice for the Bayesian inference method.

Currently developed data analysis methods for isotropic GWB studies with PTAs can be divided in two classes:
1) methods designed to set limits on the GWB characteristic strain.
2) methods designed to detect the GWB signal, and measure its parameters.
In sections 4.4.1 and 4.4.2 we describe the differences between the two approaches. Among methods that belong to the first category are those of Stinebring et al. (1990); Kaspi et al. (1994); McHugh et al. (1996); Lommen (2002); Jenet et al. (2006). Methods that are primarily focused on detecting the GWB include those of Jenet et al. (2005); Anholm et al. (2009); van Haasteren et al. (2009); Yardley et al. (2011); see also chapter 2 of this thesis. As presented in section 4.6 of this chapter, the present-day data of the EPTA, is not sufficient to make a detection of the GWB,
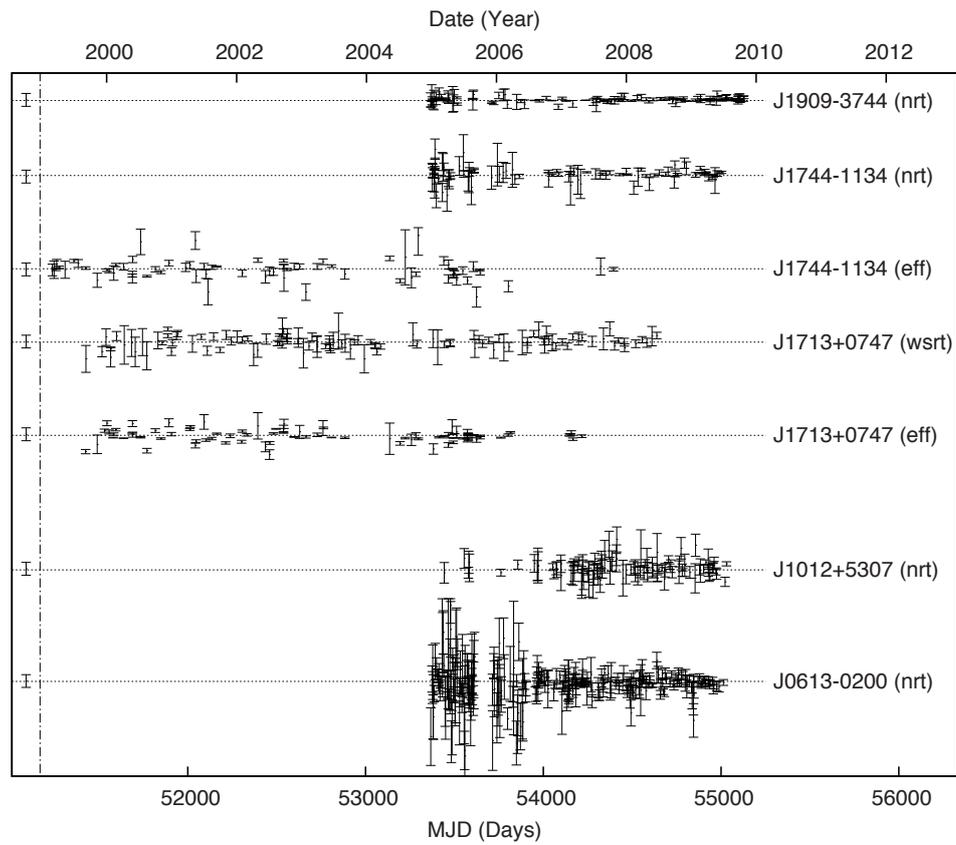
Figure 4.2: The timing residuals of all the pulsars used in the GWB limit calculation. The time in MJD is shown on the x-axis. On the left of the dash-dotted line we have placed a sample residual with an uncertainty of $1\,\mu$s.

and this is likely to hold for the data of other PTAs as well. The primary goal of the data analysis method presented in this chapter is therefore to place a limit on the GWB.

Simultaneously with the characterisation of the GWB signal, the PTA data analysis method should also provide information about the timing model and the timing noise. Because of the quality of the PTA data, the parameters of the timing model and the timing noise are best determined from the same data set that we are analysing to obtain information about the GWB signal. We discuss how this is done in the different PTA data analysis methods in section 4.4.3.

4.4.1 Detecting the GWB

As described in chapter 2, the premise of GWB detection is that the GWB induces well-defined correlations between the timing residuals of different pulsars. The correlations between timing residuals of two pulsars induced by the GWB depend only on the separation angle between those two pulsars. This correlation curve is known as the Hellings & Downs curve (see Figure 1.3). Similar to the GWB signal which is expected to be very "red" (i.e. low frequency), the unknown timing noise for each pulsar is expected to have a very red component. Timing noise is poorly understood, and currently we only have empirical models. Therefore, we cannot reliably distinguish between the GWB or timing noise using data of one pulsar. The only unique characteristic of the GWB is apparent when considering data of many pulsars: a signal with identical statistical properties in the data of all pulsars, correlated according to the Hellings & Downs curve.

All current PTA GWB detection schemes rely on the unique feature of the GWB signal that it induces correlations between timing residuals of all pulsars. Different approaches to the detection problem include:
1) Measuring the correlations between all pulsar pairs. If GWB signal is present in the data, the correlations should match the Hellings & Downs curve (Jenet et al., 2005). Also, by considering the cross-power spectrum, the amplitude of the GWB can be determined, given a value for the spectral index of the power-law GWB signal (Demorest, 2007; Yardley et al., 2011).
2) Searching for a signal in the data of all pulsars that is correlated according to the Hellings & Downs curve. Such a search takes the correlations induced by the GWB as a given, and measures the parameters of the signal correlated according to the Hellings & Downs curve (Anholm et al., 2009; van Haasteren et al., 2009, see also chapter 2 of this thesis).

In approach (1), one is primarily interested in the correlations that are induced by an unknown signal between all pulsars, whereas in approach (2) one is primarily interested in the parameters of the GWB signal. When attempting to present a re-

liable GWB detection, an analysis that takes the first approach should yield results of the GWB-induced correlations between all pulsar pairs that are accurate enough to exclude the possibility that these correlations are induced by any other source. An analysis that takes the second approach should include other correlated signals in the analysis, such that a reliable distinction can be made between a GWB signal that is correlated according to the Hellings & Downs curve, and other signals that are not correlated according to the Hellings & Downs curve. Because methods that take the first approach rely on cross-power measurements of pulsar pairs, these methods are not suited for placing limits on the GWB with data of only a few pulsars.

4.4.2 Placing limits

Methods that aim to place a limit on the GWB amplitude all try to answer the same question: assuming there is a GWB-induced signal present in the observations, what is the largest amplitude this signal can have that is still consistent with the data. Again, there are two approaches to this problem:

1) The incoherent analysis, where one solely focuses on the power spectrum of the timing residuals of the pulsars, neglecting the GWB-induced correlations between pulsar pairs (Stinebring et al., 1990; Kaspi et al., 1994; McHugh et al., 1996; Lommen, 2002; Jenet et al., 2006; Demorest, 2007).

2) The coherent analysis, where one not only includes the certain power spectrum of the GWB-induced signal in the timing residuals, but also the correlations between residuals of pulsar pairs. This is essentially what is done in this chapter. We note that it is also possible to place a coherent GWB-limit by only considering the cross-power correlations between pulsar pairs (Demorest, 2007; Yardley et al., 2011). However, in that case, the power-spectrum information of the single-pulsar time series is neglected, which makes these methods suboptimal.

In the case where only observations of a single pulsar are available, the incoherent analysis should be as sensitive as the coherent analysis. However, when observations of multiple pulsars is available, the most stringent limits are produced when doing a coherent analysis.

The difference between a coherent limit-placing analysis, and a GWB detection scheme, is that it is not necessary in the limit-placing analysis to convincingly show that the GWB signal is correlated according to the Hellings & Downs curve. The Hellings & Downs correlations are then assumed, and used to place a stringent limit. This is what is done in this chapter: in our model we have not included sources that are similar to the GWB, but correlated with a different angular dependence than the GWB. To this date, the Bayesian inference presented in this chapter is the only published method to coherently analyse PTA data with the aim to place

upper limits on the GWB without loss of information.

4.4.3   The timing model and timing noise

Both the timing model and the timing noise parameters should be determined from the same data as what is used to infer the parameters of the GWB. This has severe implications for the sensitivity to the GWB, since the timing model (especially the pulse period and pulse period derivative) and the timing noise parameters are covariant with the GWB. The way this is dealt with is different in Frequentist and Bayesian inference techniques.

Frequentist methods (e.g. Jenet et al., 2005; Yardley et al., 2011) to measure or place limits on the GWB parameters rely on estimates of the timing model and timing noise parameters. Typically, the best estimators for the timing model and timing noise parameters enter in the detection statistic for the GWB, and the uncertainty in the GWB parameters is determined by Monte Carlo simulations.

Bayesian methods like those presented in this chapter rely on marginalisation to deal with covariances. Instead of separately producing best estimators for the timing noise, the timing model, and the GWB parameters, all these contributions are present in the Bayesian posterior distribution. All parameters but those of the GWB are then marginalised over.

In Bayesian methods, the marginalisation process is computationally quite expensive. But once correctly carried out, the resulting marginalised posterior gives a very clear presentation of the analysis with all covariances correctly taken into account. In current frequentist methods, the point estimates for the GWB and the timing noise are relatively easily produced. But the Monte Carlo simulations required to take into account the covariances between the timing model, the timing noise, and the GWB parameters require a lot of computational resources, and optimality has yet to be proved (Yardley et al., 2011).

4.5   Bayesian PTA data analysis

The analysis presented in this chapter broadly follows the procedure introduced in van Haasteren et al. (2009, see also chapter 2). The Bayesian inference method of chapter 2 relies on creating the parametrised models of the timing residuals, and forming a probability distribution function (PDF) as a function of the model parameters. All known systematic contributions of known functional form should be included in the model. In the examples used in chapter 2 the model for the systematic errors included only the quadratic contribution to the TOAs from pulsar spindowns. The multi-telescope nature of the EPTA requires more complete mod-

els for timing residuals than the one used in chapter 2. In this section we show how to build and implement these models in practice.

We first briefly review the inference method of chapter 2 in Section 4.5.1 and 4.5.2. We then present the extended model we use for the analysis of the TOAs in Section 4.5.3, after which we show how we handle TOAs coming from different observatories in Section 4.5.4.

4.5.1 Brief review of the Bayesian inference method

The set of TOAs from all pulsars forms the basic input used in the Bayesian data analysis. Many processes influence the measured TOA values; in this chapter we discriminate between deterministic processes, like quadratic spindown, and stochastic processes, like timing noise:

$$t^{\text{obs}}_{(ai)} = t^{\text{det}}_{(ai)} + \delta t^{\text{stoch}}_{(ai)}, \tag{4.1}$$

where $t^{\text{obs}}_{(ai)}$ represents the i-th TOA of pulsar a, $t^{\text{det}}_{(ai)}$ is the corresponding contribution to the TOA solely due to deterministic processes, and $\delta t^{\text{stoch}}_{(ai)}$ is the contribution due to stochastic processes.

The effects of deterministic processes are described by the set of model parameters $\vec{\eta}$: $t^{\text{det}}_{(ai)} = t^{\text{det}}_{(ai)}(\vec{\eta})$. As is done in chapter 2, we assume that the stochastic processes are Gaussian, though their spectra are not necessarily white. In such a model, the stochastic processes can be represented by the correlation matrix

$$\langle \delta t^{\text{stoch}}_{(ai)} \delta t^{\text{stoch}}_{(bj)} \rangle = C_{(ai)(bj)} = C_{(ai)(bj)}(\vec{\xi}), \tag{4.2}$$

where $\vec{\xi}$ are the model parameters.

The key distribution used in a Bayesian analysis is the likelihood function, the probability distribution of the data for a given model and its parameters. As described in chapter 2, for PTAs the likelihood takes the following form:

$$
\begin{aligned}
L\left(\vec{\theta}\right) \;\; &= \;\; P\left(\vec{\delta t} \mid \vec{\theta}\right) = \frac{1}{\sqrt{(2\pi)^n \det C}} \\
&\exp\left[-\frac{1}{2} \sum_{(ai)(bj)} (\vec{t}^{\text{obs}}_{(ai)} - \vec{t}^{\text{fit}}_{(ai)}) C^{-1}_{(ai)(bj)} (\vec{t}^{\text{obs}}_{(bj)} - \vec{t}^{\text{fit}}_{(bj)})\right],
\end{aligned}
\tag{4.3}
$$

where $\vec{\theta} = (\vec{\eta}, \vec{\xi})$, and $\vec{\delta t}$ is the difference between the observed TOAs, and the fitted TOAs. A Bayesian analysis assigns prior distributions $P_0(\vec{\theta})$ to the model parameters, and explores the parameter space of the posterior distribution (short-

handed simply as *the posterior*): $P(\vec{\theta} \mid \vec{\delta t}) = L(\vec{\theta})P_0(\vec{\theta})$.

4.5.2 Obtaining a marginalised posterior distribution

The posterior $P(\vec{\theta} \mid \vec{\delta t})$ contains information about all model parameters. We need to express the posterior as a function of only those parameters that represent the GWB. This process is called marginalisation, and consists of integrating over all other parameters. The resulting marginalised posterior is the posterior probability density of the GWB parameters.

Marginalisation of a posterior in a high-dimensional parameter space is non-trivial, and a direct numerical integration is prohibitively computationally expensive. As in chapter 2, we employ a mix of analytic integration and Markov Chain Monte Carlo (MCMC) methods to accomplish this. The marginalisation remains the computational bottleneck for the method's effectiveness, as the computational time scales with n^3, with n the total number of TOAs to be analysed.

A computational shortcut can be used by analytically marginalising over the parameters of the timing model. As shown in chapter 2, this is possible provided that the parameters represent signals of known functional form. This condition is equivalent to the requirement that the timing residuals generated by the timing model are linear with respect to its parameters: $\delta t = d(\alpha - \hat{\alpha})$, where δt is the timing residual, d is a proportionality constant, $\hat{\alpha}$ is the best fit value for the model parameter, and α is the model parameter. While this is always true for quadratic spindown as considered explicitly in chapter 2, it is generally not true for other timing model parameters. However, when the deviations of the timing model parameters from their best-fit values are small, it is a good approximation that the residuals generated by the timing model are linear with respect to the deviations from their best-fit values: $\delta t \approx d(\alpha - \hat{\alpha})$.

Analytically marginalising over the timing model is therefore possible, and by doing so the number of parameters that must be integrated over numerically by the use of MCMC is reduced greatly. Dependent on the model we use to describe the statistics of the timing residuals, the number of parameters left to explore is then just several per pulsar/backend combination. The results of the analysis can be presented as a marginalised posterior as a function of any parameter in the model, provided that this parameter was present in the MCMC run.

4.5.3 Used model for the TOAs

We divide the actual parameterisation in 3 parts:
a) The deterministic timing model.
b) The gravitational-wave background.

c) Other stochastic processes (e.g.,timing noise).
In this section we discuss how we have taken these into account in our data analysis.

As a first step, the TOAs are processed using the software package TEMPO2, in order to determine the best-fit timing model. This procedure consists of the following steps:

1. TEMPO2 requires an initial guess α_{0i} for the timing model parameters α_i in order to find timing residuals (pre-fit timing residuals).

2. It then constructs an approximation to the timing model, in which the timing residuals depend linearly on $\alpha_i - \alpha_{0i}$.

3. It finds the best-fit α_i within this linear approximation, and uses those values to update the timing residuals using the full non linear timing model (post-fit timing residuals).

4. The newly obtained parameters and corresponding timing residuals are then judged by the person performing the model fitting, and if determined necessary the newly obtained parameters can act as the initial guess for a new fitting iteration. TEMPO2 also allows adjustment and fitting of α_i one by one.

Finding the timing solution with TEMPO2 is not fully algorithmic, but typically requires someone experienced with pulsar timing analysis, who approaches the TOAs fitting in several different ways, which ensures that phase coherence is maintained and that the relevant deterministic model parameters are included properly. Though this strategy works well in practice, we should remain conscious of the possibility that different solutions might be obtained by different observers, who may also choose to include additional model parameters. [2] In Appendix B of this chapter we present the timing solutions we found for the analysed pulsars. These are the values we used as our initial guess, α_{0i}. Note that these α_{0i} and their uncertainties, although created with TEMPO2 using the same datasets that we base our upper limit on, do not include our model for the red noise. The values and uncertainties we list in Appendix B of this chapter therefore do not represent our best estimates if we were to take into account the red timing noise. Although calculating these best estimates of α_i is reasonably straightforward, these estimates are not accessible in our MCMC because we have marginalised over these parameters analytically. The calculated upper limit on the GWB, however, does include all these effects, and therefore automatically incorporates the removal of power from the low-frequency GW signal by fitting for the timing model parameters and jumps.

In the above mentioned step 2 where the timing model is linearised, we have made an important simplification that we now describe in more detail. Since we

---

[2]Qualitatively, experienced observers are rightfully so very confident in their timing solutions. Quantitatively however, the only statistical tool currently available for observers to check whether the timing solution is reasonable is the reduced χ^2 statistic. But since the error bars obtained with the cross-correlation technique cannot be fully trusted, the same holds for the χ^2 statistic.

take into account, and marginalise over, all timing model parameters present in our method, we are effectively working with the TOAs instead of just the timing residuals. However, the timing model has been linearised by TEMPO2 with respect to $\alpha_i - \alpha_{0i}$. This implies that we need to be sufficiently close to α_{0i} in the parameter space for this approximation to be valid, which means that the timing residuals derived with TEMPO2 need to be approved by the person fitting the data, before using these as inputs in the Bayesian inference method.

The stochastic component contributing to the TOAs is characterised as follows. Firstly, general relativity describes how the timing residuals of a pair of pulsars are correlated due to gravitational waves:

$$\zeta_{ab} = \frac{3}{2} \frac{1 - \cos\theta_{ab}}{2} \ln\left(\frac{1 - \cos\theta_{ab}}{2}\right) - \frac{1}{4} \frac{1 - \cos\theta_{ab}}{2} + \frac{1}{2} + \frac{1}{2}\delta_{ab}, \qquad \boxed{4.4}$$

where θ_{ab} is the angle between pulsar a and pulsar b (Hellings & Downs, 1983). The GWB spectrum is parametrised as a power-law of the form (Maggiore, 2000; Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003; Sesana et al., 2008):

$$h_c = A\left(\frac{f}{\mathrm{yr}^{-1}}\right)^\alpha, \qquad \boxed{4.5}$$

were h_c is the characteristic strain as used in Jenet et al. (2006), A is the amplitude of the signal, and α is the spectral index. This then results in a correlation matrix for the GWB (see chapter 2):

$$C^{\mathrm{GW}}_{(ai)(bj)} = \frac{-A^2\zeta_{ab}}{(2\pi)^2 f_L^{2-2\alpha}} \left\{ \Gamma(-2 + 2\alpha)\cos(\pi\alpha)(f_L\tau)^{2-2\alpha} \right. $$
$$\left. \sum_{n=0}^{\infty} (-1)^n \frac{(f_L\tau)^{2n}}{(2n)!\,(2n + 2\alpha - 2)} \right\}, \qquad \boxed{4.6}$$

where, as in chapter 2, $\tau = 2\pi|t_i - t_j|$, and f_L is a cut-off frequency, set much lower than the lowest GW frequency we are sensitive to.

Secondly, the stochastic timing noise for each individual pulsar is split into three components:
1) Individual errors of TOA determination from the cross-correlation, represented by the TOA error bars. An extra free parameter, called the EFAC value, is commonly introduced by pulsar observers in order to account for possible mis-calibration of the radiometer noise (Hobbs & Edwards, 2006); this parameter is a multiplier for all of the TOA error bars for a given pulsar.
2) An extra white noise component, independent of the error bars. This basi-

cally acts as extra non-time–dependent noise, and the parameter is often called an EQUAD parameter.

3) Red noise, consisting of a power-law spectrum in the timing residuals. This component allows for structure in the timing residuals.

All three timing noise components are uncorrelated between the pulsars.

The resulting correlation matrices from components 1, 2, and 3, as derived in chapter 2, are given by:

$$
\begin{aligned}
C^{\mathrm{err}}_{(ai)(bj)} &= E_a^2 \Delta t^2_{(ai)} \delta_{ab} \delta_{ij} \\
C^{\mathrm{WN}}_{(ai)(bj)} &= N_a^2 \delta_{ab} \delta_{ij} \\
C^{\mathrm{RN}}_{(ai)(bj)} &= \frac{-R_a^2 \delta_{ab}}{(2\pi)^2 f_L^{2-2\alpha_a}} \left\{ \Gamma(-2 + 2\alpha_a) \cos(\pi \alpha_a) (f_L \tau)^{2-2\alpha_a} \right. \\
&\qquad \left. \sum_{n=0}^{\infty} (-1)^n \frac{(f_L \tau)^{2n}}{(2n)! \, (2n + 2\alpha - 2)} \right\},
\end{aligned}
\tag{4.7}
$$

where $C^{\mathrm{err}}_{(ai)(bj)}$, $C^{\mathrm{WN}}_{(ai)(bj)}$, and $C^{\mathrm{RN}}_{(ai)(bj)}$ are the correlation matrices corresponding to the error bars, the extra white noise, and the red noise respectively, with a and b denoting the pulsar number, i and j denote the observation number, Δt is the TOA uncertainty (the error bar) as calculated in the pipeline, E_a is the scaling parameter of the error bars for the a'th pulsar (the EFAC factor), N_a is the amplitude of the white noise, R_a is the amplitude of the red timing noise, α_a is the spectral index of the red noise spectrum of pulsar a, and τ is the time difference between two observations.

4.5.4 Combining datasets

The previous section gives a complete description of the model we use to analyse the TOAs of a single pulsar, observed with one telescope. That model does not yet account for the use of different observatories. In this section we explain what we do to accomplish this.

As discussed in Section 4.2, the reduced data products are (sometimes subtlety) influenced by many different components of the reduction process. In order to account for slight offsets between TOAs, introduced by using slightly different reduction procedures on individual datasets, a calibration term needs to be introduced when merging TOAs from different observing systems. This extra calibration term takes the form of a "jump", an arbitrary phase offset between datasets, which is fit for simultaneously with other timing model parameters. We use the term dataset for any series of TOAs that can be analysed without a jump. In practice this is any series of TOAs, of the same pulsar, observed with the same hardware elements,

and processed with the same algorithms, at the same observing frequency. Here we combine 7 such datasets (those shown in Figure 4.2).

Jumps have been used routinely when combining data of different observatories and/or data recorders (e.g., Janssen, 2009). This allows us to find a single solution for the timing model of a pulsar timed by multiple observatories. However, the TOAs produced by pipelines at different observatories may have different statistical properties. In order to account for this, we allow the stochastic contributions in our model discussed in Section 4.5.3 to vary between datasets:

1) One timing model per pulsar (taken directly from TEMPO2)
2) Jumps between different datasets
3) A scaling factor for the error bars (EFAC) for each dataset
4) An extra white noise component (EQUAD) for each dataset
5) Power law red noise for each dataset

A major advantage of this approach is that it allows one to detect statistical differences between observatories that may be introduced by different algorithms/components, and then use this feedback to iteratively improve our datasets.

The analysis of the TOAs consists of two steps. In the first step TEMPO2 is used to find the timing solution for a single pulsar. This includes possible jumps between datasets. Once the timing solution is obtained, the results are passed on to the Bayesian inference method. The Bayesian method then analytically marginalises all parameters of the timing model, including jumps, while using MCMC to explore the rest of the parameter space.

4.6 Results

Now that we have developed the necessary framework to analyse the TOAs, we apply the Bayesian inference method to the observations. In the following subsections we explain in detail how we selected the five pulsars that we already mentioned in Section 4.3.2, and we present the GWB upper limit we are able to calculate using observations of those pulsars.

