On Pulsar Radio Emission

Ph.D. Thesis in Astronomy

at the Faculty of Physics
of the Ludwig-Maximilians-Universität München

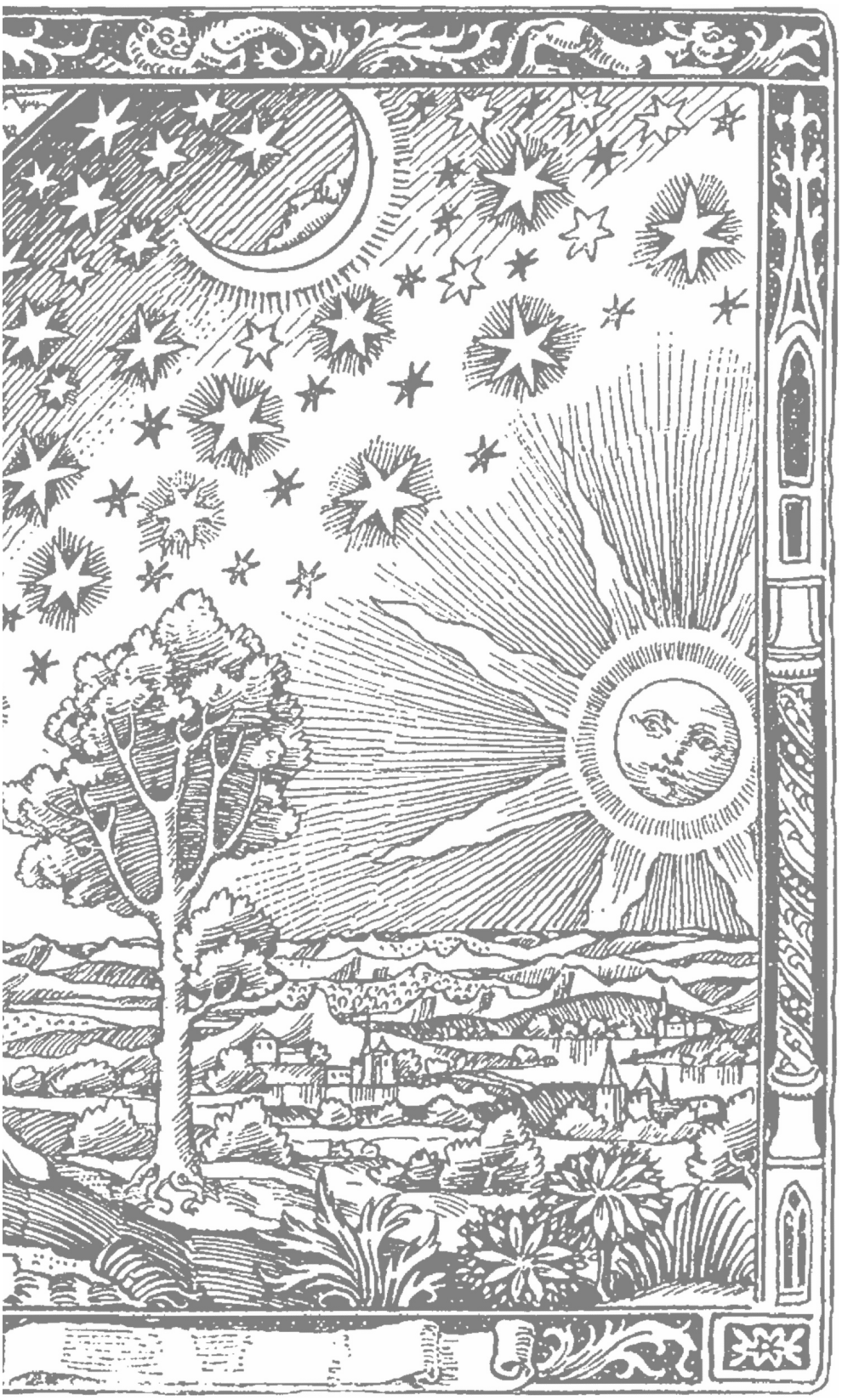
submitted by

Maximilian Imgrund
from Lohr am Main, Germany

Munich, March 2016

First evaluator: Prof. Dr. Harald Lesch
Second evaluator: Prof. Dr. Michael Kramer