4.6.1 Selecting the most constraining datasets

For any pulsar, obtaining the timing solution and timing residuals is the first step after obtaining the TOAs. The timing residuals of the pulsars used in this chapter are shown in Figure 4.2, and the parameters of the timing model are shown in Appendix B of this chapter. The timing model also includes several jumps as some of these pulsars have been observed with several European telescopes. The timing solutions we find are quite consistent with the values already published in the literature. Given that we are solving for 56 parameters, it is to be expected that one

or two parameters deviate at the 2-σ level. The only unexpected outlier we find is the proper motion in right ascension of J1713+0747, which deviates from Splaver et al. (2005) by over 5-σ. Given that we are combining data of several telescopes, and that we do not take into account our red noise models in listing these timing solutions, we postpone exploring this difference to future work where the focus lies on investigating the statistics of the timing model parameters in the presence of red noise. Such an investigation is beyond the scope of this manuscript.

With the model of the systematic contributions in place, we first perform the analysis separately for each of the datasets and obtain the posterior probability distribution for their intrinsic noise parameters, specified in Equation (4.7) of the previous section. Note that at this stage of the analysis the contribution from a GWB is not yet included. We determine a marginalised posterior for each pulsar as a function of the following parameter combinations:
1) EFAC vs. EQUAD
2) Red noise amplitude vs. red noise spectral index
In both cases, the posterior is marginalised over all parameters but two, and the resulting 2-dimensional distribution is displayed as contours at the 1-, 2-, and 3-σ level (the regions where respectively 68%, 95%, and 99.7% of the volume of the posterior is enclosed).

As an example we consider the TOAs of pulsar J1713+0747, which consist of data taken with Effelsberg and Westerbork. Here we focus on the marginalised posterior distributions that represent information about the Effelsberg TOAs; these distributions and the residuals are shown in Figs 4.3 and 4.4. A traditional non-Bayesian analysis of the Effelsberg TOAs consists of a fit to the timing model with Tempo2, which yields the best-fit parameters, the corresponding uncertainties, and a reduced χ^2 statistic. The reduced χ^2 is defined as:

$$\chi^2 = \frac{1}{n-m} \sum_{i=1}^{n} \frac{\left(t_i^{\mathrm{obs}} - t_i^{fit}\right)^2}{\epsilon^2 \sigma_i^2}, \qquad \boxed{4.8}$$

where n is the number of observations, m is the number of free parameters in the least-squares fit, t_i^{obs} is the observed TOA, t_i^{fit} is best-fit value of the TOA, σ_i is the TOA uncertainty of t_i^{obs}, and ϵ is the EFAC value. It is usual practice to set the EFAC such that the reduced $\chi^2 = 1$, which is accomplished by: $\epsilon = \sqrt{\chi^2(\epsilon = 1)}$. For the J1713+0747 Effelsberg TOAs, we have $\chi^2(\epsilon = 1) = 18.9$ and therefore $\epsilon = 4.35$.

As can be seen from Figure 4.4, a non-zero red noise component is required to describe the TOAs. The EQUAD parameter is consistent with 0-amplitude according to Figure 4.3, while the EFAC parameter is significantly lower than what

a TEMPO2 $\sqrt{\chi^2}$ estimate would give. This tells us that no separate white-noise component is required to describe the TOAs: all the uncorrelated scatter can be assigned to the error bars on the TOAs. It is also of interest that in this case the EFAC parameter is much smaller, and indeed much closer to 1, than the more traditional estimator $\sqrt{\chi^2}$. The data is reasonably well-modelled by just the presence of red noise.

It is also worth noting that, due to practicalities having to do with hardware changes at the observatories, the TOAs of J1713+0747 end at an earlier epoch than the other 4 pulsars. Although in the future the inclusion of this data will obviously benefit the sensitivity to the GWB, we note that the GWB limit is not negatively effected by this lack of overlap of the TOAs between pulsars.

The analysis of the TOAs of the other pulsars yields similar, but slightly different results. As can be seen in Appendix B of this chapter, some pulsars do have non-negligible white noise, and some do appear to have EFAC values significantly different from 1. As of yet we do not have a complete explanation for the exact form of the marginalised posteriors.

We present the marginalised posterior as a function of the red noise parameters in an intuitive way: as pointed out in Section 4.3.2 we use the same units for the red noise amplitude and red noise spectral index as we use for the GWB parameters. For the analysis of TOAs of just one pulsar, the red noise can now be thought of as if it was generated solely by a GWB with a certain amplitude and spectral index. In this case, the marginalised posterior for the red noise parameters shows us what upper limit we are able to place on the GWB amplitude as a function of spectral index.

We choose a 3-σ threshold of $R_a \leq 10^{-13}$ at a spectral index of $\alpha_a = -2/3$. Based on the marginalised posteriors of all the EPTA pulsars, we can decide whether a particular dataset can put a constraint on the GWB lower than this or not. Using this threshold we include five pulsars in our final analysis. These five significantly outperform the other pulsars in terms of how well they can limit the GWB amplitude, and we do not expect to gain any significant sensitivity by including more pulsars in our current archival data sets. The residuals of the pulsars we use in our combined analysis are shown in Figure 4.2. More datasets will be added after some extensive and detailed recalibration procedure of existing datasets.

4.6.2 GWB upper limit

Now that we have selected our pulsars that can significantly contribute to a GWB limit, we are in the position to infer the amplitude and spectral index of the GWB. Our model of the combined data of the five pulsars we selected in Section 4.6.1 consists of all sources we included in the analysis for the single pulsars, and an
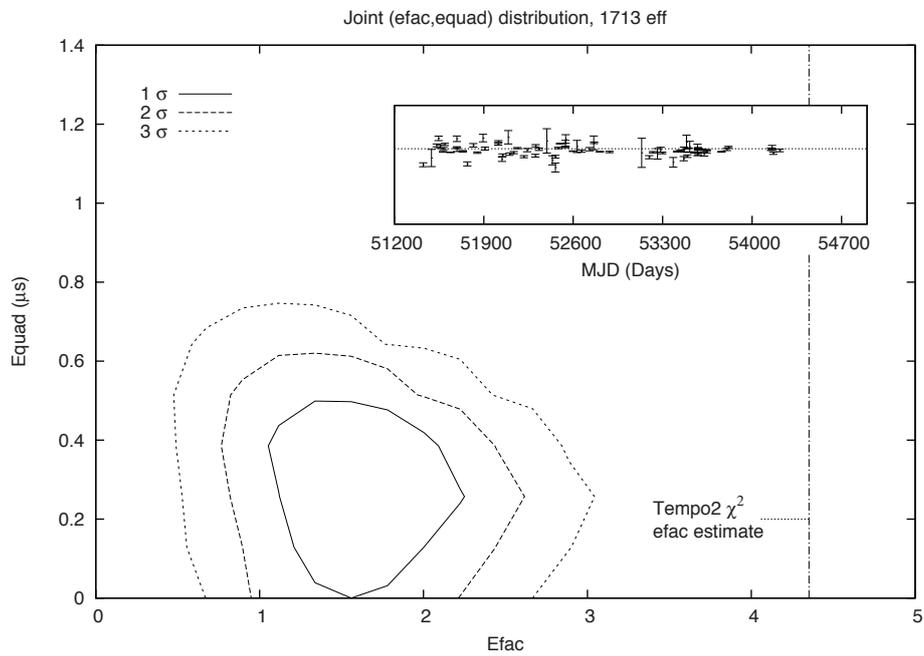
Figure 4.3: The marginalised posterior of J1713+0747 (Effelsberg), as a function of the EFAC and EQUAD parameters. The contours are at the 1, 2, and 3-σ level, indicating a respective volume inside that region of 68%, 95%, and 99.7%.
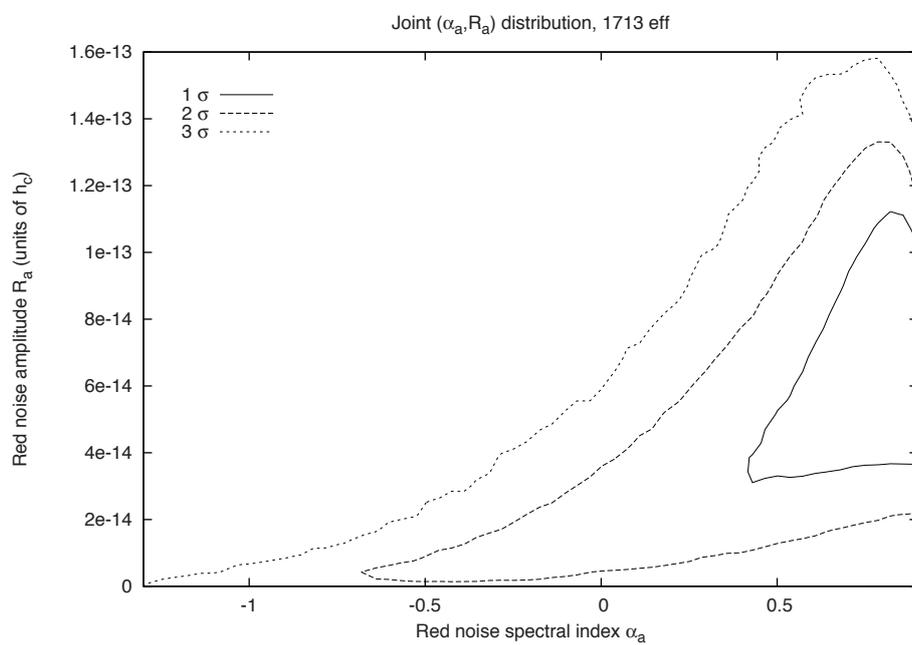
Figure 4.4: The marginalised posterior of J1713+0747 (Effelsberg), as a function of the power-law red noise parameters: the amplitude and the spectral index. The contours are at the 1, 2, and 3-σ level, indicating a respective volume inside that region of 68%, 95%, and 99.7%.
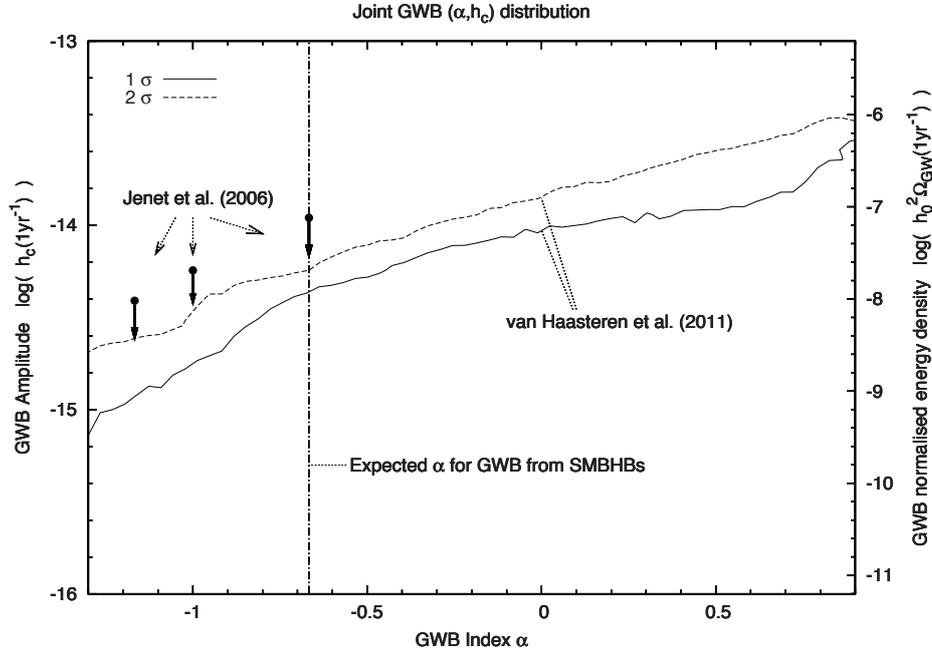
Figure 4.5: The marginalised posterior distribution as a function of the GWB amplitude, and spectral index. The contours marked by 'van Haasteren et al. (2011)' are the results of this chapter at the 1-σ and 2-σ level, indicating a respective volume inside that region of 68%, and 95%. The vertical dash-dotted line at $\alpha = -2/3$ shows where we expect a GWB generated by supermassive black-hole binaries. The most recent published limits are shown as the three upper limit arrows pointing down, marked by 'Jenet et al. (2006)'.

extra source that corresponds to the GWB. As discussed in Section 4.5.3, the GWB source is a power-law correlated between pulsars as described by Equation (4.4).

As before, we use MCMC to sample the posterior distribution while analytically marginalising over the timing model; now the analytic marginalisation happens simultaneously for the timing models of the five pulsars. In Figure 4.5 we present the posterior, marginalised over all parameters except the GWB amplitude and spectral index. In the same figure we also show the PPTA published values of the GWB limit (Jenet et al., 2006). For the expected spectral index for a GWB generated by a large number of supermassive black-hole binaries, $\alpha = -2/3$, we find a 95% confidence GWB upper limit of $h_c(1\text{yr}) \leq 6 \times 10^{-15}$. This is smaller by a factor of 1.8 than the previously published PPTA limit.

As a cross-check with other codes, and to verify that we are definitely sensitive

to the level of the limit we have calculated here, we perform an additional test. We use the TEMPO2 plug-in GWbkgrd (Hobbs et al., 2009) to generate simulated timing residuals as produced by a GWB with an amplitude of $h_c(1\text{yr})$. We then create a new set of TOAs, consisting of the values of the simulated timing residuals added to the values of the observed TOAs of the five pulsars that we have analysed in this section. We then redo the whole analysis. Current PTAs aim to reach sensitivities in the order of $h_c(1\text{yr}) = 10^{-15}$ in the future (Jenet et al., 2005), which is over five times more sensitive than the limit we achieve here. In the case that the GWB just happens to be at the 2-σ level of our current limit, we demonstrate what such a fivefold increase in sensitivity could do for our ability to measure the GWB parameters by adding a signal of $h_c(1\text{yr}) = 30 \times 10^{-15}$ to our current TOAs. The result is shown in Figure 4.6. We find that the results are consistent with the input parameters of the simulated GWB, and that we can reliably detect a GWB in this case[3]. The values of the GWB parameters we have used to simulate the GWB lie within the 1-σ credible region of Figure 4.6.

4.7 Implications and outlook

The analysis performed in this chapter puts an upper limit on a GWB with a power-law characteristic strain spectrum $h_c = A(f/\text{yr}^{-1})^\alpha$. In the literature, upper limits are typically quoted for various values of α, where the considered α depends on the physics responsible for generation of the GWB. A useful feature of our approach is that we are able to measure α for a strong enough GWB (see chapter 2 for a discussion). The extra degree of freedom in our model, α, necessarily changes the interpretation of the posterior to some extent. We interpret the 2-σ contour in our plot of the marginalised posterior as the upper limit on the GWB as a function of α. Fixing α and re-evaluating the 2-σ limit based on the posterior for A only does not significantly alter our results.

In this section, we briefly discuss the implications of the new upper limits in the context of two different mechanisms for generation of the GWB: binaries of supermassive black holes, and cosmic strings. We also place the obtained limit in a context with respect to other PTA projects, and we discuss how we expect this limit to evolve in the near future.

---

[3]We note that, although such a detection is consistent with a GWB, we would need more pulsars to exclude the possibility that some other effect is causing the correlated signal we detect here.
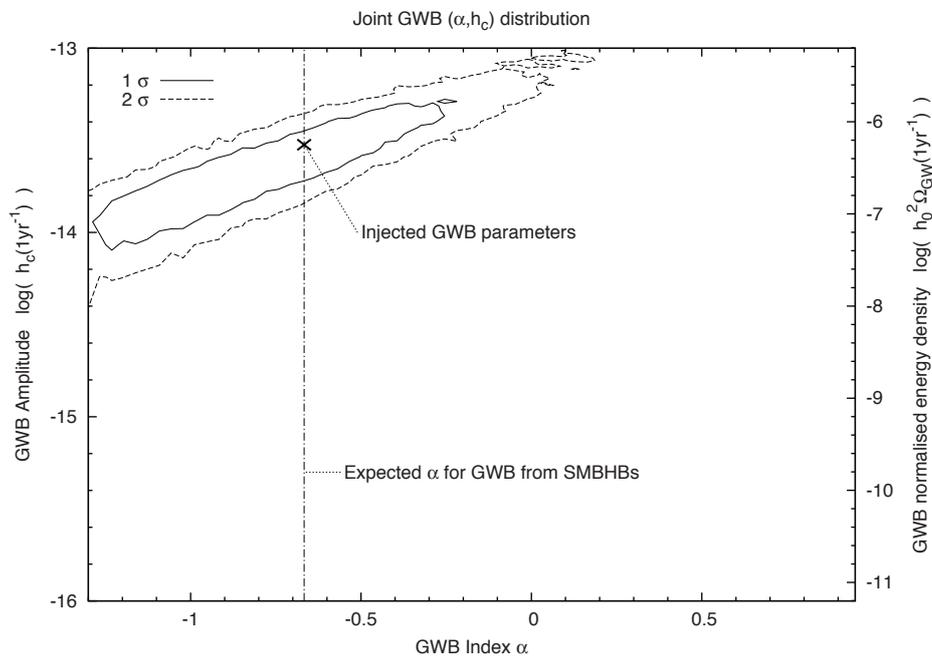
Figure 4.6: Same marginalised posterior distribution as in 4.5, but here we have injected the residuals of a simulated GWB with amplitude $h_c(1\mathrm{yr}) = 30 \times 10^{-15}$ in the data.

4.7.1 Supermassive black hole binaries

Several authors discuss the characteristic strain spectrum generated by an ensemble of supermassive black holes (SMBHBs) distributed throughout the Universe (Begelman et al., 1980; Phinney, 2001; Jaffe & Backer, 2003; Wyithe & Loeb, 2003). They show that the characteristic strain spectrum generated by such black hole binaries can well be approximated by a power-law:

$$h_c = h_{1\mathrm{yr}}\left(\frac{f}{\mathrm{yr}^{-1}}\right)^{-2/3}, \qquad (4.9)$$

where $h_{1\mathrm{yr}}$ is a model-dependent constant. Though the form of the characteristic strain, the power-law, is quite general among the different SMBHB assembly models the authors use in their work, the parameterisations and assumptions about other physical quantities differ between all investigators. The predicted $h_{1\mathrm{yr}}$ therefore differs depending on what SMBHB assembly scenarios were assumed.

Recently, Sesana et al. (2008) have extensively investigated a wide variety of assembly scenarios, including those considered in Jenet et al. (2006). For our purposes in this chapter, their most important result is an estimate of $h_{1\mathrm{yr}}$ for all models[4]. In calculating this value, they take into account the uncertainties of the key model parameters for different scenarios, and come up with $h_{1\mathrm{yr}} \approx 2\times 10^{-16} - 4\times 10^{-15}$. We are less than a factor of two away from this range, so we foresee that we can start to rule out some models in the near future.

Two more results of Sesana et al. (2008) are interesting with respect to the limit presented in this chapter. The first is that the frequency dependence of the GWB is expected to be steeper than a power-law $\propto f^{-2/3}$ for frequencies $f \gtrsim 10^{-8}$ Hz. The steepness depends on the chosen model. We have incorporated a varying spectral index α in our current analysis, and since we are not yet able to detect the GWB, we postpone a more thorough investigation of the exact dependence of h_c on f to later work with even better datasets. The second interesting result is that in the frequency range of 10^{-8} Hz $\leq f \leq 10^{-7}$ Hz, the GWB might be dominated by single sources. In that case, a search for just a certain characteristic strain spectrum is not appropriate, and we note that further investigation is required in this regard.

4.7.2 Cosmic strings

Several authors have suggested that oscillating cosmic string loops will produce gravitational waves (Vilenkin, 1981; Damour & Vilenkin, 2005; Ölmez et al.,

---

[4]The model for the GWB that Sesana et al. (2008) use is a broken power-law. Their $h_{1\mathrm{yr}}$ therefore has a slightly different meaning, and our quoted value should be taken as a crude approximation.

2010). Damour & Vilenkin (2005) have used a semi-analytical approach to derive the characteristic strain h_c of the GWB generated by cosmic strings:

$$
\begin{aligned}
h_c(f) \quad = \quad & 1.6 \times 10^{-14} c^{1/2} p^{-1/2} \epsilon_{\text{eff}}^{-1/6} \\
& \times (h/0.65)^{7/6} \left(\frac{G\mu}{10^{-6}} \right)^{1/3} \left(\frac{f}{\text{yr}^{-1}} \right)^{-7/6} ,
\end{aligned} \tag{4.10}
$$

where μ is the string tension, G is Newton's constant, c is the average number of cusps per loop oscillation, p is the reconnection probability, ϵ_{eff} is the loop length scale factor, and h is the Hubble constant in units of $100 \text{km s}^{-1} \text{Mpc}^{-1}$. Usually, the dimensionless combination $G\mu$ is used to characterise the string tension. Theoretical predictions of string tensions are $10^{-11} \leq G\mu \leq 10^{-6}$ (Damour & Vilenkin, 2005).

From the above expression for the characteristic strain generated by cosmic strings, we see that this is again a power-law, but now with $\alpha = -7/6$. Using a standard model assumption that $c = 1$, the facts that p and ϵ_{eff} are less than one, and that h is expected to be greater than 0.65, we can safely use our derived upper limit on h_c for $\alpha = -7/6$ to limit the string tension: $G\mu \leq 4.0 \times 10^{-9}$. This already places interesting constraints on the theoretical models, and in a few years the EPTA will be able to place much tighter restrictions in the case of a non-detection of a GWB: with only a factor of five decrease of the upper limit, we would be able to completely exclude the $10^{-11} \leq G\mu \leq 10^{-6}$ range mentioned in Damour & Vilenkin (2005).

4.7.3 Comparison with other PTA projects, and prospects

The EPTA data we have analysed in this chapter consists of only a small subset of pulsars from the complete ensemble of pulsars that is observed by the EPTA. The observatories of the EPTA have been continuously upgrading their systems and their data processing pipelines. The limit on the GWB produced by EPTA data will therefore improve substantially in the near future. But even though the limit $h_c(1\text{yr}) \leq 6 \times 10^{-15}$ we present here is significantly lower than the limit published by Jenet et al. (2006), it is worth placing this value into context.

Because of the redness of the GWB signal, the rms of the timing residuals induced by the GWB increases sharply with the duration of the PTA experiment. For an spectral index $\alpha = -2/3$, the rms scales as $\sigma^{GWB} \propto T^{5/3}$, where σ^{GWB} is the rms of the GWB-induced timing residuals, and T is the duration of the experiment. Since PTA observations have been rapidly improving over the past several years, we conclude that the Jenet et al. (2006) results are not representative for the PPTA data quality. Also the greater timespan of high-quality data makes the PPTA much

more sensitive to the GWB now than it was in 2006.

The sensitivity to the GWB should be compared between the three projects PPTA, NANOGrav, and EPTA, by performing the same analysis on their respective datasets: there may still be noise contributions to the TOAs that can be mitigated, and the effect on GWB sensitivity due to the differences between data reduction pipelines is not completely understood. We note here that all three projects have their strengths. The PPTA was the first to organise itself and has been timing a full array of 20 pulsars regularly for over five years, and it has access to the southern hemisphere pulsars. NANOGrav has the largest telescopes at their disposal, and the EPTA has many telescopes which, next to their regular observing schedule, will be coherently combined in the near future (Ferdman et al., 2010).

Most likely, the first detection of GWs by a PTA will occur as a result of a joint IPTA effort. We use the graphs of chapter 2 to create rule of thumb that sensitivity of a PTA to the GWB with a fixed T scales as $\sqrt{n}$, with n the total number of TOAs. Because the GWB signal increases in strength with time as $\sigma^{GWB} \propto T^{5/3}$, we can be optimistic about prospects for detection of the GWB by PTAs. Given that we have only used five pulsars, most of which were observed for five years, and that there are three currently organised PTAs, we can hope to reach a sensitivity to the GWB by the IPTA in five years an order of magnitude greater than we have accomplished in this chapter. Although this prediction is probably optimistic due to our selection of best pulsars in this chapter, and because it neglects the mostly unknown level of red timing noise which is expected to be present in all millisecond pulsars, we conclude that the sensitivity of PTAs to the GWB will greatly increase in the near future.

4.8 Conclusion and discussion

In this chapter we have developed the methodology on how to handle combined PTA datasets of several telescopes and how to robustly calculate a corresponding upper limit on the GWB. Our Bayesian approach has handled in a straightforward way different data sets of varying duration, regularity, and quality. The current upper limit on the GWB, calculated with EPTA data, is $h_c \leq 6 \times 10^{-15}$ in the case of $\alpha = -2/3$, as predicted for a GWB created by an ensemble of supermassive BH binaries. More generally, the analysis has resulted in a marginalised posterior as a function of the parameters of the GWB: the GWB amplitude and the spectral index.

Due to hardware and software upgrades at the EPTA observatories, and due to the ever increasing time baseline of the data, we expect the sensitivity to increase greatly over the next few years. Especially the combination of the EPTA data sets

with the data of the other PTA projects seems promising.

The raw telescope data must first undergo careful reduction and modelling before it can be analysed by the Bayesian inference method. We have provided some discussion of these processes and have motivated our choice of model for the TOAs. As part of our analysis, we have studied the probability distribution of the pulsar noise parameters, and highlighted the crucial importance of precise characterisation of the red component of pulsar timing noise.

Appendix B

Here we show the timing solutions of all datasets used in this chapter, combined with the posterior distributions for the timing noise.

Pulsar name	J0613−0200	J1012+5307	J1713+0747
Fit and data set			
Telescopes used	NRT	NRT	EFF & WSRT
MJD range	53367 - 55012	53443 - 55030	51426 - 54637
Number of TOAs	280	107	195
Rms timing residual (ns)	912	769	396
Reduced χ^2 value	1.00	1.00	1.13
Epoch	54189	54236	53031
Measured Quantities			
Right ascension, α (J2000)	06:13:43.97385(4)	10:12:33.43241(10)	17:13:49.530782(3)
Declination, δ (J2000)	−02:00:47.0720(12)	+53:07:02.665(2)	+07:47:37.52343(8)
Pulse freq., ν (s^{-1})	326.600562095168(13)	190.26783448248(14)	218.811840486637(30)
Derivative of pulse freq., $\dot{\nu}$ (s^{-2}) . .	−1.02281(3)×10^{-15}	−6.1998(4)×10^{-16}	−4.0836(2)×10^{-16}
PM in RA, μ_α (mas yr^{-1})	1.90(4)	3.17(7)	5.017(12)
PM in DEC, μ_δ (mas yr^{-1})	−10.31(9)	−24.96(9)	−3.96(3)
Parallax, π (mas)	—	—	0.915(7)
Dispersion measure, DM (cm^{-3}pc)	38.77700	9.0176	15.9907
Binary model	DD	ELL1	DD
Orbital period, P_b (d)	1.19851257534(5)	0.60462272322(4)	67.8253309255(20)
Derivative of orbital period, $\dot{P}_b$ . . .	—	—	—
Epoch of periastron, T_0 (MJD) . . .	54189.019(6)	—	53014.9592(7)
Projected sm. axis of orbit, x (lt-s)	1.09144417(8)	0.58181742(13)	32.34242015(7)
Longitude of periastron, ω_0 (deg) .	47.1(1.6)	—	176.2109(12)
Orbital eccentricity, e	5.47(15) ×10^{-6}	—	7.49312(13) ×10^{-5}
Time of ascending node (MJD) . . .	—	54236.2078302(3)	—
EPS1 (ϵ_1), $e \sin \omega$	—	1.18(5)×10^{-5}	—
EPS2 (ϵ_2), $e \cos \omega$	—	2.20(5)×10^{-5}	—
Sine of inclination angle, $\sin i$	—	—	—
Companion mass, M_c ($M_\odot$)	—	—	—
Assumptions			
Clock correction procedure		TT(TAI)	
Solar system ephemeris model . . .		DE405	

Table 4.2: The timing solutions for the pulsars used in this chapter before applying the Bayesian inference method. These solutions are determined using Tempo2, which uses the International Celestial Reference System and Barycentric Coordinate Time. As a result this timing model must be modified before being used with an observing system that inputs Tempo format parameters. See Hobbs et al. (2006) for more information. Note that the figures in parentheses are the nominal 1-σ Tempo2 uncertainties, with EFACs included, and therefore do not include the red noise model. In the GWB limit calculation these respective parameters are marginalised over. Also, the dispersion measure quoted here results from combining these observations with EPTA data of other frequencies. These DM values are used in the dedispersion, but we didn't include all observations in our GWB analysis. We therefore have not fit for the DM here, and an error estimate cannot be

Pulsar name	J1744−1134	J1909−3744
Fit and data set		
Telescopes used	EFF & NRT	NRT
MJD range	51239 - 55001	53366 - 55127
Number of TOAs	159	113
Rms timing residual (ns)	444	134
Reduced χ^2 value	1.05	1.00
Epoch	53120	54247
Measured Quantities		
Right ascension, α (J2000)	17:44:29.391592(7)	19:09:47.437982(5)
Declination, δ (J2000)	−11:34:54.5762(6)	−37:44:14.3176(2)
Pulse freq., ν (s^{-1})	245.426119777227(4)	339.31568732355(1)
Derivative of pulse freq., $\dot{\nu}$ (s^{-2}) . .	$-5.3817(4)\times10^{-16}$	$-1.614853(8)\times10^{-15}$
PM in RA, μ_α (mas yr^{-1})	18.817(10)	−9.490(11)
PM in DEC, μ_δ (mas yr^{-1})	−9.30(6)	−35.89(4)
Parallax, π (mas)	2.602(10)	1.01(7)
Dispersion measure, DM (cm^{-3}pc)	3.13632	10.37877
Binary model	—	ELL1
Orbital period, P_b (d)	—	1.53349947490(6)
Derivative of orbital period, $\dot{P}_b$ . . .	—	$3.5(5)\times10^{-13}$
Epoch of periastron, T_0 (MJD) . . .	—	—
Projected sm. axis of orbit, x (lt-s)	—	1.89799108(11)
Longitude of periastron, ω_0 (deg) .	—	—
Orbital eccentricity, e	—	—
Time of ascending node (MJD) . . .	—	54247.169903748(15)
EPS1 (ϵ_1), $e \sin\omega$	—	$6.4(5.5)\times10^{-8}$
EPS2 (ϵ_2), $e \cos\omega$	—	$-3(3)\times10^{-8}$
Sine of inclination angle, $\sin i$	—	0.9980(3)
Companion mass, M_c ($M_\odot$)	—	0.208(7)
Assumptions		
Clock correction procedure	TT(TAI)	
Solar system ephemeris model . . .	DE405	

Table 4.3: Same as table 4.2.
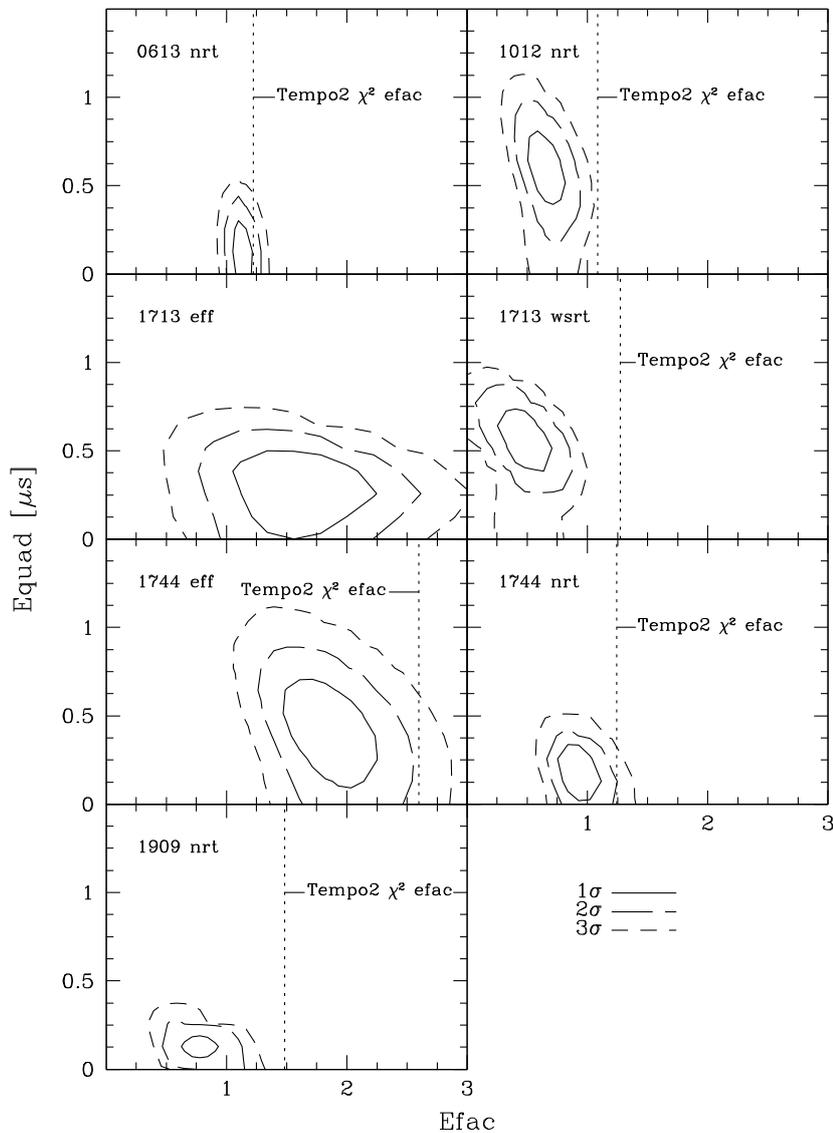
Figure 4.7: The marginalised posteriors of all datasets, as a function of the EFAC and EQUAD parameters. The contours are at the 1, 2, and 3-σ level, indicating a respective volume inside that region of 68%, 95%, and 99.7%. For the J1713-0747 posterior, the TEMPO2 χ^2 estimate is not shown because it has the off-scale value of 4.4.
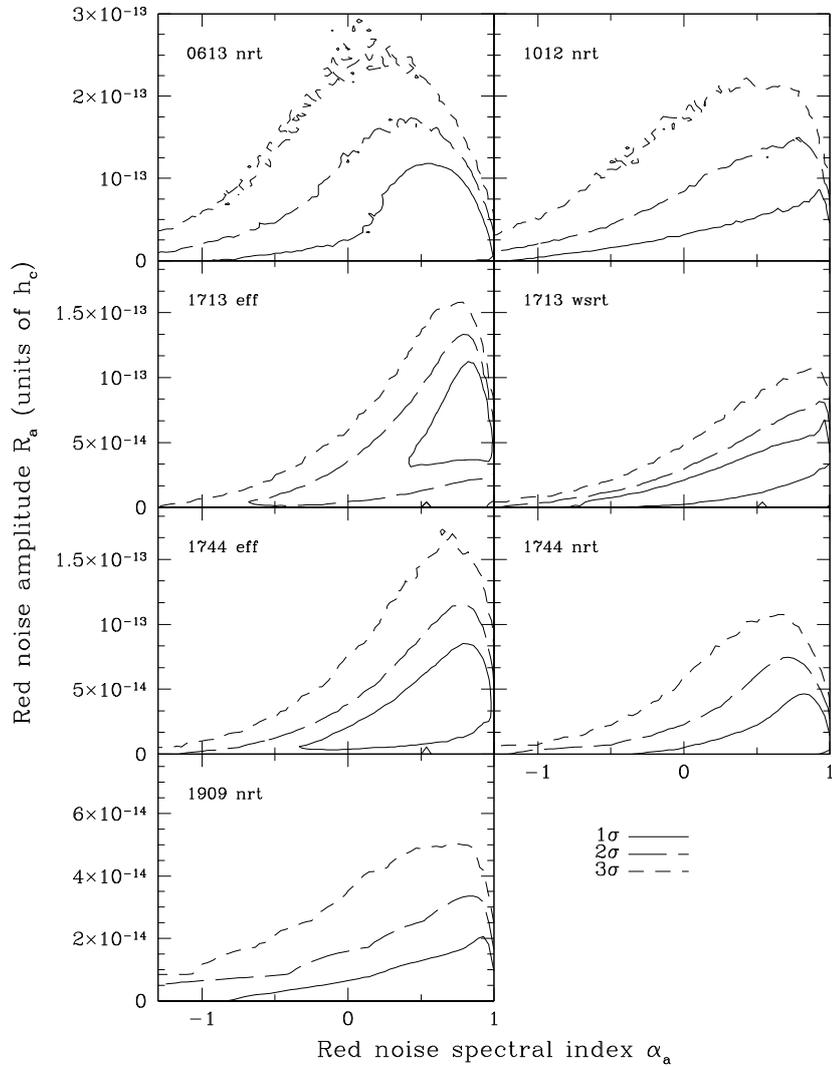
Figure 4.8: The marginalised posterior of all datasets, as a function of the power-law red noise parameters: the amplitude and the spectral index. The contours are at the 1, 2, and 3-σ level, indicating a respective volume inside that region of 68%, 95%, and 99.7%. The more negative the value of α, the steeper the power-law spectrum, with the spectrum approaching a white spectrum at the right of the plot. We also note that the amplitude of the red noise cannot be trivially scaled linearly to an rms value of the timing residuals.

5

Marginal likelihood calculation with MCMC methods

The most important kind of freedom is to be what you really are. [...] There can't be any large-scale revolution until there's a personal revolution, on an individual level. It's got to happen inside first.

<div align="right">Jim Morrison</div>

Abstract

Markov Chain Monte Carlo (MCMC) methods have revolutionised Bayesian data analysis over the years by making the direct computation of posterior probability densities feasible on modern workstations. However, the calculation of the prior predictive, the marginal likelihood, has proved to be notoriously difficult with standard techniques. In this chapter a method is presented that lets one calculate the marginal likelihood using nothing but the results from standard MCMC algorithms, like Metropolis-Hastings. This new method is compared to other methods like nested sampling, and outperforms the latter in some cases. One of the toy problems considered in this chapter is the analysis of mock pulsar timing data, as encountered in pulsar timing array projects. This method is expected to be useful as well in other problems in astrophysics, cosmology and particle physics.

<div align="right">To be submitted to MNRAS</div>

5.1   Introduction

Bayesian inference has proved over the years to be a very powerful tool in many branches of science as it gives a very clear prescription of how to analyse datasets without loss of information, even for very complicated models. On the practical side, in performing Bayesian analysis two difficult problems often emerge:

CHAPTER 5. MARGINAL LIKELIHOOD CALCULATION WITH MCMC METHODS

1. Producing the posterior probability density functions (PDFs) for several interesting parameters requires marginalisation, i.e. the integration of the full joint posterior probability distribution function (PDF) over most model parameters. In a majority of cases this must be done numerically, and it is common to employ a Markov Chain Monte-Carlo (MCMC) algorithm for performing the integration.
2. Model selection requires the calculation of the Bayes factor: the ratio of the marginal likelihood (ML) of two competing models. This marginal likelihood, sometimes also called the evidence, is the normalisation constant required to have the likelihood times the prior PDF (when normalised called the posterior PDF) integrate to unity when integrating over all parameters. The calculation of this value can be notoriously difficult using standard techniques.

The use of Markov Chain Monte Carlo (MCMC) algorithms, such as the Metropolis-Hastings algorithm, has become extremely popular in the calculation of the marginal posterior PDFs. MCMC algorithms allow one to sample from posterior distribution of complicated statistical models, greatly reducing the effort involved in evaluating numerical integrals. MCMC algorithms typically sample from a posterior distribution in a way that does not require explicit normalization. This is a virtue when only one model is of interest and only parameter estimation must be pursued. But it means that most MCMC algorithms do not directly address problem (2).

Over the years, several methods capable of calculating the ML have been developed (including Newton & Raftery, 1994; Earl & Deem, 2005; Skilling, 2004; Feroz et al., 2009). Typically, the technical challenges of producing the posterior PDFs, and the ML, that these methods try to overcome can be divided in three stages:
1. Finding the high probability density (HPD) regions of the posterior in parameter space.
2. Maximising the posterior to find the best estimators.
3. Sampling the HPD regions as accurately as possible.
It is very unlikely that one algorithm will outperform all others in all three aspects, and complicated problems may require a combination of algorithms to be efficient. Results of ML calculations can rarely be checked for sanity without additional major effort, usually involving running the same or a different algorithm for the same problem to verify a result. Especially for integrals in very high dimensional spaces, which are particularly challenging, crosschecks between algorithms may be desired.

In this chapter we present a method to calculate the ML from the MCMC chains of regular MCMC algorithms. This method can be applied to chains that have already been run, provided that the values of the likelihood times the prior have been saved together with the values of the parameters at each point in the chain.

By using MCMC chains, the method is especially efficient in stage (3), sampling the HPD region accurately if found correctly. In problems where sampling the HPD region efficiently is the greatest challenge, this new method could be of particular interest. The error on the ML can be calculated using a batch means procedure.

The outline of the chapter is as follows. In Section 5.2 we briefly review the basic aspects of Bayesian inference for parameter estimation and model selection. Then we provide the reader with some necessary details on MCMC in Section 5.3, where we also outline the new algorithm to calculate the ML. In Section 5.4 we assess the strengths and weaknesses of the new method relative to those that use nested sampling or parallel tempering. In Section 5.5 we test all competing algorithms on some toy problems like the analysis of pulsar timing data as encountered in pulsar timing array projects.

5.2   Bayesian inference

Bayesian inference methods provide a clear and consistent approach to parameter estimation and model selection. Consider a model, or Hypothesis, H with parameters $\vec{\Theta}$ for a dataset $\vec{d}$. Then Bayes' theorem states that

$$P\left(\vec{\Theta} \mid \vec{d}, H\right) = \frac{P\left(\vec{d} \mid \vec{\Theta}, H\right) P\left(\vec{\Theta} \mid H\right)}{P\left(\vec{d} \mid H\right)}, \tag{5.1}$$

where $P(\vec{\Theta}) := P(\vec{\Theta} \mid \vec{d}, H)$ is the posterior PDF of the parameters, $L(\vec{\Theta}) := P(\vec{d} \mid \vec{\Theta}, H)$ is the likelihood function, $\pi(\vec{\Theta}) := P(\vec{\Theta} \mid H)$ is the prior, and $z := P(\vec{d} \mid H)$ is the marginal likelihood.

The ML is the factor required to normalise the posterior over $\vec{\Theta}$:

$$z = \int L\left(\vec{\Theta}\right) \pi\left(\vec{\Theta}\right) \mathrm{d}^m \Theta, \tag{5.2}$$

where m is the dimensionality of $\vec{\Theta}$. The ML can then be used to calculate the so-called odds ratio in favor of model H_1 over H_0, which allows one to perform model selection:

$$\frac{P\left(H_1 \mid \vec{d}\right)}{P\left(H_0 \mid \vec{d}\right)} = \frac{z_1}{z_0} \frac{P\left(H_1\right)}{P\left(H_0\right)}, \tag{5.3}$$

where $P(H_0)$ and $P(H_1)$ are the prior probabilities for the different models . As with the prior for a model parameters, the prior probability for a model should be chosen to reflect the available information, if any.

In parameter estimation problems, one is interested in the posterior PDF, marginalised over all nuisance parameters. In this case, knowing the marginalised likelihood is not required, since the resulting marginalised posterior PDF can be normalised after the integration. However, when model choice itself is uncertain, the ML is no longer an uninteresting normalisation constant, but a key quantity that allows us to perform model selection.

Being the average of the likelihood over the prior distribution, the ML is larger for a model if more of its parameter space is likely and smaller for a model with large areas in its parameter space having low likelihood values. Even if the likelihood function has high peaks, in order to increase the ML these peaks must compensate for the areas in its parameter space where the likelihood is low. Thus the ML automatically implements Occam's razor: a simpler theory with compact parameter space will have a larger ML than a more complicated model, unless the latter is significantly better at explaining the data.

5.3 Markov Chain Monte Carlo

Markov Chain Monte Carlo methods (MCMC) can be used to sample from very complicated, high dimensional distribution; for Bayesian inference it is usually the posterior PDF. The method presented in this chapter could be useful for integration problems other than ML calculation, so we use the more general $f(\vec{\Theta})$ to denote this unnormalised function. The samples drawn from a distribution proportional to this function can then be used to perform the integrals we need for the marginal posterior PDFs and, as we show, for the ML. The exact mechanism that produces these samples can differ between MCMC methods and is irrelevant for the purposes of this work, but the result is always a large number of samples distributed according to $f(\vec{\Theta})$. The main advantage of this is that we do not have to sample from the entire volume of the parameter space or the prior, but only from a small fraction of it: the high probability density (HPD) regions. Especially for functions in high-dimensional parameter spaces this feature is crucial for efficient ML evaluation.

In Section 5.3.2 we show that all information needed to calculate the ML is already present in the samples of the MCMC. Then we give a practical example of how to calculate the integral in practice in Section 5.3.4.

5.3.1 The harmonic mean estimator

A very simple ML estimator can be derived from Equation (5.1):

$$
\begin{aligned}
\frac{1}{z} &= \frac{P\left(\vec{\Theta} \mid \vec{d}, H\right)}{L\left(\vec{\Theta}\right)\pi\left(\vec{\Theta}\right)} = \int \frac{P\left(\vec{\Theta} \mid \vec{d}, H\right)}{L\left(\vec{\Theta}\right)} \mathrm{d}^m \Theta \\
&= \left\langle \frac{1}{L\left(\vec{\Theta}\right)} \right\rangle_P,
\end{aligned}
\tag{5.4}
$$

where we have multiplied the right hand side by $\pi(\vec{\Theta})$, and integrated over all parameters. The expectation is an expectation over the posterior PDF. Equation (5.4) is then used to form the harmonic mean (HM) estimator of the ML (Newton & Raftery, 1994):

$$
\hat{z}_{HM} = \frac{1}{\frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{L(\vec{\Theta_i})}\right)},
\tag{5.5}
$$

where the summation is over all N MCMC samples.

Although the HM estimator follows directly from a true identity, it has some statistically undesirable properties. The HM estimator does converge to the correct value, but the variance of the estimator can be infinite. This is because the likelihood can reach values arbitrarily close to zero with nonzero probability in the case that these values are permitted by the prior. This is especially important if a peak in the likelihood is smaller than a corresponding peak in the prior PDF, i.e. when the data is informative. As a result, in the case that the data is informative, the HM estimator converges much slower than estimators that follow the central limit theorem (Wolpert, 2002).

5.3.2 The truncated harmonic mean estimator

In this chapter we advocate a modification to the HM estimator by eliminating the probability that we include samples in our estimator for which the likelihood gets arbitrarily close to zero. In the derivation in this section, we consider the general function $f(\vec{\Theta}) = zp(\vec{\Theta})$, where $p(\vec{\Theta})$ is a PDF proportional to $f(\vec{\Theta})$. We are interested in the integral over all parameters of $f(\vec{\Theta})$. Analogous to Equation (5.4), we now integrate $1/z$, but we now perform the integral only over a HPD region of $f(\vec{\Theta})$, given by $\mathcal{V} \subset \mathbb{R}^n$ with volume V. This gives:

$$
\frac{V}{Z} = \int_{\mathcal{V}} \frac{p(\vec{\Theta})}{f(\vec{\Theta})} \mathrm{d}^m \Theta.
\tag{5.6}
$$

This equation can be used to form an estimator similar to the HM estimator as follows. First note that the PDF $p(\vec{\Theta})$ is not normalised on $\mathcal{V}$. By defining the PDF $p_{\mathcal{V}}(\vec{\Theta}) = w p(\vec{\Theta})$, we can write a truncated harmonic mean estimator for the ML:

$$\frac{V}{Z} = w \int_{\mathcal{V}} \frac{p_{\mathcal{V}}(\vec{\Theta})}{f(\vec{\Theta})} \mathrm{d}^m \Theta = w \left\langle \frac{1}{f(\vec{\Theta})} \right\rangle_{p_{\mathcal{V}}}. \qquad (5.7)$$

Now, assuming that $N_{\mathcal{V}}$ samples of the total N MCMC samples are within $\mathcal{V}$, the HPD region of $f(\vec{\Theta})$, we can write down the new estimator:

$$\hat{z} = \frac{V}{\frac{N_{\mathcal{V}}}{N} \left(\frac{1}{N_{\mathcal{V}}} \sum_{\vec{\Theta}_i \in \mathcal{V}} \frac{1}{f(\vec{\Theta})} \right)}, \qquad (5.8)$$

where $N_{\mathcal{V}}/N$ appears as an estimator for w. Equation (5.8) is the estimator we propose as an improvement to the HM estimator in this chapter.