Date of oral exam: 14.06.2016



Für meinen lieben Küchen-Pauli

Zusammenfassung

Diese Arbeit möchte einen Beitrag zum Verständnis der Radiostrahlung von Pulsaren leisten. Pulsare sind Neutronensterne mit einem Radius von etwa 10^6cm und einer Masse von etwa eins bis drei Sonnenmassen, die mit einer Periode zwischen Sekunden und Millisekunden rotieren. Sie besitzen immense Magnetfelder von 10^8 bis 10^{13} Gauss. Diese Felder tragen dazu bei, die Rotationsenergie im Wesentlichen in Dipolstrahlung, Röntgenstrahlung und dem Treiben eines Windes umzusetzen. Es wird weniger als ein Tausendstel des gesamten Energieverlustes in Radiostrahlung umgewandelt. Dies geschieht jedoch durch einen kollektiven Plasmaprozess in kohärenter Weise und auf einer Zeitskala von einer Nanosekunde und weniger.
Da auf dem Gebiet bereits seit fast einem halben Jahrhundert geforscht wird, beschäftigen wir uns in Kapitel 1 mit den erarbeiteten theoretischen Konzepten und Ideen für einen Strahlungsprozess und das Aussehen der sogenannten Magnetosphäre, dem mit Plasma gefüllten Bereich um einen Neutronenstern, als Ganzes. Wir zeigen auf, dass zwar viele grundlegende Fragen zur Zufriedenheit geklärt sind, aber insbesondere solche, die den Rahmen für den Radiostrahlprozess bilden, einer großen Unsicherheit unterliegen. Vor allem ist es immer noch unklar, aus welcher Quelle die Radioemission ihre Energie schöpft. Die frühen Theorien um die Arbeiten von Goldreich and Julian [1969] und Ruderman and Sutherland [1975] prognostizieren das Auftreten hoher elektrischer Felder, die einen starken elektrischen Strom zu erzeugen vermögen. Um die Energie für die Radioemission bereit zu stellen, sind hingegen eher mild relativistische Teilchenenergien und ein geringer Strom geeignet. Wie die Umwandlung von Strom in Flussenergie geschieht, ist jedoch unklar. Vielmehr stehen vielen früheren Modellen neuere Simulationen entgegen, die ebenfalls keinen relativistischen Fluss nahe des Pulsars vorhersagen.
Wir untersuchen die beobachtete Strahlung und deren Form, insbesondere im Hinblick auf die aufgezeigten theoretischen Modelle in Kapitel 2. Dort erfahren wir, dass die Radioemission wahrscheinlich auf extrem kurzen Zeitskalen, vergleichbar mit der inversen Plasmafrequenz, erzeugt wird. Wir arbeiten heraus, warum dies hohe Ansprüche an den theoretischen Emissionsprozess stellt und de facto nur noch einen Kandidaten plausibel erscheinen lässt. Wir schließen, dass wesentliche grundlegende Fragen den Energietransport und die Energiequelle des Strahlprozesses betreffend zum derzeitigen Stand der Theorie unbeantwortet bleiben. Auch die raumzeitliche Komprimierung der benötigten Energie auf wenige Zentimeter und Nanosekunden bleibt ungeklärt, insbesondere im Hinblick darauf, dass nur ein Bruchteil der theoretisch zur Verfügung stehenden Energie umgewandelt wird.
Da die für die Kompression relevanten Fluktuationen auf einer Mesoskala zwischen Nanosekunden und Mikro- bis Millisekunden stattfindet, sollte man diese Fluktuationen auch mit Beobachtungsdaten nachweisen können. Um dies zukünftig zu ermöglichen, beschließen wir in Kapitel 3 die Statistik des Empfängers der Radiostrahlung genauer zu untersuchen, da dies auch für andere Bereiche der Pulsarforschung relevant ist.
Die in Kapitel 4 beschriebenen Ergebnisse zeigen, dass die entwickelte Bayes'sche Methode in