5.3.3   Analogy with tessellations

In order to provide a more intuitive approach to the estimator presented in Section 5.3.2, we try to motivate the use of that estimator in an entirely different way in this section. Consider the function:

$$f(x, y) = \exp\left(-ax^2 - by^2\right), \qquad (5.9)$$

where a and b are arbitrary model parameters. For the values $a = 1/5$ and $b = 2/5$, a Metropolis algorithm with $N = 40000$ samples yield a distribution of samples similar to Figure 5.1. We use the notation $\vec{\Theta}_i = (x_i, y_i)$ to indicate the ith sample. We would now like to calculate the integral

$$z = \int \int f(x, y) \, \mathrm{d}x \mathrm{d}y. \qquad (5.10)$$

In any MCMC algorithm, the function values $f(\vec{\Theta}_i)$ for each point are evaluated, and usually these values are stored together with the values of the parameters. These function values can be used to calculate the integral if we treat the MCMC samples in parameter space as an irregular grid. A straightforward way to do this is to calculate the Voronoi tessellation for the samples in the parameter space, an example of which is given in Figure 5.2. The samples are then the centres of the
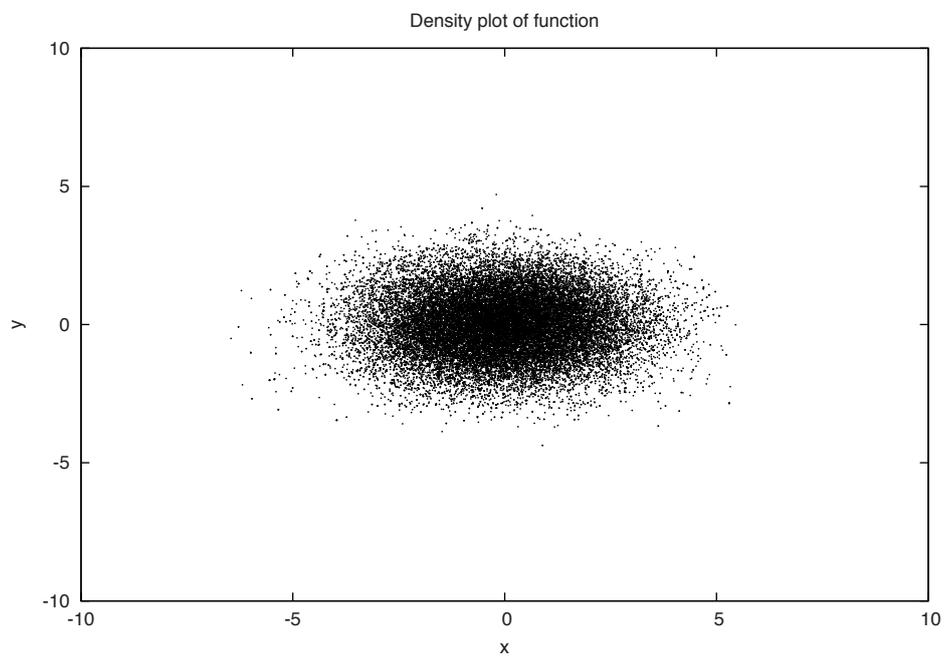
Figure 5.1: A scatter plot of of 40000 samples, drawn using a Metropolis algorithm from the function $f(x, y) = \exp(-ax^2 - by^2)$, with $a = 1/5$ and $b = 2/5$.
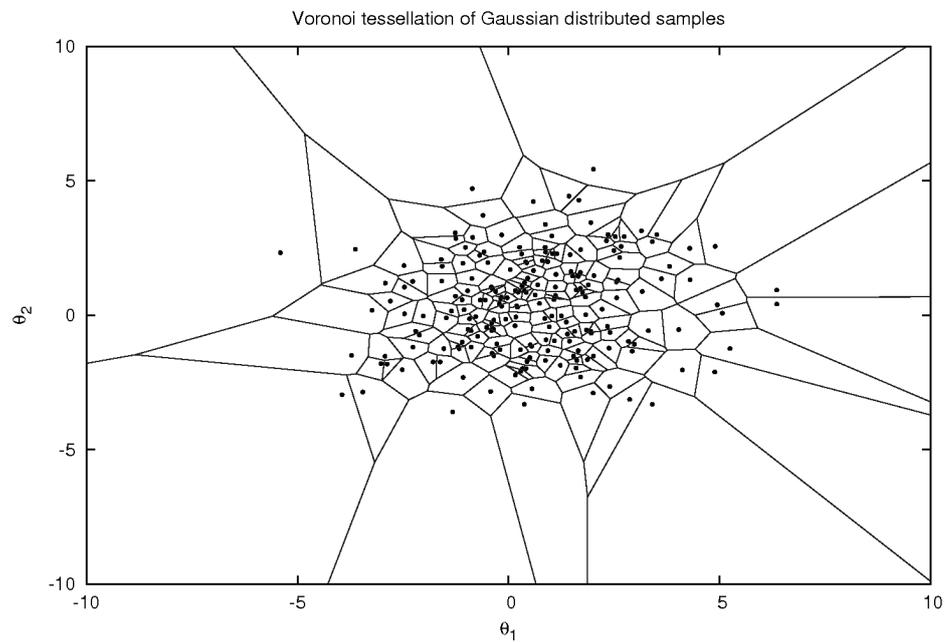
Figure 5.2: An example of a Voronoi tessellation. Here we have taken 200 samples from a 2-dimensional Gaussian distribution as the centres of the Voronoi diagram.

Voronoi cells, and the integral can be calculated as follows:

$$z \approx \sum_i f\left(\vec{\Theta}_i\right) O_i,$$

(5.11)

where O_i is the surface of the ith Voronoi cell, and we only sum over closed cells (with finite surface). This procedure converges to the correct value for large number of samples N, and for all tessellations, not just when we use Voronoi cells. In practice, although Voronoi tessellations can be computed for any dimensionality, this becomes computationally prohibitive in practice for problems with high dimensionality (Edelsbrunner & Shah, 1996). This procedure does illustrate that all information needed to evaluate the integral z is present in the MCMC chain.

We now proceed to show that we can link the estimator of Equation (5.8) to the voronoi tessellation. First, we consider the HPD region $\mathcal{V}$ of $f(\vec{\Theta})$, and write the volume V as a sum of the volumes of the voronoi cells:

$$V = \sum_{i:\vec{\Theta}\in\mathcal{V}} O_i.$$

(5.12)

These cells comprise an inhomogeneous Poisson Voronoi Tessellation (PVT). Analytical expressions about the statistics of inhomogeneous PVTs are quite difficult to establish; most results are based on large simulations. Analytically it is known that $\langle O_i \rangle \to 1/$ density of points for $N \to \infty$, for cells away from the boundary (Barr, 2008). For finite N this expectation is biased. We ignore the bias, and introduce a proportionality constant α_N as follows:

$$\langle O_i \rangle = \frac{\alpha_N}{f(\vec{\Theta}_i)}.$$

(5.13)

By only considering the HPD region $\mathcal{V}$, we do not have to deal with cells with infinite volume. We can estimate α_N analogous to Equation (5.11), but now with the expectation of O_i:

$$\sum_{i:\vec{\Theta}\in\mathcal{V}} f(\vec{\Theta})\langle O_i \rangle = \alpha_N \sum_{i:\vec{\Theta}\in\mathcal{V}} 1 = \alpha_N N_{\mathcal{V}} \approx zw,$$

(5.14)

where we have used notation as in Section 5.3.2. Since the expectation of a sum is the sum of the expectation, and using Equations (5.14 & 5.13): we also have:

$$V = \sum \langle O_i \rangle \approx z \frac{w}{N_{\mathcal{V}}} \sum_{i:\vec{\Theta}\in\mathcal{V}} \frac{1}{f(\vec{\Theta})}.$$

(5.15)

This expression directly leads us to the estimator of Equation (5.8) that we have
derived in the previous section.

5.3.4   A practical algorithm

In this section we construct a simple practical algorithm for numerically calculat-
ing integrals using regular MCMC methods. As stated in Section 5.3.3, we need
to define a HPD region of the parameter space that is sufficiently populated by
MCMC samples. In the case of multimodal posteriors, we should take samples
that 'belong' to one specific mode. Clustering algorithms like X-means (Pelleg &
Moore, 2000), and G-means (Hamerly & Elkan, 2003) can be used for this.

Assuming that we have found M samples that belong to a specific mode, we
can then define a HPD region as follows. Assume that the M samples are sorted
according to their function value $f(\vec{\Theta}_i)$, with $f(\vec{\Theta}_0)$ the highest value. We then first
approximate the maximum of $f(\vec{\Theta})$ as:

$$\vec{\mu} = \frac{1}{k} \sum_{i=1}^{k} \vec{\Theta}_i, \qquad \boxed{5.16}$$

where $k = aM$ is a small fraction of M, dependent on the shape of the mode. We
use $a = 1/20$ in several cases. We now define our HPD region $\mathcal{V}$ as an ellipsoid
with centre $\vec{\mu}$, and covariance matrix C, defined by:

$$C = \frac{1}{n} \sum_{i=1}^{n} \left(\vec{\Theta}_i - \vec{\mu}\right)\left(\vec{\Theta}_i - \vec{\mu}\right)^T, \qquad \boxed{5.17}$$

where $n = bM$ is a fraction of M, again dependent on the shape of the mode. We
use $b = 1/5$ in several cases. All samples within this ellipsoid of size $\sqrt{r^2 \det C}$
satisfy:

$$\left(\vec{\Theta} - \vec{\mu}\right)^T C^{-1} \left(\vec{\Theta} - \vec{\mu}\right) \leq r^2. \qquad \boxed{5.18}$$

We adjust the parameter r^2 such that $l = cM$ samples satisfy this relation. It is cru-
cial that we choose l to be as large as possible, while still satisfying the requirement
that the entire ellipsoid is sufficiently populated with samples. If l is too small, we
will not achieve very high accuracy for our integral z, but if l is too large, we have
underpopulated parts of the ellipsoid which results in the wrong outcome. We use
$c = 1/3$ in several cases.

We now have all the ingredients needed to calculate the integral z, as the Vol-

ume of our k-dimensional HPD region is given by:

$$V = \frac{r^k \pi^{\frac{k}{2}}}{\Gamma\left(1 + \frac{k}{2}\right)} \sqrt{\det C}.$$ (5.19)

This, combined with Equation (5.8) allows us to evaluate the integral.

We would like to note that this prescription is merely one of many that could be used. In this case we have defined our HPD region as an ellipsoid located at a maximum of our integrand, but any other HPD region will suffice.

5.3.5 Error estimates

An ideal MCMC algorithm generates samples according to a PDF proportional to $f(\vec{\Theta})$. Consider the number of samples inside the HPD region $\mathcal{V}$. This number follows a Poissonian distribution with mean and variance equal to cM. Using that we can obtain a rough estimate for the error in the integral:

$$\Delta z = \frac{1}{\sqrt{cM}} z.$$ (5.20)

We note that this Equation (5.20) is too simplistic for the truncated harmonic mean estimator we introduce in chapter, but it can be used as an order of magnitude guide to how well a particular simulation will do. Even if the estimator follows the central limit theorem, which we do not show here, we show in Section 5.5 that the error estimate should depend on the samples of the MCMC algorithm that have been used. Many MCMC algorithms produce correlated samples, and in that case the error estimate of the ML estimator should also increase (Roberts et al., 1997).

Having efficiency as one of our goals, we would like to produce reliable error-bars on our numerical integral using the correlated MCMC samples. We propose to use a batch means method (Efron, 1979). In the case of several chains, one can use the spread in the estimates based on different chains to estimate the error of the integral. In the case of a single chain, we propose to divide the chain in 10 succeeding parts, and use the spread of the integral in those 10 parts to estimate the error of the numerical integral. In Section 5.5.4 we test the error estimates discussed here extensively.

Another point noting is that we expect the truncated harmonic mean estimator of Equation (5.8) to be slightly biased. Because the denominator of Equation (5.8) is an unbiased estimator, this does not hold for the ML estimator itself. We investigate the bias of the ML estimator in Section 5.5.4.

5.4   Comparison to other methods

Although the algorithm developed in this chapter is quite generally applicable, in practice there are caveats for each integration problem that one needs to be aware of in order to choose the right integration algorithm. In this section we discuss the strengths and the weaknesses of the method developed in this chapter in combination with the Metropolis Hastings MCMC algorithm, and we compare it to two other methods widely used in astronomy: Nested sampling, and Thermodynamic Integration based on Parallel Tempering (TI-PT).

For all algorithms, the main criteria that we cover are efficiency, accuracy and robustness.

5.4.1   Metropolis Hastings

The main advantage of the Metropolis Hastings (MH) algorithm in relation to ML evaluation is that of parameter space reduction; the full parameter space over which we would like to evaluate the integral is too large, which is why we cannot use a regular grid. The MH algorithm will mostly sample in the HPD regions, the regions of parameter space where the function value $f(\vec{\Theta})$ is high enough to significantly contribute to the integral. In general this can be done by drawing samples from a distribution $P(\vec{\Theta})$ that resembles the function $f(\vec{\Theta})$, with the optimal choice being $P(\vec{\Theta}) = f(\vec{\Theta})$. MH is specifically designed to satisfy this optimal relation, making the MH algorithm optimally efficient from a theoretical point of view (Roberts et al., 1997). In Figure 5.1 we show an example of samples generated with a MH algorithm, released on a 2-dimensional Gaussian function.

The drawback of MH is that one can never be sure that all important parts of the parameter space have been covered. Some functions have many distinct peaks, the so-called modes of the function. The MH algorithm often has difficulty to make the chain move from one mode to another. If the function $f(\vec{\Theta})$ is highly multimodal, we must make sure that the sample density ratio between the modes is right and that we have reached all the important modes in the entire parameter space.

An additional practical challenge with the method developed in this chapter is to make sure that we have constructed a HPD region that is sufficiently populated with samples. In the case of a not very peaked, or highly degenerate function this requires special care.

5.4.2   Nested sampling

Nested sampling is a Monte Carlo technique aimed at accurate evaluation of numerical integrals, while staying efficient (Skilling, 2004). The algorithm is especially

well suited for problems that have huge parameter spaces, and very complicated multi modal distributions. Nested sampling starts with sampling from the original parameter space; in the case of ML calculation this is equivalent to sampling from the prior distribution of the parameters. The density with which the parameter space is sampled is adjustable in the form of an number n_L of so-called live points; the number of points evenly distributed among the part of the parameter space we are exploring. At the start of the algorithm, the live points are evenly distributed over the entire parameter space. Then the live points are replaced one-by-one under the restriction that the newly sampled live points are higher than the lowest one we have not replaced yet. Effectively this is the same as shrinking the parameter space by a fixed factor every time we replace a new live point, ultimately sampling only the part of the function close to the maximum.

The main advantage of nested sampling is that one generates samples from the original parameter space, thereby completely eliminating the main deficiency of the MH algorithm: all modes of the function $f(\vec{\Theta})$ are sampled from, provided that we choose n_L high enough, i.e. that we have a high enough sampling density in the parameter space. And due to the shrinking of the parameter space by a fixed factor with the draw of each new live point, the size of the parameter space decreases exponentially. This way all of the parameter space is explored in a relatively manageable way, making nested sampling quite efficient in searching the parameter space for HPD regions.

The main disadvantage of nested sampling is that the samples are not drawn from a distribution that closely resembles the function $f(\vec{\Theta})$ we want to integrate. This method will therefore never reach the efficiency that the MH algorithm offers. In Figure 5.3 we show an example of samples generated with nested sampling, applied to a 2-dimensional Gaussian function.

5.4.3 Parallel tempering

Thermodynamic Integration based on Parallel Tempering (TI-PT) is an algorithm spawned from the desire to solve the problems of MH and similar MCMC methods. TI-PT algorithms possess better mixing properties, allowing the chain to "escape" local extrema, and allow one to calculate the complete integral, or in our case the ML (Earl & Deem, 2005). Let us briefly review TI-PT in this section, without discussing it in too much detail.

The main idea of TI-PT is that of parameter space exploration by adding an imaginary inverse temperature β to the system, changing the integrand of our integral to:

$$f_\beta\left(\vec{\Theta}\right) = \left(f(\vec{\Theta})\right)^\beta .$$

(5.21)

121
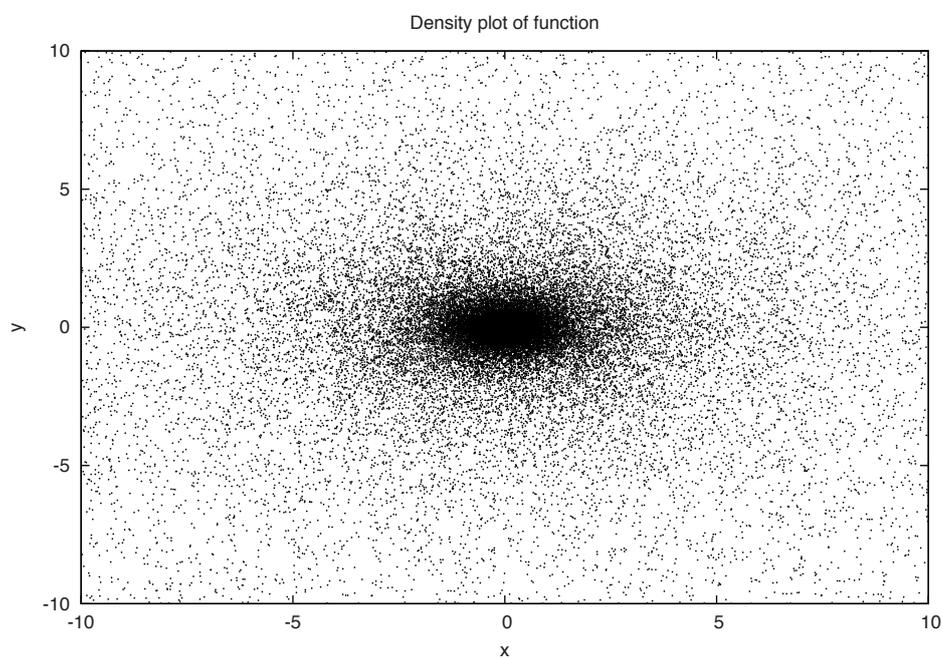
Figure 5.3: A scatter plot of 40000 samples, drawn using a nested sampling algorithm with 5000 live points from the function $f(x, y) = \exp(-ax^2 - by^2)$, with $a = 1/5$ and $b = 2/5$.

Then many MCMC chains are released in the parameter space, each with a different temperature $\beta \in [0, 1]$. A clever swapping system is employed, allowing the chain with $\beta = 1$ - the "true" chain - to swap parameters with chains of higher temperature every now and then provided that such a high-temperature chain was able to reach a point in parameter space with high $f(\vec{\Theta})$. This trick allows the coldest system, the "true" chain with $\beta = 1$, to escape local extrema.

The integral over $f(\vec{\Theta})$ is calculated by using all chains, not just the one with $\beta = 1$, as follows. We first define a partition function:

$$Z(\beta) = \int d\vec{\Theta} f_\beta \left(\vec{\Theta} \right), \quad\quad (5.22)$$

which has a logarithmic derivative of:

$$\frac{d}{d\beta} \log \left(Z(\beta) \right) = \frac{1}{Z(\beta)} \frac{d}{d\beta} Z(\beta) = \left\langle \log \left(f(\vec{\Theta}) \right) \right\rangle_\beta, \quad\quad (5.23)$$

where $\langle . \rangle_\beta$ is the expectation value of a quantity over a distribution proportional to $f_\beta(\vec{\Theta})$. Since we know that our desired integral can be expressed as $z = Z(1)$, Equation (5.23) is all we need to calculate it:

$$\log \left(Z(1) \right) = \log \left(Z(0) \right) + \int_0^1 d\beta \left\langle \log \left(f(\vec{\Theta}) \right) \right\rangle_\beta. \quad\quad (5.24)$$

The observant reader will have noted that we have neglected to mention the size $Z(0)$ of the parameter space that we explore with our chains. The high temperature chain with $\beta = 0$ is unbounded by the function $f(\vec{\Theta})$, and therefore will transverse the entire parameter space. We should make sure that we limit the size of the parameter space as much as possible, without missing any peaks of $f(\vec{\Theta})$.

The main advantages of TI-PT are that it explores the entire parameter space, even in the presence of strong local peaks, and that the ML can be calculated. These advantages are accompanied with the large computational costs of the extra chains with $\beta \neq 1$, which has resulted in the need for alternative methods like nested sampling. As MultiNest, a specific nested sampling algorithm, is supposed to outperform TI-PT in virtually all cases (Feroz et al., 2009), we compare our method to MultiNest only in Section 5.5.

5.5   Applications and tests

In this section we consider several toy models, and we apply several integration algorithms to these toy models as to compare the algorithms. In this whole section

we have two ways to use the MH algorithm. To produce correlated samples, we use a MH algorithm where we use the entire chain. To produce uncorrelated samples, we use the same algorithm, but we only store one in every j samples produced by the algorithm. We choose j high enough that there is negligible correlation between 2 succeeding used samples; in the case of a 4-dimensional Gaussian as in Section 5.5.1 we use $j = 100$.

5.5.1 Toy model 1: a high-dimensional Gaussian

We first consider a problem that is a typical example of what MH algorithm are efficient in: a highly peaked, high-dimensional function. The curse of dimensionality prohibits any direct numerical integration scheme on a fine grid, but analytical integration is possible. Consider the multiplication of N Gaussian functions,

$$f_1(\vec{x}) = \prod_{i=1}^{N} \sqrt{\frac{a_i}{2\pi}} \exp\left(-\frac{1}{2}a_i x_i^2\right),$$ (5.25)

with a_i the width of the Gaussian in the ith direction. Now let us perform a volume preserving coordinate transformation using a random orthogonal matrix R, generated with a QR decomposition of a random matrix, as follows: $\vec{\Theta} = R\vec{x}$. If we introduce a matrix $A_{ii}^{-1} = a_i = 1 + i$ with $A_{ij}^{-1} = 0$ for $i \neq j$, we then have a highly degenerate high-dimensional Gaussian function:

$$f_1\left(\vec{\Theta}\right) = \frac{1}{(2\pi)^{N/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}\vec{\Theta}^T C^{-1}\vec{\Theta}\right),$$ (5.26)

where we have introduced the covariance matrix $C = RAR^T$.

We now apply in turn the MultiNest nested sampling algorithm and the algorithm developed in this chapter combined with MH to the multi-dimensional Gaussian for various number of dimensions.

For the MultiNest algorithm, we use a number of live points $L = 50N$ with N the number of dimensions, and we have set the sampling efficiency to $e = 0.3$ and the ML tolerance to $t = 0.1$ as advocated in (Feroz et al., 2009).

We set up the Metropolis-Hastings algorithm to have an acceptance ratio equal to the optimal value of 23.4% (Roberts et al., 1997). The parameters of the algorithm advocated in Section 5.3.4 have been set to: $a = 1/20, b = 1/5, c = 1/3$. We have used the number of samples N used by the MultiNest algorithm as the number of samples in our MCMC chain. However, we realise that the MultiNest algorithm might be improved in the future, as the algorithm is still under construction. The lowest efficiency (used samples / drawn samples) we encountered for

		log (z) for different algorithms		
n	# N	MultiNest	Unc. MCMC	Cor. MCMC
2	2902	-0.17 ± 0.18	-0.018 ± 0.025	0.03 ± 0.025
4	7359	0.20 ± 0.17	0.007 ± 0.024	-0.01 ± 0.03
8	24540	0.17 ± 0.17	-0.01 ± 0.01	0.02 ± 0.03
16	10^5	0.05 ± 0.18	0.001 ± 0.006	0.004 ± 0.03
32	10^6	0.43 ± 0.17	0.004 ± 0.004	-0.015 ± 0.010
64	4.10^6	1.07 ± 0.77	-0.0004 ± 0.0007	-0.02 ± 0.016

Table 5.1: The log-integral values of the function f_1 of Equation (5.26). The analytically integrated value is $\log z = 0$ for all values of N.

Jump 1&2	Jump 3&4	Jump 5&6	Jump 7&8
$\Theta_1 \pm 4\pi$	$\Theta_1 \pm 2\pi$	$\Theta_1 \pm 2\pi$	$\Theta_1 \pm 0$
$\Theta_2 \pm 0$	$\Theta_2 \pm 2\pi$	$\Theta_2 \mp 2\pi$	$\Theta_2 \pm 4\pi$

Table 5.2: The possible jumps in the eggbox MCMC toy-problem.

this toy-problem was $e = 0.08$; we therefore estimate that the error-bars could be decreased with a factor of maximally $\sqrt{1/0.08} \approx 3.5$. The results of this toy-model are shown in table 5.1.