der Tat konventionelle Verfahren zur Parameterextraktion aus gemessenen Daten in Präzission und Genauigkeit übertreffen kann. So gewichtet die Methode Rotationsphasenmessungen anders als konventionelle Verfahren und gibt den Einzelmessungen eine bessere Fehlerschätzung. Dies ist von großer Relevanz für die Gravitationswellensuche mit sogenannten Pulsar-Timing-Arrays, da die Aussagekraft der Gesamtmessung wesentlich vom Verständnis der Genauigkeit der Einzelmessung abhängt.
Die Arbeit an Einzeldaten mit Bayesschen Methoden zeigt aber auch schnell numerische Grenzen auf. So ist es wünschenswert, Analysen unter Einbeziehung aller Einzeldaten zu ermöglichen. Hierzu wurde eine Softwarebibliothek geschrieben, die gerade nicht benötigte Einzeldaten auf die Festplatte auslagert und dies auch für große Datenmengen (wesentliche Bruchteile der Festplattenkapazität, und nicht des Arbeitsspeichers) ermöglicht (siehe Kapitel 5). Die Bibliothek ist so allgemein geschrieben, dass sie auch in anderen datenintensiven Gebieten der Forschung Anwendung finden kann.
Während wir damit Grundlagen zur Auswertung von Fluktationsmodellen durch Beobachtungsdaten gelegt haben, nähern wir uns dem Problem von der theoretischen Seite aus in Kapitel 6. Wir schlagen vor, dass die energetische Kopplung der Radiostrahlung magnetischer Natur sein könnte, da diese auch in der Physik der Sonnenausbrüche relevant ist. Wir argumentieren sehr allgemein, dass aufgrund der topologischen Begebenheiten in Pulsarsystemen die Rotation des Neutronensterns magnetische Energie in das System pumpt. Diese Energie kann dann durch den Zerfall von Strömen wieder freigesetzt werden. Wir zeigen auf, dass bereits die Annihilation von Elektronen und Positronen ausreichen könnte, um Radioemission auf nicht zu vernachlässigenden Energieskalen zu erzeugen. Dieser Mechanismus benötigt keine relativistischen Energien und unterliegt somit nicht dem Problem, hohe kinetische Energien zu fordern. Wir folgern, dass die bestehenden Lücken in der Theorie des Radioemissionsprozesses in Zukunft möglicherweise geschlossen werden könnten, wenn wir Beobachtungsdaten statistisch genauer auswerten und vor Allem am grundlegenden Problem des Energietransportes forschen.
Die vorliegende Arbeit ist auch ein Beispiel dafür, dass man sich der sehr theoretischen Frage des Emissionsprozess derart nähern kann, dass die Forschungsergebnisse auch anderen Bereichen der Forschung direkt nutzen können.