5.5.2 Toy model 2: egg-box function

Just as in FHB09, we now consider a highly multimodal two-dimensional problem for which the function resembles an egg-box. The function is defined as:

$$f_2\left(\vec{\Theta}\right) = \exp\left[2 + \cos\left(\Theta_1\right)\cos\left(\Theta_2\right)\right]^5, \qquad (5.27)$$

where we set the domain of the function equal to $[0, 10\pi]$ for both parameters. The shape of this function is shown in Figure 5.4. This is a typical problem where difficulties arise for traditional MCMC algorithms. Many solutions have been proposed for situations like this (Newman & Barkema, 1999), but in practice one needs to have additional information about the problem for any of those solutions to be reliable. For the sake of clarity of this chapter, we do not concern us with the practical implementation of the MCMC algorithm. We assume that a suitable trick can be found for the problem at hand so that the algorithm proposed in this chapter can be used. For the eggbox toy-model we will use a jump-technique. At each iteration of the MCMC algorithm, there is a small probability, 1% in this case, that the chain will jump to a neighbouring mode. The available jumps are shown in table 5.2.

The MultiNest algorithm is ideally suited for this problem, as this is a low-
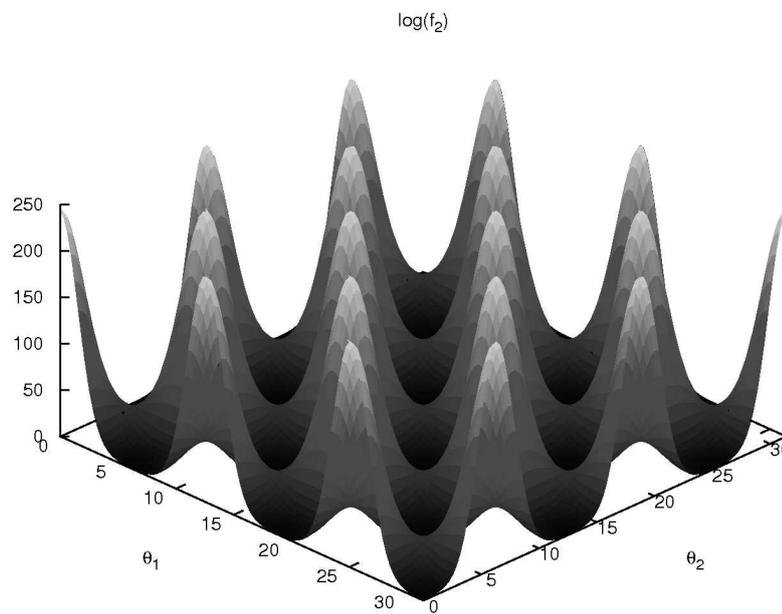
Figure 5.4: Toy-model 2: the eggbox of Equation (5.27)

	log(z) for different algorithms		
# N	MultiNest	Unc. MCMC	Cor. MCMC
60210	240.19 ± 0.05	240.19 ± 0.027	240.23 ± 0.05

Table 5.3: The log-integral values of a single mode of the eggbox function of Equation (5.27) The fine-grid integrated value is log $z = 240.224$

dimensional multimodal function. With enough live samples all modes should be easily discovered, and the peaks are symmetric and well-behaved. We run Multi-Nest on this problem with the same parameters as in FHB09: We use $L = 2000$ live points, efficiency $e = 0.3$, and tolerance $t = 0.1$.

Table 5.3 shows the results of this analysis[1]. For evaluating the integral with the MCMC chain, we have taken a total of $N = 60210$ samples as was done with Multi-Nest, but we have used only the samples of one single peak in Equation (5.8). The number of samples in a single peak is 2/25 of the total number of samples, leading to loss of accuracy. Though more sophisticated methods can be constructed by, say, averaging the ML estimators for several HPD regions for all individual peaks, we show here that we only need to find one single HPD region that is sufficiently populated to calculate a reliable value for the integral.

5.5.3 Application in Pulsar Timing

In pulsar timing data analysis, one often encounters datasets of which the exact statistics are not well known. Bayesian model selection would provide the ideal tool to obtain information about the power spectra present in the data. van Haasteren et al. (2009, see also chapter 2) give a full description of the Bayesian data analysis for pulsar timing arrays, but their work lacks a method to calculate the ML from the MCMC samples. In this section, we use a pulsar timing mock dataset to show that the method developed in this chapter is well-suited for ML calculation in pulsar timing problems. We also use MultiNest to analyse this dataset, and we compare the two results.

For the purposes of this chapter, the description of the pulsar timing posterior distribution is kept brief; full details can be found in chapter 2. The data of pulsar timing experiments consists of the arrival times of pulsar pulses (TOAs), which arrive at the earth at highly predictable moments in time (Hobbs et al., 2006). The deviations from the theoretically predicted values of these TOAs are called the timing-residuals (TRs). These TRs are the data we concern ourselves with in this example.

---

[1]The true value presented here is different from the one noted in FHB09. We have incorporated hypercube transformation into our results.

Consider $n = 100$ TRs, denoted as $\vec{\delta t}$, observed with intervals between succeeding observations of 5 weeks. Assume that we are observing a stable and precise millisecond pulsar with timing accuracy about $\sigma = 100$ns (the error bars on the TRs). Usually σ not precisely known, since pulsar timers generally assume that their estimate of the error bar is slightly off. Several datasets of millisecond pulsars also seem to contain correlated low frequency noise (Verbiest et al., 2009). We therefore also allow for some correlated timing-noise in the data, with a power-spectrum given by:

$$S(f) = r^2 \gamma \exp(-\gamma f),$$ (5.28)

where f is the frequency, r is the amplitude of the correlated timing-noise in ns, and γ is the typical size of the structures that appear in the data due to this correlated timing-noise. Following chapter 2, we can now write the likelihood function for the TRs as a multi-dimensional Gaussian:

$$P\left(\vec{\delta t} \mid \sigma, r, \gamma\right) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp\left(-\frac{1}{2}\vec{\delta t}^T C^{-1} \vec{\delta t}\right),$$ (5.29)

where C is an $(n \times x)$ matrix, with elements defined as:

$$C_{ij} = \sigma^2 \delta_{ij} + r^2 \frac{\gamma^2}{\gamma^2 + \tau_{ij}^2},$$ (5.30)

with δ_{ij} the Kronecker delta, and τ_{ij} is the time difference between observation i and observation j.

Simulating a mock dataset from such a Gaussian distribution is quite straightforward; for details see chapter 2. We now analyse a mock dataset, shown in Figure 5.5, generated with parameters: $\sigma = 100$ns, $r = 100$ns, and $\gamma = 2$yr. We assume uniform prior distributions for all parameters: $\sigma \in [0, 1000]$ns, $r \in [0, 1000]$ns, and $\gamma \in [0, 10]$yr. The posterior is then sampled using both MultiNest and a Metropolis-Hastings algorithm, resulting in marginalised posterior distributions as shown in Figure 5.6 &5.7. ML values are in good agreement between the two methods:

$$
\begin{aligned}
z_{\text{MCMC}} &= \exp(1523.12 \pm 0.17) \\
z_{\text{MultiNest}} &= \exp(1522.93 \pm 0.15).
\end{aligned}
$$ (5.31)

For both methods, the same number of samples has been used: $N = 9617$.
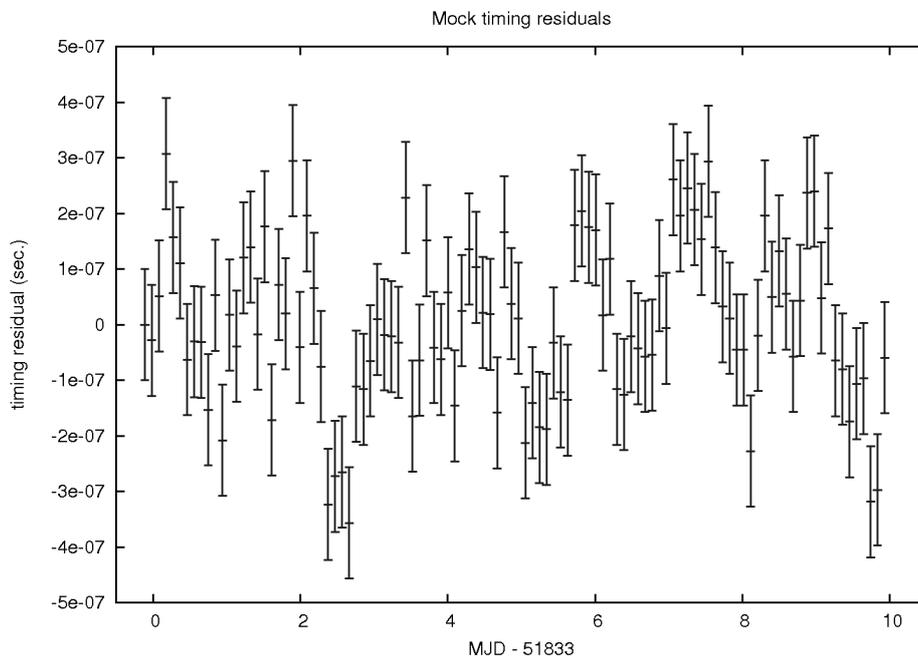
Figure 5.5: The mock timing-residuals we have analysed with both MultiNest and the method developed in this chapter
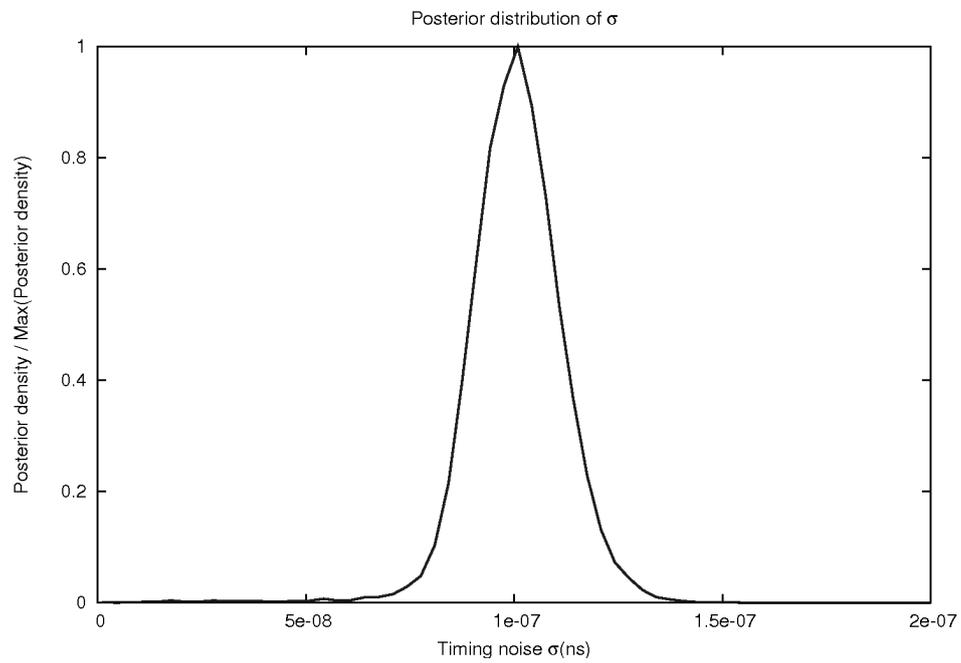
Figure 5.6: The marginalised posterior of the *a* parameter, sampled using a Metropolis-Hastings MCMC method.

Figure 5.7: The marginalised posterior of the r and γ parameters, sampled using a Metropolis-Hastings MCMC method.

5.5.4   Assessing bias

We now test the accuracy and the bias of the algorithm by running it many times on the same toy-problem, and then considering the statistics of the ensemble. We found the 16-dimensional Gaussian of Section 5.5.1 to be an illustrative example. Just as in table 5.1, we take $N = 10^5$ and $c = 0.3$, and then we run $n = 10^4$ Metropolis-Hastings chains on this toy-problem. For the ith chain we then calculate the integral z_i and batch means error estimate σ^{BM}. We have presented the results of this analysis as a histogram of z_i values in Figure 5.8. Several useful quantities that characterise the ensemble are:

$$\bar{z} \;=\; \frac{1}{n}\sum_{i=1}^{n} z_i = 0.980$$

$$\bar{\sigma} \;=\; \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2} = 0.028$$

$$\bar{\sigma}^{\text{BM}} \;=\; \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\sigma_i^{\text{BM}}\right)^2} = 0.027, \tag{5.32}$$

where $\bar{z}$ is the integral average, $\bar{\sigma}$ is the rms of the integral values, and $\bar{\sigma}^{\text{BM}}$ is the rms value of the batch means errors.

Figure 5.8 shows that the batch means error estimate is quite correct, since $\bar{\sigma} \approx \bar{\sigma}^{\text{BM}}$. However, though smaller than $\bar{\sigma}$, there is a significant deviation in the value of of $\bar{z}$ compared to the true value $z = 1$. This shows that the estimator indeed has a small bias, as noted in Section 5.3.5.

In order to investigate the behaviour of the bias of $\bar{z}$, we perform 3 additional tests. In all 3 cases, we construct a new ensemble of MCMC chains identical to the ensemble above, except for one parameter. The differences with the above mentioned ensemble are:

1. Instead of a correlated MCMC chain, we use an MCMC chain of uncorrelated samples, produced by performing a regular MCMC but only storing every 100th sample.

2. $N = 10^7$, instead of $N = 10^5$. This results in much more samples in the HPD region.

3. $c = 0.7$, instead of $c = 0.3$, which also results in more samples in the HPD region.

We note that these results are merely illustrative of the behaviour of the ML estimator, since the bias of the estimator is problem dependent.

We present the results of this analysis as the values of Equation (5.32) in table
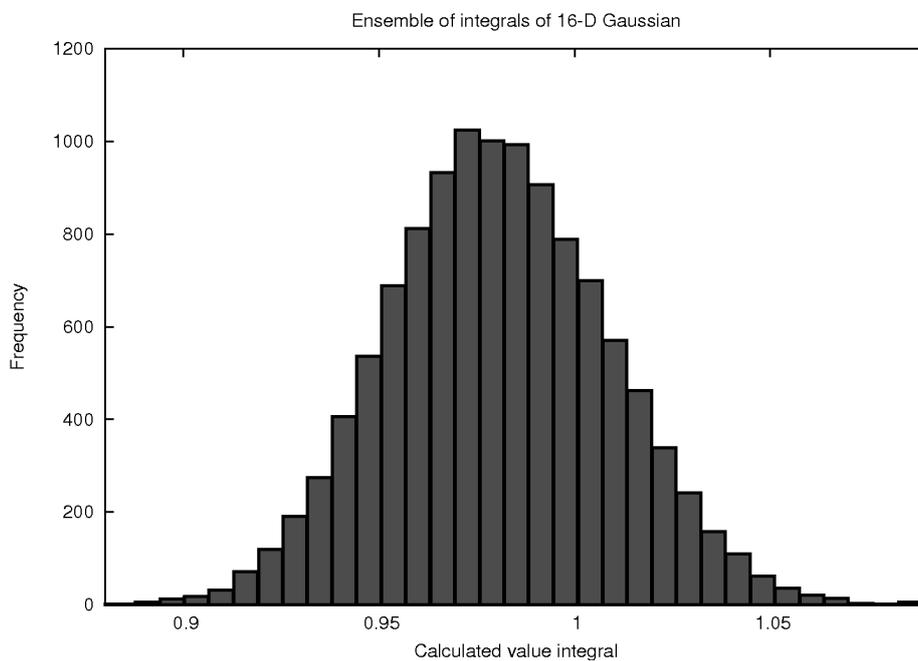
Figure 5.8: Histogram of the frequency of calculated integral using the method developed in this chapter. We have taken the 16-dimensional Gaussian of Equation (5.26) as integrand. Here we have analysed it with $N = 10^5$, and $c = 0.3$. This histogram has mean and standard deviation: $\bar{z} = 0.980$, $\bar{\sigma} = 0.028$. The rms of the batch means error of the ensemble was $\bar{\sigma}^{\text{BM}} = 0.027$

Cor./Unc.	N	c	$\bar{z}$	$\bar{\sigma}$	$\bar{\sigma}^{\mathrm{BM}}$
Ensemble statistics for different parameters

Cor./Unc.	N	c	$\bar{z}$	$\bar{\sigma}$	$\bar{\sigma}^{\mathrm{BM}}$
Cor.	10^5	0.3	0.980	0.028	0.027
Unc.	10^5	0.3	1.000	0.006	0.006
Cor.	10^7	0.3	1.000	0.003	0.003
Cor.	10^5	0.7	0.994	0.020	0.034

Table 5.4: The statistics of Equation (5.32) for various ensembles of $n = 10^4$ MCMC runs on a 16-dimensional Gaussian. The first chain has the same parameters as used in section 5.5.1. The other chains differ in either number of samples per chain N, the size of the HPD region c, or whether or not the samples in a chain are correlated. Note that the error in the uncorrelated chain equals the limit of $\sigma = \sqrt{1/cN} = 0.006$.

5.4. All adjustments seem to reduce the bias of the estimator. Several notes are in order:

1. The only reason we can increase c is because we know exactly what the integrand looks like. In practical applications this is probably not an option. Also note that batch means error increases, indicating that estimate of the integral is less stable.

2. The fact that the uncorrelated chain performs as well as described by Equation (5.20) shows that the algorithm suffers significantly under the use of correlated MCMC chains.

3. Increasing the number of samples in a chain makes the calculated integral more reliable, as the HPD region is more densely populated. Note that this large chain is build up of small chains that would yield a more biased value.

5.6 Discussion and conclusions

We develop and test a new algorithm that uses MCMC methods to accurately evaluate numerical integrals that typically arise when evaluating the ML. The new method can be applied to MCMC chains that have already been run in the past so that no new samples have to be drawn from the integrand, provided that we can define a high probability density region of the integrand well enough based on the MCMC samples. We test the new algorithm on several toy-problems, and we compare the results to other methods: nested sampling and thermodynamic integration based on parallel tempering. We conclude that the new algorithm could be of great value for high-dimensional, peaked problems. When applicable, the new algorithm can outperform other algorithms, provided that the MCMC has been

properly executed. This new algorithm is therefore expected to be useful in astro-physics, cosmology and particle physics.

We have demonstrated that the new algorithm suffers under the use of corre-lated MCMC chains, produced by using the entire chain of a particular MCMC method like Metropolis-Hastings. If the new algorithm is used in combination with an uncorrelated chain, the accuracy of the numerical integral can reach the value: $\sigma = z/\sqrt{N}$, with σ the uncertainty, z the value of the integral, and N the number of MCMC samples in the high probablility density region. Using corre-lated MCMC samples can significantly increase the integral uncertainty, and longer MCMC chains are needed for the integral to converge. We have also shown that the new estimator is slightly biased, where the bias is problem dependent. Additional tests to assess convergence and bias are required.
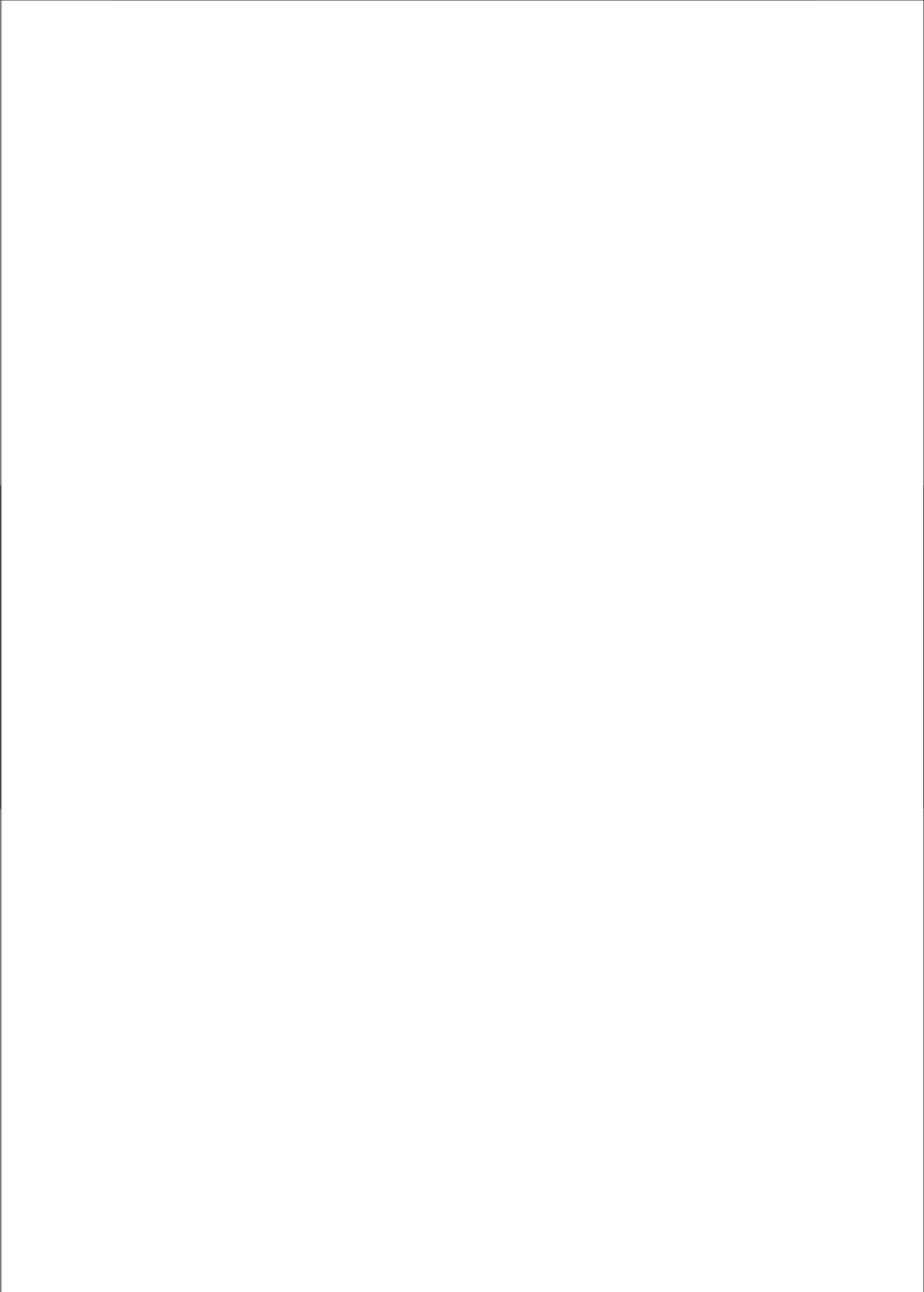
6

Nederlandstalige samenvatting

Het overgrote deel van wat wij op dit moment weten van het Universum zijn we te weten gekomen door met telescopen te kijken naar *electromagnetische straling*. Voorbeelden van dit soort straling zijn zichtbaar licht, radiogolven, infrarood straling, en ultraviolet licht. Het verschil tussen deze voorbeelden is de golflengte van de electromagnetische golven waaruit deze verschillende soorten straling bestaan. Algemeen gesproken bestaat electromagnetische straling uit golven in het electromagnetisch veld, waarbij de golflengte van die golven bepaalt wat voor soort straling het is. Op die manier is er een heel spectrum van electromagnetische straling mogelijk (zie figuur 6.1 voor een overzicht), en voor elke soort straling is een andere telescoop nodig om die straling waar te nemen. Net zoals we met onze menselijke ogen geen radio signalen kunnen waarnemen, hebben we ook een andere telescoop nodig om Röntgen straling te meten dan om zichtbaar waar te nemen.

Naast electromagnetische golven, zijn we ook in staat om dingen over het Universum te weten te komen door te kijken naar deeltjes die onze atmosfeer binnenkomen. Uiteraard kan dat zijn in de vorm van meteorieten, maar ook microscopisch kleine deeltjes komen elke seconde de dampkring van de aarde binnen. Onder deze deeltjes bevinden zich elementaire deeltjes zoals electronen/positronen (beta-straling), muonen (zware broertjes van de electronen), He^{2+} kernen (alpha-straling), neutrino's, en in principe alles wat er maar in de natuur te vinden is. Al deze deeltjes kunnen worden gedetecteerd, en leveren een schat aan informatie op over het Heelal.