Summary

This work intends to contribute to the understanding of the radio emission of pulsars. Pulsars are neutron stars with a radius of about 10^6cm and a mass of about one to three solar masses, that rotate with a period between seconds and milliseconds. They exhibit tremendous magnetic fields of 10^8 to 10^{13} Gauss. These fields facilitate the conversion of rotational energy to mainly dipole radiation, x-ray emission and the pulsar wind. Less than a thousandth of the total energy loss is being emitted as radio emission. This contribution however is generated by a collective plasma radiation process that acts coherently on a time scale of nanoseconds and below.
Since the topic has been an active field of research for nearly half a century, we introduce the resulting theoretical concepts and ideas for an emission process and the appearance of the so called "magnetosphere", the plasma filled volume around a pulsar, in Chapter 1. We show that many basic questions have been answered satisfactorily. Questions concerning the emission process, however, suffer some uncertainty. Especially the exact energy source of the radio emission remains unclear. The early works of Goldreich and Julian [1969] and Ruderman and Sutherland [1975] predict high electric fields to arise that are capable of driving a strong electric current. To supplement the energy to power the radio emission, rather mildly relativistic particle energies and a moderate current are favourable. How the system converts current into flow is unclear. In fact, the earlier theories are opposed by recent simulations that also do not predict a relativistic flow near the pulsar.
We examine the observed radiation and its form, especially in light of the illustrated models in Chapter 2. We notice that the radio emission is generated in extremely short time scales, that are comparable to the inverse of the plasma frequency. We elaborate why this places high demands on the theoretical models leaving in fact only one viable candidate process. We conclude that profound questions of energy flow and energy source remain unanswered by current theory. Furthermore, the compression of available energy in space and time to a few centimetres and nanoseconds remains unclear, especially when facing the fact that only a small fraction of the theoretically available energy is being converted.
Since the fluctuations relevant for the compression of the energy take place on an intermediate scale of nanoseconds to micro- and milliseconds, it should be possible to detect these observationally. To facilitate this, we decide to analyse the statistics of the receiver equation of radio radiation in Chapter 3, also since this is relevant to other topics of pulsar research.
The results presented in Chapter 4 show that the developed Bayesian method excels conventional methods to extract parameters from observation data in both precision and accuracy. The method for example weights rotation phase measurements differently than conventional techniques and assigns a more accurate error estimation to single measurements. This is of great relevance to gravitational wave search with so called "pulsar timing array", as the validity of the total measurement is substantially dependent on the understanding of the accuracy assigned to the single observations.
However, the work on single observation data with Bayesian techniques also exemplifies the

numerical limits of this method. It is desirable to enable algorithms to include single observation data in the analysis. Therefore we developed a runtime library that writes out currently unneeded data to hard disk, being capable to manage huge data sets (substantial fractions of the hard disk space, not the main memory) in Chapter 5. This library has been written in a generic form so that it can be also used in other data-intensive areas of research.
While we thereby lay the foundations to evaluate fluctuation models by observational data, we approach the problem from theoretical grounds in Chapter 6. We propose that the energetic coupling of radio emission could be of magnetic origin, as this is also a relevant mechanism in solar flare physics. We argue in a general way that the rotation of the pulsar pumps energy into the magnetic field, due to topological reasons. This energy can be released again by current decay. We show that already the annihilation of electrons and positrons may suffice to generate radio emission on non-negligible energy scales. This mechanism is not dependent on relativistic flow and thus does not suffer from the problem of requiring high kinetic particle energies.
We conclude that the existing gaps in the theory of the radio emission process could possibly be closed in the future, if we analyse observational data statistically more accurate and especially if we put more effort into understanding the problem of energy transport.
This thesis serves as an example that scientific investigation of a very theoretical question such as the origin of radio emission can lead to results that may be used directly in other areas of research.

Contents

CHAPTER 1

The problem of pulsar radio emission

1.1 A brief introduction to pulsars

Pulsars are fast rotating neutron stars possessing large magnetic fields of 10^8 to 10^{13} Gauss and rotation periods of milliseconds to seconds [Manchester et al., 2005]. The first pulsars were discovered by Hewish et al. [1968] detecting the pulsed radio emission. Besides our main subject, the radio emission, they are also sources of γ-rays. The energy driving these emissions is drawn from the rotational energy of these compact (radii of 10^6cm) objects of about one to two solar masses. While the millisecond population is expected to be spun up by mass accretion and thus called *recycled* [Bhattacharya and van den Heuvel, 1991, Tauris and van den Heuvel, 2006], periods on the order of a second can be easily explained by a toy model of angular momentum and energy conservation. Magnetic flux conservation leads to the high values of the magnetic field, which dominates structure of the surrounding plasma, the so called *magnetosphere*. The currents and magnetic fields exert a torque on the neutron star slowing it down in time and carrying away rotational energy. The radio emission only accounts for a tiny fraction ($< 0.1\%$) of the total energy dissipated, while most energy is expected to be lost by dipole radiation and the pulsar wind. However there are rare examples where the observed output matches the theoretical output measured by the change in rotation period. One of these examples is the Crab pulsar [Lyne and Graham-Smith, 2012].