Helaas kan niet het gehele Universum met behulp van deze straling en deeltjes bekeken worden. Sommige delen van het Unversum zijn op dit moment onzichtbaar voor ons. Bijvoorbeeld de oerknal. In het Heelal gaat de volgende regel op: hoe verder je kijkt, hoe langer geleden het is dat hetgene dat je ziet gebeurd is. Dus als je maar ver genoeg terugkijkt zou je in principe de oerknal kunnen zien. Helaas weten we dat het Universum in het vroege begin (slechts 380.000 jaar na de oerknal) ondoorzichtig was, net als een glas melk. We kunnen dus niet verder terugk-
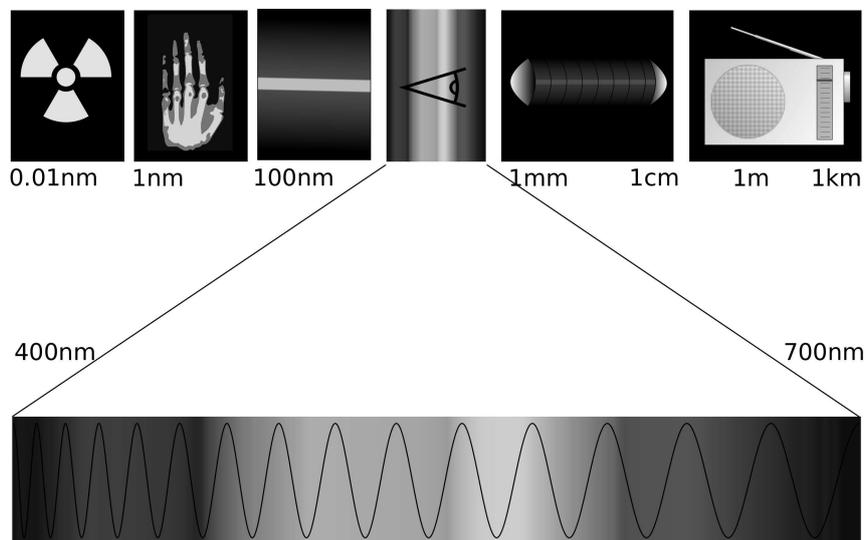
Figure 6.1: Het electromagnetisch spectrum, korte golflengten aan de linkerkant, lange golflengten aan de rechterkant. Boven, van links naar rechts: γ-straling Röntgen straling, ultraviolet licht, zichtbaar licht, infrarood licht, en radiogolven. Gemaakt door Tatoute, gereproduceerd met toestemming.

ijken dan deze magische grens. Ook onzichtbaar voor ons zijn samensmeltende zwarte gaten. Zwarte gaten zijn objecten waar niets uit kan ontsnappen zodra het erin valt, dus ook licht niet. De enige manier waarop een zwart gat invloed uitoefend op zijn omgeving is via zijn zwaartekracht.

Dit zijn slechts voorbeelden van delen van of gebeurtenissen in het Universum die wij op het moment niet kunnen zien omdat we niet over het juiste gereedschap beschikken. Net zoals het met Röntgen stralen mogelijk is om te zien wat er in het menselijk lichaam gebeurd, is het misschien wel mogelijk om met een ander stel ogen anders naar het Universum te kijken en deze interessante gebeurtenissen te bekijken. Daarom zijn wetenschappers altijd op zoek naar nieuwe manieren om precies dat te doen. Dit proefschrift gaat over één van die nieuwe manieren: het gebruik en meten van gravitatiegolven.

Gravitatiegolven zijn nog nooit direct waargenomen. Theoretisch gezien zijn ze een noodzakelijk gevolg van de relativiteitstheorie. Zelfs Newton heeft in zijn boek "principia" al een opmerking geplaatst die opgevat kan worden als een verklaring dat een theorie zonder gravitatiegolven niet fysisch is. Zoals we verderop zullen bespreken hebben we observationeel gezien al overtuigend bewijs dat gravitatiegolven inderdaad ook echt bestaan. Voor dit bewijs is in 1993 de Nobelprijs in de natuurkunde uitgereikt. We moeten alleen nog de juiste telescoop uitvinden om ze te bekijken.

Er zijn diverse grote projecten op dit moment bezig met de speurtocht naar gravitatiegolven. Sommige projecten proberen dit te bereiken met behulp van gebouwde detectoren. Het onderzoek in dit proefschrift betreft Pulsar Timing Arrays (PTAs), projecten om gravitatiegolven waar te nemen door naar pulsars (ronddraaiende neutronensterren die elke omwenteling een radiosignaal naar de aarde sturen) te kijken.

Zwaartekracht en gravitatiegolven

Relativiteitstheorie

Zwaartekracht is een fascinerend verschijnsel. Het is een groot onderdeel van ons dagelijks leven, en in onze belevenis is het zo normaal dat we er niet over nadenken. Sinds Newton de universele wet van de zwaartekracht opstelde kunnen we zelfs vrijwel alle dagelijkse en minder dagelijkse verschijnselen verklaren met wonderbaarlijke nauwkeurigheid, en voor lange tijd werd gedacht dat er weinig extra over zwaartekracht te ontdekken viel. De theorie was af. Maar onder de motorkap blijkt een wonderlijke wereld te schuilen die ons voorstellingsvermogen te boven gaat.

De algemene relativiteitstheorie, opgesteld door Einstein bijna een eeuw geleden, geeft naast een nieuwe beschrijving van de zwaartekracht ook een volledig

andere interpretatie van het Universum. Tijd en ruimte zijn één geheel, het zogenaamde ruimtetijd continuüm, en zijn niet onafhankelijk van elkaar te beschrijven. Dit idee, origineel geïntroduceerd door Minkowski na het bestuderen van Einsteins speciale relativiteitstheorie, geeft een meetkundige interpretatie aan het begrip tijd. Algemene relativiteitstheorie beschrijft hoe deze vierdimensionale eenheid, de ruimtetijd, vervormt door de aanwezigheid van massa en energie. Het beste voorbeeld is de planeet waar wij op wonen. De aarde is een enorme concentratie van massa, die volgens de theorie dus het ruimtetijd continuüm om zich heen beïnvloedt. Sterker nog. Alle massa, dus ook dit bescheiden proefschrift, vervormt de ruimtetijd, maar slechts in extreme gevallen zoals de aarde is deze vervorming zo sterk dat het merkbaar wordt. De vervorming van de ruimtetijd om de aarde is wat wij alledaags ervaren als zwaartekracht.

Deze beschrijving dat zwaartekracht een kromming is van de ruimtetijd vereist wat extra uitleg. Immers, hoe kan een niet-bewegend object dan toch aangetrokken worden? Het antwoord ligt erin dat onze dagelijkse ervaringen niet heel handig zijn om de werkelijkheid te beschrijven. Het is mogelijk om de ruimtetijd op zo'n manier meetkundig te beschrijven, dat alle objecten met dezelfde "snelheid" door de ruimtetijd bewegen[1]. De snelheid zoals wij die in ons dagelijks 3-dimensionale leven meten, is dan slechts een andere richting in deze vierdimensionale ruimtetijd. Net zoals voertuigen die elkaar op een kruispunt tegenkomen een ander begrip hebben van links, rechts, vooruit, en achteruit, hebben objecten met een andere richting in de ruimtetijd een ander begrip van tijd en ruimte. Alle objecten bewegen dus even snel door de ruimtetijd[2]. Een object op aarde dat niet beweegt ten opzichte van de aarde, beweegt dus nog steeds door de ruimtetijd, alleen is die beweging op dat moment in dezelfde richting door de ruimtetijd als de aarde. Deze richting is wat wij als "tijd" zouden beschrijven. Met deze beschrijving is het wat aannemelijker dat een gekromde ruimte invloed heeft op alle objecten. Immers, alles beweegt, dus een gekromde ruimte verandert de richting van elk object in die gekromde ruimte. Als je beweegt via een kromme lijn, dan verandert je richting. En als twee objecten in de ruimtetijd een andere richting hebben, dan bewegen ze ten opzichte van elkaar.

---

[1]In deze beschrijving is de snelheid door de 4-dimensionale ruimtetijd de zogenaamde 4-vector van de snelheid. Deze is gedefinieerd als $u = \frac{\mathrm{d}x}{\mathrm{d}\tau}$, waarbij x de 4-vector van de positie is, en τ is de eigentijd van het object.

[2]Deze snelheid is de lichtsnelheid. Hierbij meten we tijd in meters: het aantal meter dat een lichtstraal zou afleggen in een bepaalde tijd.

Het testen van de relativiteitstheorie

De relativiteitstheorie is uitgebreid getest de afgelopen eeuw, en tegenwoordig is het een algemeen geaccepteerde beschrijving van de realiteit. We weten nu bijvoorbeeld van metingen tijdens een totale zonsverduistering dat lichtstralen van sterren die langs de zon bewegen precies zo afgebogen worden, dat het lijkt dat de sterren op een andere plek staan, precies zoals algemene relativiteit voorspelt. We kunnen met atoomklokken meten dat voor bewegende voorwerpen de tijd inderdaad langzamer verstrijkt. Maar hoe succesvol de relativiteitstheorie ook mag zijn, we hebben goede redenen om aan te nemen dat de relativiteitstheorie niet geldig is op elke schaal.

Theoretisch gezien voorspelt de relativiteitstheorie het bestaan van singulariteiten: punten in ruimtetijd waar de kromming van de ruimtetijd oneindig groot wordt, en waar materie samengeperst is tot oneindig kleine omvang. Een voorbeeld van zo'n singulariteit is het centrum van een zwart gat. In het electromagnetisme is het probleem van singulariteiten opgelost door de wetten van de electrodynamica de verenigen met de quantum theorie, resulterend in de quantum electrodynamica (QED). De QED is de meest nauwkeurig geteste fysische theorie van dit moment, en het succes daarvan heeft wetenschappers ervan overtuigd dat iets soortgelijks met de relativiteitstheorie ook gedaan moet worden. Dit blijkt echter een zeer moeilijke opgave te zijn.

Observationeel gezien zijn er zeer overtuigende aanwijzingen dat het Universum een vorm van energie bevat, zogenaamde "donkere energie", waar wij de oorsprong nog niet van kunnen verklaren. Het is niet ondenkbaar dat dit komt doordat de algemene relativiteitstheorie niet geldig is op deze enorme schaal.

Vanwege de hierboven genoemde redenen verwachten we dus dat op een bepaalde schaal, zij het heel grote schaal, zij het heel kleine schaal, de natuur zich niet meer gedraagt volgens de wetten van de algemene relativiteitstheorie. En daarom is het nog steeds een prestigieus onderdeel van de hedendaagse wetenschap om alle aspecten van de algemene relativiteitstheorie te testen. Dit is nog steeds niet gelukt voor alle voorspellingen van de theorie, omdat een aantal effecten zéér klein zijn.

Een voorbeeld van een recent experiment dat een paar van de meest subtiele effecten van algemene relativiteitstheorie testte is Gravity Probe B (GP-B). GP-B was een NASA missie met als doel het bevestigen van het geodetisch effect en frame-dragging. Het geodetisch effect uit zich in objecten die in een baan rond een erg massief object bewegen. Bijvoorbeeld een NASA satelliet die in een baan om de aarde beweegt. De relativiteitstheorie voorspelt dat een object in een baan om de aarde een klein beetje draait bij elke omwenteling ten gevolge van de kromming van de ruimtetijd. Dit effect heet het geodetische effect, of geodetische

precessie, en is een ander effect dan de klassieke precessie die bijvoorbeeld de aarde ondergaat. Frame-dragging is het effect dat een draaiend massief object de ruimtetijd "met zich meetrekt" in zijn draaiing. Wederom is de aarde een mooi voorbeeld. Een object in de buurt van de aarde zal zeer licht om zijn as gaan draaien als gevolg van frame-dragging. GP-B probeerde beide effecten, het geodetisch effect en frame-dragging, te meten met behulp van een collectie van vier zeer nauwkeurige gyroscopen die in een satelliet in een baan om de aarde zijn gebracht. Resultaten van de voorlopige data analyse van het GP-B experiment laten zien dat zowel het geodetisch effect als frame-dragging bevestigd zijn tot op respectievelijk 0.1% en 15% nauwkeurigheid.

Gravitatiegolven

De zoektocht naar gravitatiegolven hoort ook in het rijtje thuis van effecten van algemene relativiteitstheorie die nog waargenomen moeten worden. Zoals hiervoor beschreven, is het ruimtetijd continuüm waarin wij ons bevinden een eenheid die beïnvloed wordt door de inhoud ervan. Echter, deze inhoud beweegt: alle objecten in het Universum kunnen bewegen. Bijna noodzakelijkerwijs betekent dit dat de ruimtetijd zelf ook moet bewegen. Dit is zeer sterk verbonden met het verschijnsel gravitatiegolven. Het volgende voorbeeld kan dit misschien wat verduidelijken.

In het electromagnetisme zullen twee geladen deeltjes een kracht op elkaar uitoefenen. Als voorbeeld beschouwen we een positief geladen deeltje, B, en een negatief geladen deeltje op een antenne, A (zie figuur 6.2). Als we B naar boven bewegen, dan zal vanwege de aantrekkingskracht tussen A en B, ook A naar boven bewegen op de antenne. De vraag is nu: hoe lang duurt het voordat A gaat bewegen, nadat we B naar boven hebben bewogen? Gebeurt dat direct, of verstrijkt er eerst wat tijd voordat A wat merkt van de nieuwe positie van B? Het antwoord is dat het electromagnetisch veld zich met de lichtsnelheid aan zal passen, en deze aanpassing in het electromagnetisch veld zal zich als een golf door de ruimte (of beter gezegd: ruimtetijd) uitbreiden. Als we B nu continu, als een oscillator, heen en weer blijven bewegen, dan zal A dat ook gaan doen zodra de golf de antenne bereikt heeft. B zendt nu electromagnetische golven uit die zich met de lichtsnelheid richting A bewegen. A wordt vervolgens beïnvloed door de electromagnetische golven, en gaat ook bewegen.

Net zoals de electromagnetische golven uit het vorige voorbeeld golven zijn in het veld dat de kracht tussen twee geladen deeltjes veroorzaakt, zijn gravitatiegolven in algemene relativiteitstheorie golven in datgene dat zwaartekracht veroorzaakt tussen twee massa's. Gravitatiegolven zijn dus golven in de ruimtetijd. Anders
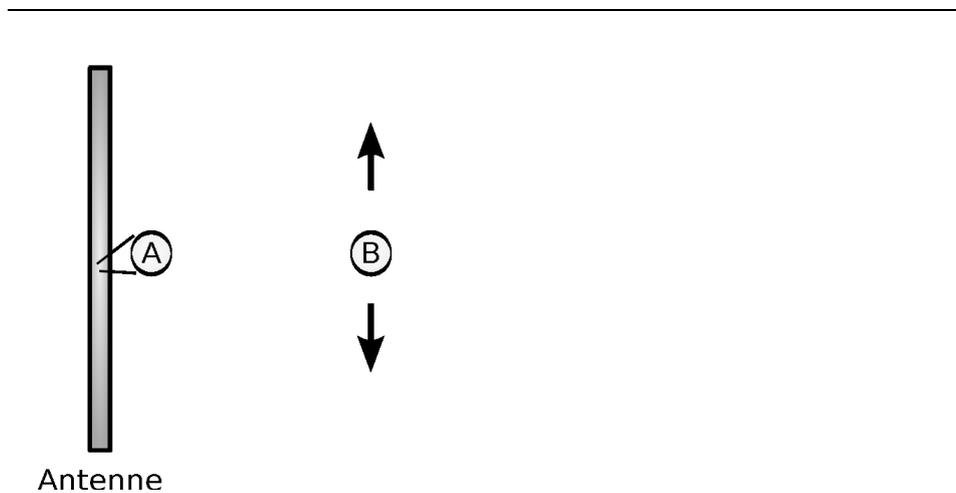
Antenne

Figure 6.2: Twee geladen deeltjes. Een negatief geladen deeltje, A, op een antenne, en een positief geladen deeltje, B, dat electromagnetische golven uitzendt.

gezegd, gravitatiegolven zijn vervormingen van de ruimtetijd, veroorzaakt door bewegende massa en energie, die zich door het Universum voortbewegen met de lichtsnelheid. Net als electromagnetische golven hebben gravitatiegolven een richting, en een golflengte.

Zwarte gaten

Zwarte gaten zijn objecten waarvan het zwaartekrachtsveld zo sterk is, ofwel, waarvan de kromming van de ruimtetijd zo groot is, dat niets er uit kan ontsnappen zodra het erin is gevallen. Zwarte gaten ontstaan zodra er een enorme hoeveelheid massa in een zéér kleine ruimte geperst wordt. Typisch kan dit gebeuren wanneer een erg massieve ster al zijn brandstof heeft opgebruikt, en aan het eind van zijn leven in elkaar stort onder zijn eigen zwaartekracht. Voor elk zwart gat is een afstand tot het centrum uit te rekenen, de zogenaamde Schwarzschild straal, vanaf waar het voor een invallend object niet meer mogelijk is om terug te keren zodra dit punt gepasseerd is. Deze afstand tot het zwarte gat is het "point of no return". Ter illustratie: als we de aarde in elkaar zouden persen tot een zwart gat, dan zou de Schwarzschild straal op 8.9mm afstand van het centrum liggen.

Hoewel het bestaan van zwarte gaten lange tijd onzeker is geweest, hebben we tegenwoordig zeer overtuigend bewijs van hun bestaan. Het mooiste voorbeeld daarvan is het centrum van ons eigen sterrenstelsel, de Melkweg. We kunnen de positie van de sterren in de buurt van het centrum van de Melkweg zeer nauwkeurig waarnemen, en uit de afgelegde banen van deze sterren blijkt dat ze allemaal om één zeer massief, en zeer klein onzichtbaar object bewegen. De massa van dit

object is ongeveer 4 miljoen zonsmassa's. Er is dus een onzichtbaar object dat 4 miljoen keer zo zwaar is als onze zon, op een zeer kleine hoeveelheid ruimte geperst[3]. Dit soort enorme zwarte gaten worden supermassieve zwarte gaten genoemd, en ze worden verondersteld in de centra van (bijna) alle sterrenstelsels voor te komen.

Supermassieve zwarte gaten zijn van belang voor PTAs, omdat aangenomen wordt dat deze objecten verantwoordelijk zijn voor gravitatiegolven die waarneembaar zijn met PTAs. Sterrenstelsels ontstaan in het algemeen in groepen, en gedurende de evolutie van het Universum hebben veel van die sterrenstelsels ontmoetingen met elkaar, vaak zelfs met botsingen tot gevolg. Een botsing tussen sterrenstelsels kan na een gecompliceerde interactie leiden tot een samensmelting van de twee sterrenstelsels. Na deze vereniging is een nieuw sterrenstelsel geboren. De supermassieve zwarte gaten van de twee oorspronkelijke sterrenstelsels zullen door interacties met hun omgeving naar het centrum van het nieuwgevormde sterrenstelsel zinken, waar ze elkaar dan tegenkomen. We hebben aanwijzingen dat dit scenario een zeer vaak voorkomend verschijnsel is in de evolutie van het Universum.

Twee zwarte gaten die elkaar dicht genoeg genaderd zijn zullen onder invloed van elkaars zwaartekracht een binair systeem vormen, waarbij de twee in een baan om elkaar heen bewegen. In dit proces worden gravitatiegolven uitgezonden. Hoewel dit in principe voor alle objecten geldt, niet alleen voor zwarte gaten, is het uitzenden van gravitatiegolven juist voor binaire zwarte gaten van belang voor hun evolutie. Dit komt vanwege de extreme massa's van de twee objecten. Hoe zwaarder een object, hoe groter zijn gravitatieveld, ofwel, hoe sterker de ruimtetijd gekromd is om het object. Hoe sterker de kromming van de ruimtetijd is tijdens de beweging van de twee zwarte gaten, hoe meer gravitatiegolven door het systeem wordt uitgezonden. Door het uitzenden van gravitatiegolven komen de zwarte gaten nog dichter bij elkaar, totdat ze op een gegeven moment op elkaar botsen en daarbij samen een nog groter zwart gat vormen.

De PTAs zijn gevoelig voor gravitatiegolven van de interactie tussen supermassieve zwarte gaten. Twee verschillende processen zijn van belang. Als eerste zorgt het om elkaar heen bewegen van de zwarte gaten voor een signaal voor PTAs. Als de twee objecten zo dicht bij elkaar zijn gekomen dat het de zwarte gaten ongeveer een paar maanden tot een paar jaar kost om één keer om elkaar heen te bewegen, dan zenden ze gravitatiegolven uit met frequenties die optimaal zijn voor PTAs om te detecteren. Een complicatie is echter dat het niet slechts één paar su-

---

[3]We kunnen niet exact zeggen hoe groot dit object is, omdat we het niet direct kunnen waarnemen. Wat we wel hebben kunnen afleiden is dat dit object zich op een hoeveelheid ruimte bevindt kleiner dan ons zonnestelsel.

permassieve zwarte gaten is in het Universum dat een signaal uitzendt voor PTAs om te detecteren. Volgens onze huidige voorspellingen zijn het er zo veel, dat we niet kunnen zien aan de observaties hoeveel systemen de signalen uitzenden, en waar ze zijn. Wat we slechts kunnen zien is een "achtergrond" van golven, net zoals het in een zwembad niet meer te zien is welke zwemmer welk golfje heeft geproduceerd.

Een tweede gravitatiegolf effect dat door samensmeltende zwarte gaten geproduceerd wordt is het gravitational-wave memory effect. Dit effect is een gevolg van het daadwerkelijke botsen en verenigen van de zwarte gaten. Laten we het vergelijken met een gewichtje dat aan een veer hangt. In eerste instantie hangen we het gewichtje stil aan de veer. Dan trekken we het gewichtje flink naar beneden, en door de veerkracht gaat het gewichtje op en neer bewegen. Na een tijdje komt het gewichtje weer tot stilstand, en het gewichtje hangt op dezelfde hoogte als in het begin. Echter, we kunnen in het begin ook erg hard aan het gewichtje trekken, wat de veer permanent een beetje zal uitrekken. Het gewichtje zal dan naderhand iets lager komen te hangen dan in het begin. Dit is te vergelijken met wat er met de ruimtetijd gebeurt na een samensmelting van twee zwarte gaten: de ruimtetijd zal zich niet herstellen tot dezelfde toestand als voorheen. In plaats daarvan zal de ruimtetijd in bepaalde richtingen een beetje "uitgerekt" en in andere richtingen een beetje "samengeperst" blijven. Dit permanente effect blijkt meetbaar te zijn met PTAs.

Pulsars en pulsar timing

Pulsars

Pulsar Timing Arrays zijn projecten die gebruik maken van radiogolf observaties van pulsars: een speciale klasse van neutronensterren. Neutronensterren kunnen op twee verschillende manieren ontstaan:
1) Als een erg massieve ster aan het eind van zijn leven door zijn brandstof heen is, zal deze onder invloed van zijn eigen zwaartekracht in elkaar storten. Bij dit in elkaar storten komt zoveel energie vrij dat de buitenste lagen van de ster met een enorme ontploffing worden uitgestoten. Dit is een zogenaamde supernova explosie, en in zeer korte tijd wordt meer energie uitgestraald dan door de zon wordt uitgezonden gedurende zijn hele leven (zo'n 10 miljard jaar in totaal). De kern van de in elkaar gestortte ster blijft over, en deze kern noemen we een neutronenster of bij een zwaardere ster een zwart gat.
2) Sterren die minder massief zijn dan ongeveer 9 zonsmassa's zullen niet ontploffen als een supernova. Deze sterren zullen aan het eind van hun leven krimpen tot een zogenaamde witte dwerg: een zeer warm compact object. Als een witte

dwerg in een dubbelster systeem staat, waarbij dus beide sterren om elkaar heen bewegen, kan het voorkomen dat de witte dwerg materie van zijn begeleidende ster naar zich toetrekt. Dit kan bijvoorbeeld gebeuren als de begeleidende ster in een rode reus verandert, en dus enorm in grootte toeneemt. Als de witte dwerg daardoor zwaarder wordt dan ongeveer 1.4 zonsmassa's, dan zal dit een kettingreactie in de witte dwerg veroorzaken waardoor deze ontploft als een supernova. Ook het resultaat van deze ontploffing zal een neutronenster zijn.

Neutronensterren zijn zware en erg compacte objecten. Een typische neutronenster is ongeveer anderhalve zonsmassa's zwaar, en heeft een straal van zo'n 10 kilometer. Omdat neutronensterren zijn begonnen als een enorme ster, en daarna zijn gekrompen tot slechts een heel klein deel van hun originele grootte, hebben neutronensterren naast hun enorme masse nog andere extreme eigenschappen. Allereerst draaien ze snel rond. Net zoals een kunstschaatser die zijn armen intrekt sneller gaat ronddraaien, zo is ook een neutronenster sneller gaan ronddraaien tijdens het vormingsproces. Verder blijft bij het vormen van neutronensterren hun totale magnetische flux behouden, wat betekent dat de sterkte van het magneetveld door de samenpersing enorm is geworden. De combinatie van een zeer sterk magneetveld, en het snel ronddraaien, zorgt ervoor dat neutronensterren een bundel electromagnetische golven uitzenden elke keer dat ze ronddraaien. Voor sommige neutronensterren is deze bundel zichtbaar vanaf de aarde. Deze neutronensterren noemen we pulsars. Pulsars zijn dus het kosmische equivalent van een vuurtoren, waarbij we een electromagnetische puls ontvangen elke keer dat de pulsar ronddraait.

Pulsar timing

Als we de pulsen die uitgezonden worden door pulsars met een telescoop waarnemen, dan valt het gelijk op dat de individuele pulsen altijd van elkaar verschillen. Geen twee zijn hetzelfde. Maar als we de pulsen een lange tijd blijven observeren kunnen we er een systeem in ontdekken. Het blijkt dat het gemiddelde van heel veel pulsen altijd heel stabiel is. Een voorbeeld hiervan is gegeven in figuur 6.3. Deze gemiddelde puls is van groot belang voor pulsar timing, omdat deze gemiddelde puls zelfs over een tijdsspanne van vele jaren zo goed als niet verandert. Deze gemiddelde puls heeft over het algemeen een heel specifieke vorm en is uniek voor elke pulsar. De reden waarom deze gemiddelde puls zo stabiel is, en waar deze vorm door bepaald wordt is nog onderdeel van huidig onderzoek. Maar deze stabiele gemiddelde pulsen worden al wel decennia lang door de wetenschap gebruikt als een bron van informatie; in feite zijn pulsars namelijk met hun stabiele rotatie en stabiele gemiddelde puls heel erg nauwkeurige klokken, die overal in het Universum staan.
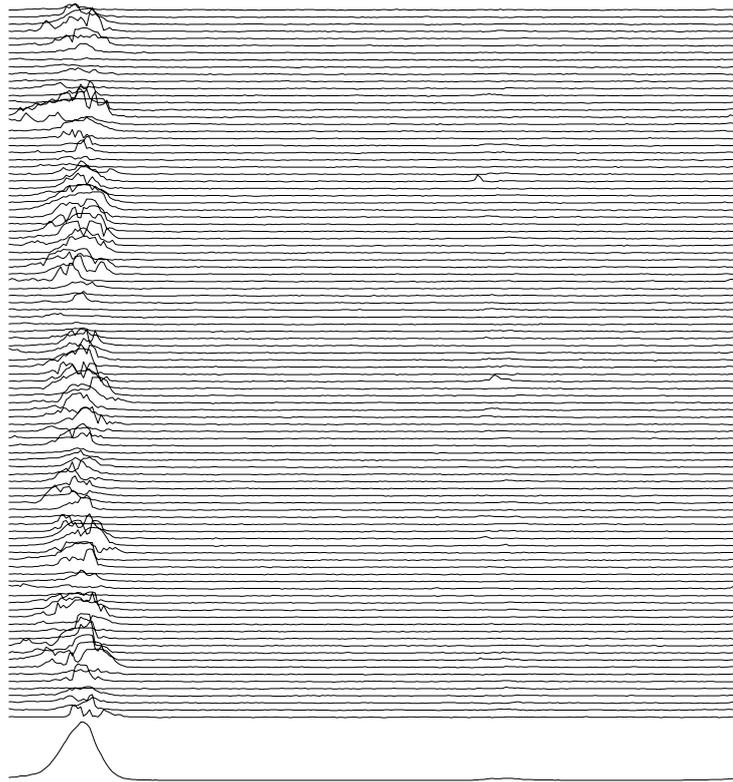
Figure 6.3: Een figuur met de pulsen van de 253 milliseconde pulsar B0950+08. We zien hier de intensiteit van de pulsen ten opzichte van de fase van de omwenteling van de pulsar. Een verzameling enkele pulsen zijn te zien aan de bovenkant van de figuur, wat laat zien dat de pulsen sterk variëren in termen van puls profiel en intensiteit. Een gemiddeld puls profiel van 5 minuten is te zien aan de onderkant van de figuur. De ongeveer 1200 pulsen benaderen het reproduceerbare standaardprofiel van deze pulsar al redelijk. Deze observaties zijn genomen met de Green Bank Telescoop in West Virginia (Verenigde Staten). Figuur komt van Stairs (2003), gereproduceerd met toestemming.

Voor pulsar timing doeleinden, en voor PTAs in het bijzonder, is een bepaalde soort pulsar extra geschikt: de zogenaamde milliseconde pulsars (MSPs). Dit zijn oude pulsars die door het aantrekken van materie van een begeleidende ster zeer snel zijn gaan ronddraaien. Waar het voor een normale pulsar typisch een seconde duurt voordat deze één keer is rondgedraaid om zijn as, kost dat een MSP slechts een paar milliseconden. De snelst ronddraaiende MSP die tot nu toe gevonden is, genaamd PSR J1748-2446, draait per seconde 716 keer om zijn as. Op de evenaar van deze pulsar is de ronddraaisnelheid ongeveer 24% van de lichtsnelheid.

Pulsar timing houdt zich bezig met het gebruik van deze nauwkeurige klokken als meetinstrumenten. Typisch wordt een MSP slechts een keer per paar weken geobserveerd, waarbij een stabiele gemiddelde puls wordt gemeten. Met deze gemiddelde puls kan in de beste gevallen tot minder dan 100ns (één tien miljoenste van een seconde) bepaald worden wanneer de pulsen op aarde aankomen. Ter vergelijking, 100ns is de tijd die het licht nodig heeft om 30m af te leggen. Door deze extreme nauwkeurigheid, en door de stabiliteit van de MSPs, is het een keer per paar weken observeren genoeg. Om een voorbeeld te geven: stel we zijn een half jaar lang niet in staat om een bekende MSP met een ronddraaisnelheid van 500Hz te observeren, en na een half jaar kijken we er weer naar met een telescoop. Hoewel deze pulsar in de tussentijd bijna 8 miljard keer is rondgedraaid zonder dat we ernaar gekeken hebben, kunnen we *exact* zeggen hoeveel keer dit is geweest.

Door dit gebruik van MSPs als nauwkeurige klokken met een heel goed bepaalde frequentie, kunnen we hun beweging heel goed in kaart brengen. Dit werkt hetzelfde als het meten van de snelheid van auto's op de snelweg met een lasergun. Sommige pulsars staan in zeer extreme omgevingen, zoals bijvoorbeeld de pulsars die in dubbelster systemen staan en elke paar uur om een andere neutronenster heen bewegen. In dit soort systemen is de kromming van de ruimtetijd zodanig groot, dat het nauwkeurig in kaart brengen van de baan van de pulsar het mogelijk maakt om de algemene relativiteitstheorie te testen. Dit is precies waar Hulse & Taylor in 1993 de Nobelprijs in de natuurkunde voor hebben gekregen. Hulse & Taylor hebben met het timen van de binaire pulsar PSR B1913+16, meestal de Hulse-Taylor binary pulsar genoemd, laten zien dat algemene relativiteit dit systeem uiterst goed beschrijft, en dat het inkrimpen van de baan van dit systeem komt door het uitzenden van gravitatiegolven.

Pulsar Timing Arrays

Zoals hiervoor besproken is een gravitatiegolf een vervorming van de ruimtetijd die zich door het Universum verplaatst met de lichtsnelheid. Het effect van een gravitatiegolf is het uitrekken en inkrimpen van de ruimtetijd, wat ook invloed heeft op de voortplanting van electromagnetische golven zoals de pulsen van een MSP. Een

puls van een MSP zal dus iets eerder of iets later aankomen dan verwacht onder invloed van een gravitatiegolf, en dit is precies wat we meten met pulsar timing: de aankomsttijd van de pulsen. Het effect van de gravitatiegolven is dus te zien aan de observaties van MSPs, en het is nu de kunst om dit effect er zo goed mogelijk uit te halen. Dit is wat geprobeerd wordt met de PTA projecten. Omdat de gravitatiegolven tot nu toe nog niet gemeten zijn, zijn PTAs in principe altijd bezig met de volgende drie dingen:

1) De instrumenten hier op aarde verbeteren zodat de aankomsttijden van de pulsen beter bepaald kan worden.

2) Meer MSPs vinden, en vaker waarnemen. Meer data betekent meer signaal. En we hebben nog lang niet alle MSPs in ons Melkwegstelsel gevonden; er is grote hoop dat nog meer zeer stabiele MSPs worden gevonden in de nabije toekomst.

3) Betere data analyse methoden ontwikkelen. De data analyse van pulsar timing is vrij complex, omdat er met erg veel zaken rekening gehouden moet worden. Zeer veel bekende en onbekende processen beïnvloeden de metingen, en deze moeten op de juiste manier in rekening gebracht worden om de gravitatiegolven ook daadwerkelijk uit de waarnemingen te kunnen halen.

Dit proefschrift

In dit proefschrift zetten we een methode uiteen om de waarnemingen van MSPs te analyseren met als doel daar een signaal van gravitatiegolven in te ontdekken. De data analyse van PTAs is gecompliceerd, en de alternatieve methoden die bestonden voor de ontwikkeling van het werk in dit proefschrift hielden niet voldoende rekening met alle processen die de waarnemingen beïnvloeden. Met de methode uit dit proefschrift kunnen de waarnemingen geanalyseerd worden, waarbij met de volgende complicaties correct rekening gehouden wordt:

1) Waarnemingen worden sporadisch gedaan voor MSPs. Typisch eens in de paar weken, maar dit hangt af van het waarneemrooster op de verschillende telescopen. De waarnemingen zijn niet regelmatig.

2) Een waarneming bestaat in eerste instantie uit data van typisch een half uur. Dit wordt in een ingewikkeld proces omgezet tot slechts één datapunt: een aankomst tijd van de gemiddelde puls. Het proces dat hiervoor zorgt kan variëren per telescoop, en het is niet gegarandeerd dat deze procedure optimaal is verlopen.

3) Bepaalde effecten worden bepaald uit dezelfde waarnemingen als waar we de gravitatiegolven proberen uit te halen. De MSPs bewegen bijvoorbeeld ten opzichte van de aarde: ze bewegen door ons Melkwegstelsel, vaak staan ze in een systeem met een begeleidende ster waar ze omheen bewegen, en de aarde beweegt om de zon. Deze bewegingen worden in kaart gebracht met de timing waarnemin-

gen. Ook wordt de rotatiesnelheid van de MSPs bepaald uit deze observaties.

4) Zoals eerder besproken is een belangrijke bron van gravitatiegolven de achtergrond van gravitatiegolven zoals veroorzaakt door een combinatie van heel veel supermassieve zwarte gaten. Hoewel we de vorm van dit signaal vrij goed kunnen beschrijven, is het nog steeds nodig om een deel van dit signaal te karakteriseren met de PTA observaties. Dit is een complexe procedure.

5) Hoewel de MSPs zich als zeer stabiele klokken gedragen, is er nog onzekerheid over hun lange termijn gedrag op het gebied van stabiliteit. Verwacht wordt dat ze over lange tijdschalen minder stabiel zijn, en het in kaart brengen van dit lange termijn gedrag is belangrijk.

6) Om een zo groot mogelijke gevoeligheid te hebben voor gravitatiegolven moeten alle waarnemingen gebruikt worden die we tot onze beschikking hebben. Dit zullen altijd observaties van meerdere verschillende telescopen zijn, en zelfs bij datasets van één telescoop kan het voorkomen dat apparatuur op de telescoop veranderd is gedurende de loop van het experiment. Alleen al de verandering van de lengte van één kabel heeft al invloed op de resultaten, en al deze complicaties moeten correct behandeld worden.

Het testen van de methode

In hoofdstuk 2 beschrijven we de methode waarmee alle bovenstaande problemen correct worden aangepakt, en waarmee een achtergrond van gravitatiegolven kan worden gedetecteerd. We testen deze methode op meerdere manieren op gesimuleerde observaties, en deze simulaties leveren de volgende resultaten op:

1) Met de methode kunnen zoals verwacht gravitatiegolven gemeten worden.

2) Het is zeer belangrijk om de lange termijn stabiliteit van de MSPs te modelleren en in de analyse methode mee te nemen. De gevoeligheid voor gravitatiegolven is daar zeer sterk van afhankelijk.

3) We geven grafische overzichten van de gevoeligheid voor gravitatiegolven als functie van verschillende parameters, zoals het aantal waargenomen pulsars, en de nauwkeurigheid waarmee de pulsars te timen zijn.

Deze resultaten zijn zoals we hadden verwacht. Het enige nadeel van de methode tot zo ver is dat er erg veel rekenkracht nodig is om de berekeningen uit te voeren. Dit probleem moet in de toekomst opgelost worden, omdat we verwachten dat de datasets van de toekomst veel meer observaties zullen bevatten. De rekenkracht die nodig is om een dataset te analyseren hangt sterk af van de hoeveelheid waarnemingen.

Gravitational-wave memory

Zoals hiervoor al beschreven is het gravitational-wave memory effect ook in principe meetbaar met PTAs. Dit is pas recent gerealiseerd door drie onafhankelijke groepen. Wij waren één van die groepen, en hoofdstuk 3 is onze beschrijving hiervan. In dit hoofdstuk laten we zien dat PTAs inderdaad dit signaal kunnen meten, en hoe dit signaal er in de data uit zou zien. De technieken in dit hoofdstuk zijn gebaseerd op de methode uit hoofdstuk 2, waardoor we dus van dezelfde voordelen kunnen profiteren bij de zoektocht naar het gravitational-wave memory effect in de waarnemingen. We analyseren de gevoeligheid van de PTAs voor het gravitational-wave memory effect grondig, en we testen dit vervolgens op gesimuleerde data. Gebaseerd op simulaties van de evolutie van het Universum voorspellen we dat de kans dat we met PTAs samensmeltingen van supermassieve zwarte gaten zullen zien de komende jaren vrij laag is: zo tussen de 1% en 10% als we aannemen dat we ongeveer 10 jaar aan waarnemingen tot onze beschikking hebben. Maar het is niet ondenkbaar dat onze simulaties van het Universum iets afwijken van de realiteit, wat dit gravitational-wave memory een signaal maakt dat het waard is om naar te zoeken.

Analyse van Europese PTA data

De methode die we in hoofdstuk 2 introduceren is uitermate geschikt om echte waarnemingen mee te analyseren, zelfs als daar nog geen detecteerbaar signaal in zit. In het geval dat we (nog) niets vinden in de waarnemingen, noemen we het resultaat een upper-limit. Een upper-limit is een bovengrens voor de sterkte van het signaal dat eventueel toch in de data verstopt zit. Het enige dat we dan kunnen zeggen over dit signaal, is dat áls het er is, dat het niet groter is dan deze waarde.

In hoofdstuk 4 beschrijven we hoe we de waarnemingen van de Europese Pulsar Timing Array (EPTA) analyseren met onze methode. Een van de grote uitdagingen bij het verwerken van de data van de EPTA is het feit dat deze waarnemingen met meerdere telescopen, en met verschillende apparatuur en verschillende verwerkingsmethoden, geproduceerd zijn. We hebben nauw samengewerkt met alle EPTA groepen om alle verschillen tussen de verschillende observatoria goed in kaart te brengen, en om een algemene manier op te stellen om hier in de data analyse rekening mee te houden. In hoofdstuk 4 beschrijven we hoe we dit gedaan hebben, en in principe is dit werk dus een recept voor andere PTA groepen hoe dit op hun datasets ook gedaan kan worden. We laten zien dat áls er een achtergrond van gravitatiegolven bestaan, dat het signaal niet sterker is dan de waarde $h_c \leq 6 \times 10^{-15}$. Dit is nog net boven het regime waar we de achtergrond verwachten op basis van simulaties van het Universum, dus vooral de komende jaren gaan

spannend worden. We zullen namelijk door middels van meer en nauwkeurigere waarnemingen een grotere gevoeligheid voor gravitatiegolven hebben.

Het berekenen van de marginal likelihood

De analyse methoden die we introduceren in dit proefschrift zijn voorbeelden van een zogenaamde Bayesiaanse analyse. Bayesiaanse analyse is een manier om data te analyseren die gebaseerd is op het theorema van Bayes: een zeer algemene manier van data analyse waarbij volgens stricte regels van de waarschijnlijkheidsleer gewerkt wordt. Bayesiaanse analyse heeft veel voordelen, waaronder dat zeer complexe modellen moeiteloos behandeld kunnen worden. Hoewel deze benadering van data analyse al bekend was in de eerste helft van de 20ste eeuw, is het pas echt populair geworden rond de afgelopen eeuwwisseling. Dit komt doordat er zeer veel rekenkracht nodig is om dit type methode toe te passen. Zo ook voor de methode die beschreven is in de hoofdstukken 2, 3, en 4. Om de upper-limit te produceren op de achtergrond van gravitatiegolven uit hoofdstuk 4 hebben we ongeveer 1000 uur aan rekenkracht gebruikt.

In het algemeen, als men aan data analyse doet, dan neemt men eerst een model aan. Gegeven dat het model correct is, kan men met de waarnemingen of metingen iets zeggen over de eigenschappen van dat model. Maar het is niet altijd duidelijk of dat model wel de goede beschrijving geeft van de werkelijkheid. Misschien is er wel een ander model dat de werkelijkheid beter beschrijft. Daarom zijn er in het algemeen, en dus ook in de Bayesiaanse analyse, twee typen resultaten die belangrijk zijn:

1) De geschatte waarden van de model parameters aan de hand van de data. Hierbij wordt aangenomen dat het model correct is.

2) Een zogenaamde goodness-of-fit waarde, in de Bayesiaanse analyse meestal de marginal likelihood genoemd. Deze waarde kan gebruikt worden om te bepalen welk model de data beter verklaart als we niet zeker weten wel model de meest correcte beschrijving is van de werkelijkheid.

Vooral het uitrekenen van (2) is in de Bayesiaanse analyse een zeer grote uitdaging, en deze waarde hebben we nog niet uitgerekend voor de waarnemingen van de EPTA, omdat dat teveel rekenkracht zou kosten. Hoofdstuk 5 stelt een nieuwe rekenmethode voor om de marginal likelihood uit te rekenen met de resultaten van de rekenmethode die gebruikt is in de hoofdstukken 2, 3, en 4. Deze nieuwe rekenmethode vergt dus bijna geen extra rekenkracht, en zou een zeer goede kandidaat zijn om op PTA data toe te passen. We testen deze nieuwe methode in hoofdstuk 5 op meerdere manieren, waaronder op een set gesimuleerde waarnemingen van PTAs. Bij vergelijkingen met andere rekenmethoden vinden we goede overeenkomst van de resultaten. Echter, voordat we deze rekenmethode met

vertrouwen kunnen toepassen op echte datasets, moeten we een uitgebreidere analyse van het algoritme doen. Deze nieuwe methode moet op meer verschillende manieren getest worden voordat we deze methode kunnen aanvaarden als een volwaardig alternatief. Dit is nog werk voor de toekomst.

Toekomstig werk

Pulsar Timing Arrays en de data analyse daarvan begint een volwassen wetenschap te worden. De afgelopen tien jaar hebben we een zeer grote sprong voorwaarts gemaakt, en we weten steeds beter wat we nodig hebben om een echte gravitatiegolf detectie te maken. We zijn echter nog niet klaar.

Aan de kant van de waarnemingen valt nog veel te verbeteren. Nieuwe technologie zullen zorgen voor nauwkeurigere meetinstrumenten, en betere zoekprojecten met radiotelescopen zullen zorgen voor meer bekende milliseconde pulsars. Dit zal zorgen voor meer data, en voor meer begrip van pulsars. We begrijpen pulsars deels, maar nog niet genoeg om alle eigenschappen van de pulsars te verklaren.

Aan de kant van de data analyse, waar dit proefschrift voornamelijk over gaat, zijn ook nog een aantal specifieke verbeteringen mogelijk:

1) Hoewel we met de Bayesiaanse analyse een signaal kunnen detecteren, moeten we om de gehele wetenschappelijke gemeenschap te overtuigen dat het niets anders geweest kan zijn dan een gravitatiegolf nog wat extra moeite doen.

2) De hoeveelheid rekenkracht die we nodig hebben is te groot om ook toekomstige datasets aan te pakken. Het algoritme moet aangepast worden om sneller dezelfde resultaten te produceren.

3) Behalve voor gravitational-wave memory en de achtergrond van gravitatiegolven moet het algoritme ook in staat zijn om enkele inspirals van supermassieve zwarte gaten te detecteren. Mocht er een paar supermassieve zwarte gaten dichtbij genoeg staan dat dit paar zichtbaar is, dan moeten we daarop voorbereid zijn.

4) De methode uit hoofdstuk 5 moet uitgebreider getest en geanalyseerd worden.

Er is dus nog genoeg te doen in de nabije toekomst. Maar we zijn erg optimistisch over het potentieel van Pulsar Timing Arrays, en we weten goed wat we moeten doen om deze projecten tot een succes te maken. Misschien komt het zelfs tot een eerste detectie ooit van gravitatiegolven!
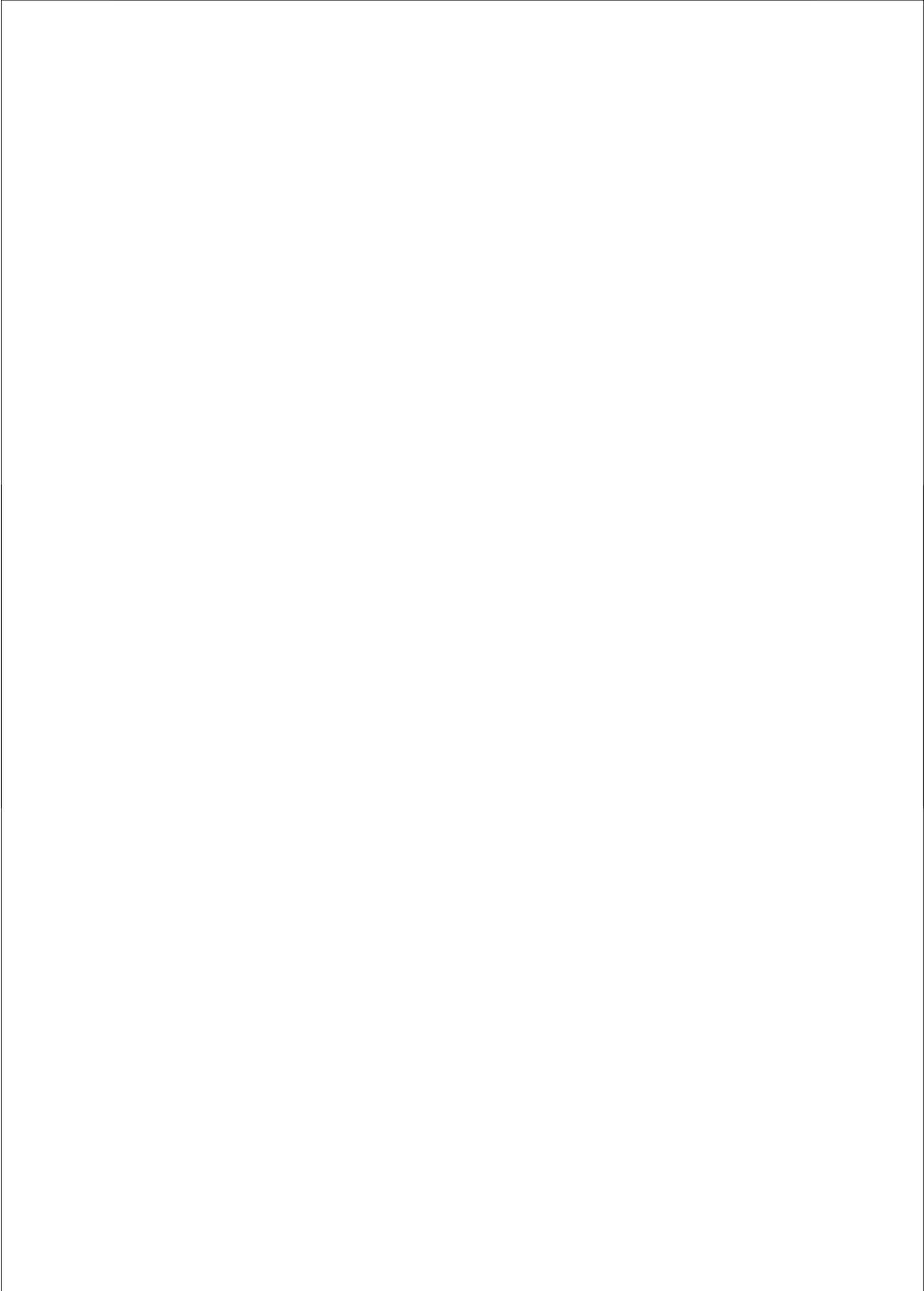
Bibliography

Abadie J., Abbott B. P., Abbott R., Abernathy M., Accadia T., Acernese F., Adams C., Adhikari R., Ajith P., Allen B., et al. 2010, Phys. Rev. D, 82, 102001

Abbott B. P. et al. (LIGO Scientific Collaboration), 2009, Reports on Progress in Physics, 72, 076901

Acernese F. et al. (Virgo Collaboration), 2006, Class. Quantum Grav., 23, S635-S642

Alpar M. A., Cheng A. F., Ruderman M. A., Shaham J., 1982, Nature, 300, 728

Anholm M., Ballmer S., Creighton J. D. E., Price L. R., Siemens X., 2009, Phys. Rev. D, 79, 8

Backer D. C., Kulkarni S. R., Heiles C., Davis M. M., Goss W. M., 1982, Nature, 300, 615

Barr C.D., 2008, PhD thesis: "Applications of Voronoi Tessellations in Point Pattern Analysis"

Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307

Berczik P., Merritt D., Spurzem R., 2005, ApJ, 633, 680

Blanchet L., Damour T., 1992, Phys. Rev. D, 46, 4304

Brady P. R., Creighton J. D. E., Thorne K. S., 1998, Phys. Rev. D, 58, 061501

Brent R., 1973, Algorithms for Minimization without Derivatives. Prentice-Hall, Englewood Cliffs, New Jersey

Britton M. C., 2000, ApJ, 532, 1240

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

Christodoulou D., 1991, Physical Review Letters, 67, 1486

Conklin J. W., the Gravity Probe B Collaboration 2008, Journal of Physics Conference Series, 140, 012001

BIBLIOGRAPHY

Cordes J. M., Kramer M., Backer D. C., Lazio T. J. W., Science Working Group for the Square Kilometer Array Team 2005, in Bulletin of the American Astronomical Society Vol. 37 of Bulletin of the American Astronomical Society, Key Science with the Square Kilometer Array: Strong-field Tests of Gravity using Pulsars and Black Holes. pp 1390–+

Cordes J. M., Shannon R. M., 2010, e-print arXiv:1010:3785v1

Crowder J., Cornish N. J., 2005, Phys. Rev. D, 72, 083005

Damour T., Deruelle N., 1986, Ann. Inst. Henri Poincaré Phys. Théor., Vol. 44, No. 3, p. 263 - 292, 44, 263

Cuadra J., Armitage P. J., Alexander R. D., Begelman M. C., 2009, MNRAS, 393, 1423

Damour T., Vilenkin A., 2005, Phys. Rev. D, 71, 063510

Demorest P. B., 2007, PhD thesis, University of California, Berkeley

Detweiler S., 1979, ApJ, 234, 1100

Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956

Earl D. J., Deem M. W., 2005, Physical Chemistry Chemical Physics (Incorporating Faraday Transactions), 7, 3910

Edelsbrunner H., Shah N. R., 1996, Algorithmica, 15, 223

Efron B., 1979, Ann. of Stat., 7, 1

Estabrook F., Wahlquist H., 1975, Gen. Relativ. Gravitation, 6, 439

Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152

Evans M. J., 2006, in J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. West ed., Proceedings of the Eighth Valencia International Meeting Discussion of Nested Sampling for Bayesian Computations by John Skilling. pp 507–512

Favata M., 2009, Phys. Rev. D, 80, 024002

Ferdman R. D., van Haasteren R., Bassa C. G., Burgay M., Cognard I., Corongiu A., D'Amico N., Desvignes G., Hessels J. W. T., Janssen G. H., Sanidas S., Smits R., Theureau G., 2010, Classical and Quantum Gravity, 27, 084014

156

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Finn L. S., 1992, Phys. Rev. D, 46, 5236

Flanagan É. É., Hughes S. A., 1998, Phys. Rev. D, 57, 4535

Foster R., Backer D., 1990, ApJ, 361, 300

Ghez A. M., Salim S., Weinberg N. N., Lu J. R., Do T., Dunn J. K., Matthews K., Morris M. R., Yelda S., Becklin E. E., Kremenek T., Milosavljevic M., Naiman J., 2008, ApJ, 689, 1044

Gillessen S., Eisenhauer F., Trippe S., Alexander T., Genzel R., Martins F., Ott T., 2009, ApJ, 692, 1075

Gillessen S., Eisenhauer F., Fritz T. K., Bartko H., Dodds-Eden K., Pfuhl O., Ott T., Genzel R., 2009, ApJ, 707, L114

Grishchuk L. P., 2005, Uspekhi Fizicheskikh Nauk, 48, 1235

Grote H. and LIGO Scientific Collaboration, 2008, Class. Quantum Grav. 25 114043

Haehnelt M. G., 1994, MNRAS, 269, 199

Haensel P., Potekhin A. Y., Yakovlev D. G., eds, 2007, Neutron Stars 1 : Equation of State and Structure Vol. 326 of Astrophysics and Space Science Library

Hamerly G., Elkan C., 2003, in Proceedings of the 17-th annual conference on Neural Information Processing Systems (NIPS) pp 281–288

Hellings R., Downs G., 1983, ApJ, 265, L39

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

Hobbs G. B., Bailes M., Bhat N. D. R., Burke-Spolaor S., Champion D. J., Coles W., Hotan A., Jenet F., et al. 2009, Publications of the Astronomical Society of Australia, 26, 103

Hobbs G. B., Edwards R. T., 2006, TEMPO2 user manual, Version 2.0

Hobbs G. B., Edwards R. T., Manchester R. N., 2006, MNRAS, 369, 655

BIBLIOGRAPHY

Hobbs, G. B. et al., 2010, Classical and Quantum Gravity, 27, 084013

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

Jaffe A., Backer D., 2003, ApJ, 583, 616

Janssen G. H., 2009, PhD thesis, University of Amsterdam

Jenet F., Hobbs G., Lee K., Manchester R., 2005, ApJ, 625, L123

Jenet F., Hobbs G., van Straten W., Manchester R., Bailes M., Verbiest J., Edwards R., Hotan A., Sarkissian J., Ord S., 2006, ApJ, 653, 1571

Jenet F. et al., 2009, eprint arXiv:0909.1058v1

Kaspi V. M., Taylor J. H., Ryba M. F., 1994, ApJ, 428, 713

Kerr R. P., 1963, Phys. Rev. Lett., 11, 237

Khan F., Just A., Merritt D., 2011, e-print arXiv:1103:0272v1

Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., et al. 2011, ApJS, 192, 18

Kormendy J., Richstone D., 1995, ARA&A, 33, 581

Kramer M., Stairs I. H., Manchester R. N., McLaughlin M. A., Lyne A. G., Ferdman R. D., Burgay M., Lorimer D. R., Possenti A., D'Amico N., Sarkissian J. M., Hobbs G. B., Reynolds J. E., Freire P. C. C., Camilo F., 2006, Science, 314, 97

Lazaridis K., Wex N., Jessner A., Kramer M., Stappers B. W., Janssen G. H., Desvignes G., Purver M. B., Cognard I., Theureau G., Lyne A. G., Jordan C. A., Zensus J. A., 2009, MNRAS, 400, 805

Levine J., 2004, Physics in Perspective (PIP), 6, 42

Lommen A. N., 2001, PhD thesis, UC Berkeley

Lommen A. N., 2002, in WE-Heraeus Seminar on Neutron Stars, Pulsars, and Supernova Remnants, ed. W. Becker, H. Lesch, & J. Trumper (Garching: MPE), 114

Lorimer D., Kramer M., 2005, Handbook of Pulsar Astronomy. Cambridge University Press

Maeder A., 1992, A&A, 264, 105

Maggiore M., 2000, Phys. Rep., 331, 283

Manchester R. N., 2006, Chinese Journal of Astronomy and Astrophysics Supplement, 6, 139

McHugh M. P., Zalamansky G., Vernotte F., Lantz E., 1996, Phys. Rev. D, 54, 5993

Menou K., Haiman Z., Narayanan V. K., 2001, ApJ, 558, 535

Merritt D., Poon M. Y., 2004, ApJ, 606, 788

Milosavljević M., Merritt D., 2001, ApJ, 563, 34

Misner C., Thorne K., Wheeler J., 1973, Gravitation. W.H. Freeman and Company, pp 872–915

Murphy Jr. T. W., Nordtvedt K., Turyshev S. G., 2007, Physical Review Letters, 98, 071102

Newman M., Barkema G., 1999, Monte Carlo Methods in Statistical Physics. Oxford University Press Inc., pp 31–86

Newton M. A., Raftery A. E., 1994, Journal of the Royal Statistical Society, Series B, 56, 3

Ölmez S., Mandic V., Siemens X., 2010, Phys. Rev. D, 81, 104028

Owen B. J., 1996, Phys. Rev. D, 53, 6749

Payne P. N., 1983, Phys. Rev. D, 28, 1894

Peebles P. J., Ratra B., 2003, Reviews of Modern Physics, 75, 559

Pelleg D., Moore A., 2000, in Proceedings of the 17-th International Conference on Machine Learning pp 727–734

Perets H. B., Alexander T., 2008, ApJ, 677, 146

Phinney E. S., 2001, eprint arXiv:astro-ph/0108028v1

Pound R. V., Rebka G. A., 1960, Phys. Rev. Lett., 4, 337

Prince T. A., LISA International Science Team 2011, in American Astronomical Society Meeting Abstracts #217 Vol. 43 of Bulletin of the American Astronomical Society, The Laser Interferometer Space Antenna (LISA). pp 347.07

BIBLIOGRAPHY

Pshirkov M. S., Baskaran D., Postnov K. A., 2010, MNRAS, 402, 417

Racca G. D., McNamara P. W., 2010, Space Sci. Rev., 151, 159

Riess A. G., Filippenko A. V., Challis P., Clocchiatti A., Diercks A., Garnavich P. M., Gilliland R. L., Hogan C. J., Jha S., Kirshner R. P., Stubbs C., Suntzeff N. B., Tonry J., 1998, AJ, 116, 1009

Roberts G., Gelman A., Gilks W., 1997, Ann. of Appl. Prob., 7, 110

Roos N., 1981, A&A, 104, 218

Sazhin M. V., 1978, Soviet Astronomy, 22, 36

Schödel R., Ott T., Genzel R., Hofmann R., Lehnert M., Eckart A., Mouawad N., Alexander T., Reid M. J., 2002, Nature, 419, 694

Schrabback T. et al., 2010, A&A, 516, A63+

Sesana A., Volonteri M., Haardt F., 2007, MNRAS, 377, 1711

Sesana A., Vecchio A., Colacino C. N., 2008, MNRAS, 390, 192

Seto N., 2009, MNRAS, 400, L38

Skilling J., 2004, in R. Fischer, R. Preuss, & U. V. Toussaint ed., American Institute of Physics Conference Series Vol. 735 of American Institute of Physics Conference Series, Nested Sampling. pp 395–405

Splaver E. M., Nice D. J., Stairs I. H., Lommen A. N., Backer D. D, 2005, ApJ, 620, 405-415

Stairs I. H., Thorsett S. E., Taylor J. H., Wolszczan A., 2002, ApJ, 581, 501

Stairs I. H., 2003, Living Reviews in Relativity, 6, 5

Stappers B. W., Kramer M., Lyne A. G., D'Amico N., Jessner A., 2006, Chinese Journal of Astronomy and Astrophysics Supplement, 6, 298

Stappers B. W., Kramer M., 2011, Bulletin of the American Astronomical Society, 43, #124.04

Stinebring D. R., Ryba M. F., Taylor J. H., Romani R. W., 1990, Phys. Rev. D, 65, 285-288

Takahashi R. and the Tama Collaboration, 2008, Class. Quantum Grav. 25 114036

Taylor J. H., Fowler L. A., McCulloch P. M., 1979, Nature, 277, 437

Taylor J. H., Weisberg J. M., 1982, ApJ, 253, 908

Taylor J. H., Weisberg J. M., 1989, ApJ, 345, 434

Taylor J. H., 1992, Philosophical Transactions of the Royal Society of London, 341, 117-134 (1992), 341, 117

Thorne K. S., 1992, Phys. Rev. D, 45, 520

Tofani G. et al., 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7012 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Status of the Sardinia Radio Telescope project

van Haasteren R., Levin Y., McDonald P., Lu T., 2009, MNRAS, 395, 1005

van Haasteren R., Levin Y., Janssen G.H., Lazaridis K., Kramer M., Stappers B.W., Desvignes G., Purver M.B., et al., 2011, MNRAS, 414, 3117

Verbiest J. P. W., Bailes M., Coles W. A., Hobbs G. B., van Straten W., Champion D. J., Jenet F. A., Manchester R. N., et al., 2009, MNRAS, 400, 951

Verbiest J. P. W., Bailes M., Bhat N. D. R., Burke-Spolaor S., Champion D. J., Coles W., Hobbs G. B., Hotan A., et al. 2010, Classical and Quantum Gravity, 27, 084015

Vilenkin A., 1981, Physics Letters B, 107, 47

Volonteri M., Haardt F., Madau P., 2003, in V. Avila-Reese, C. Firmani, C. S. Frenk, & C. Allen ed., Revista Mexicana de Astronomia y Astrofisica Conference Series Vol. 17 of Revista Mexicana de Astronomia y Astrofisica Conference Series, The Merging History of Binary Supermassive Black Holes. pp 333–334

Weisberg J. M., Taylor J. H., 2005, in F. A. Rasio & I. H. Stairs ed., Binary Radio Pulsars Vol. 328 of Astronomical Society of the Pacific Conference Series, The Relativistic Binary Pulsar B1913+16: Thirty Years of Observations and Analysis. pp 25–+

Wolpert R.L., 2002, Technical report, Stable Limit Laws for Marginal Probabilities from MCMC Streams: Acceleration of Convergence

Wyithe J., Loeb A., 2003, ApJ, 595, 614

Yardley D. R. B., Coles W. A., Hobbs G. B., Verbiest J. P. W., Manchester R. N., van Straten W., Jenet F. A., Bailes M., Bhat N. D. R., Burke-Spolaor S., Champion D. J., Hotan A. W., Oslowski S., Reynolds J. E., Sarkissian J. M., 2011, MNRAS, pp 491–+

Portegies Zwart S. F., Baumgardt H., Hut P., Makino J., McMillan S. L. W., 2004, Nature, 428, 724

Curriculum vitae

On January 17, 1983, I was born in The Hague, the Netherlands. I went to high school at Zandvliet College, The Hague. During my youth I have always had a fascination for all the physical sciences, but it wasn't until my 4th year in high school that I decided I was going to study physics when I read about quantum mechanics and the theory of relativity. I graduated from Zandvliet in 2001, after which I started my undergraduate in physics and astronomy at Leiden University.

During my studies, I developed a passion for teaching as well. Ever since my skills were high enough to help others, I have been tutoring, assisting classes, and teaching crash courses on various subjects. Few things are more rewarding than helping others grasp difficult concepts by teaching them with passion and patience. Besides the mandatory role of assistent for classes during my PhD, among the teaching duties I have carried out are physics and mathematics tutoring of high school students (Descarted Onderwijsbegeleiding), assisting various physics classes as university (Leiden University), providing multi-day mathematics crash courses for high school finalists (Examencursus), and being an after-hours mathematics teacher at some high schools in the Netherlands (word of mouth).

In 2007, I received my doctoraal (Master of Science degree) in theoretical physics at Lorentz Institute, Leiden University, for my research project on pulsar timing arrays with Yuri Levin. This research would turn out to be the progenitor for my PhD project, described in this thesis. I have worked on it from 2007 to 2011, at Leiden Observatory, under the supervision of Yuri Levin, with promotor Professor Simon Portegies zwart. During my research, I have collaborated with international groups of scientist. The most intense collaboration has resulted in me becoming part of the European Pulsar Timing Array effort and publishing a paper presenting the most stringent limits on the stochastic gravitational-wave background to date (see chapter 4 of this thesis). Collaboration with other pulsar timing groups has also started, and will continue in the future. During my PhD I have participated in, and presented my work at various international meetings in: Bad Honnef (Germany), Sydney (Australia), Brownsville (USA), Arecibo (USA), Manchester (UK), New York (USA), Rio de Janeiro (Brazil), Hanover (Germany), Besançon (France), Sardinia (Italy), Bonn (Germany), Toronto (Canada), and Melbourne (Australia). I will continue my work on gravitational-wave detection and pulsar timing arrays in Hanover at the Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut).
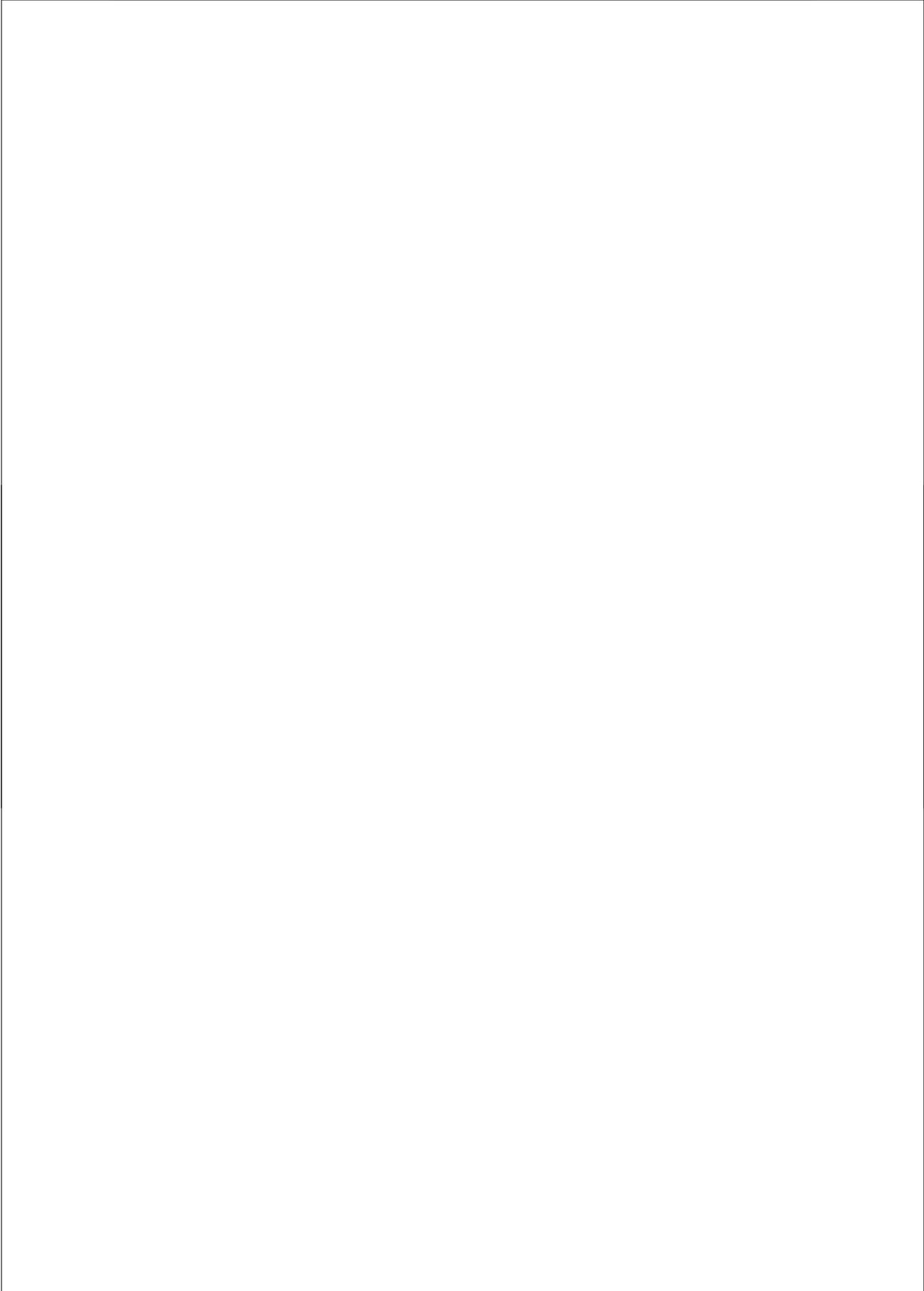
Nawoord

Deze laatste pagina van dit proefschrift wil ik graag met wat woorden van dankbaarheid vullen aan mensen die niet direct aan het onderzoek van dit proefschrift hebbben meegewerkt, maar wel invloed hebben gehad op het tot stand komen van dit proefschrift. Een van de belangrijkste lessen die ik heb geleerd tijdens mijn promotie, is dat ik geen succesvol wetenschapper wil zijn als ik daarnaast geen succesvol en rijk gevuld leven kan hebben. Zoals met alles in het leven gaat het om de gulden middenweg, en voor die les heb ik iedereen om mij heen te danken. Bovendien zorgt alles naast mijn werk behalve voor afleiding, ook voor inspiratie en goede ideeen.

The Sterrewacht, where I have spent a big portion of my daily life for years, is home to numerous colleagues that have helped me in one way or the other to produce the work you are now reading. All of you, thank you very much for allowing me to enjoy myself there! There are, however, a few people that I would like to mention by name. First my office mate Chael. Thanks for putting up with me for years, and for being the first in line to talk science to. Maarten, I hope you have come to appreciate me and my great sense of humour just as much as I have come to like our daily lunch breaks. Ann-Marie, thank you for making Melbourne as awesome as it was. We'll keep that up.

Verder wil ik al mijn cordialgenoten bedanken voor de geweldige tijd die ik altijd met ze heb. Hoewel ik nu sowieso Sainturijnen tekort doe, noem ik nog een paar mensen daarvan bij naam. Boot, je was een geweldige huisgenoot, ik zal altijd met plezier terugdenken aan mijn tijd op de Wagenstraat. P, de hoeveelheid jaren van mijn leven die ik met jou verspild heb, heb ik in het kwadraat teruggekregen aan mooie momenten. Wat had ik zonder jou aangemoeten de afgelopen jaren? Ralph, de squash was altijd een hoogtepuntje in de week, en de gesprekken waren altijd van hoog niveau. En uiteraard zijn de mooie momenten in of buiten de kroeg met de vaste bezetting Gijs, Vuijk, IJs, en Jap ook van onschatbare waarde voor mij geweest.

At Monash, the whole gang I met there, you all made it a very pleasant place to work. And thanks for introducing me to all the other things that make Melbourne a nice place to live. Amanda, thank you for your hospitality and the good times.

Remko, uiteraard kan ik jou niet vergeten hier. Ik kom snel weer 's langs voor een ouderwets avondje! Caroline, al is het maar een jaartje van mijn promotie geweest, dank je wel voor je steun en liefde gedurende die tijd. Lucy, bedankt voor de gesprekken en alle momenten van inspiratie. Sonja, lief zusje, al was je ver weg, ik heb genoten van onze skype momentjes. En als laatste natuurlijk pap en mam: woorden kunnen nooit genoeg zijn om te beschrijven wat jullie voor mij hebben gedaan en betekend. Ik houd van jullie. Allemaal.