1.2 Qualitative estimations on magnetic field and energy

Being the remnant of a supernova explosion, a neutron star still carries a fraction of the angular momentum of the star. under the assumption that the star lost parts of its mass but only a negligible amount of angular momentum per mass, we immediately deduce that the angular momentum for a small lump of mass remaining at the star scales as:

$$m\omega_{pre}r_{pre}^2 = m\omega_{post}r_{post}^2 \tag{1.1}$$

$$\Rightarrow \omega_{post} = \omega_{pre}\frac{r_{pre}^2}{r_{post}^2} \tag{1.2}$$

which can easily lead from a rotation in a months time to a rotation period of a few milliseconds. The immense rotational energy of a star of about 1.4 solar masses then amounts to:

$$E_{\mathrm{rot}} = I_{NS}\frac{\omega^2}{2} = I_{NS}\frac{2\pi^2}{P^2} \tag{1.3}$$

where I_{NS} is the moment of inertia of the neutron star and P denotes its period. There exists a lower limit to the pulsars rotation period as matter on its surface should still be gravitationally bound. We conclude that in this case the centrifugal force should be still dominated by the force exerted by gravitation:

$$\omega^2 r < \frac{GM}{r_{NS}^2} \Rightarrow P > 2\pi\sqrt{\frac{r_{NS}^3}{GM}} \tag{1.4}$$

Where M denotes the neutron star mass, $r_{NS} \approx 10^6$cm the neutron star radius and G the gravitational constant. For a neutron star of $1.5M_\odot$ we derive a critical period P_{crit} of about 0.5 milliseconds. Indeed the pulsar population is limited to the millisecond to second period regime with pulsars aggregating at the second and millisecond regime respectively. When accreting, the imposed lower limit to the rotation period restricts the transport of rotational momentum. Non-accreting pulsars slow down to end up in the second regime by the effects of dipole radiation and pulsar wind energy losses, which we will discuss briefly in the magnetosphere section 1.3 below.
Assuming the star to be a sphere of equally dense matter we can estimate the rotational energy available to be dissipated by inserting the moment of inertia for a sphere, $I = \frac{2}{5}mr^2$:

$$E_{\mathrm{rot}} = \frac{4\pi^2}{5} M_\odot \frac{r_{NS}^2}{P^2} = 1.6 \cdot 10^{52}\mathrm{erg} \tag{1.5}$$

We can calculate the power dissipated by taking the derivative of the rotational energy w.r.t. time:

$$P_{\mathrm{rot}} = -4\pi^2 I_{NS} \frac{\dot{P}}{P^3} = I_{NS}\omega\dot{\omega} \tag{1.6}$$

This quantity is called *spin-down luminosity* [Lorimer and Kramer, 2004] and denotes the total power output of the pulsar that is usually only observable by the slow-down of the period.
The collapse not only spins up the supernova remnant, but also amplifies the magnetic field of the star. This can be easily seen by assuming magnetic flux conservation over a hemisphere. as $\Phi \propto Br^2 = \mathrm{const}$, we deduce that

$$B_{post} \approx B_{pre} \frac{r_{pre}^2}{r_{post}^2} \tag{1.7}$$

enlarging the magnetic field of the progenitor star to values of 10^8 to 10^{13} Gauss. As is known from electrodynamics, a rotating dipole will emit dipole radiation of a strength characterized by the absolute magnetic moment m, the angle α between the magnetic moment and the spin axis α and its cyclic frequency $\omega = 2\pi\frac{1}{P}$. Following the argumentation of Lorimer and Kramer [2004] we may estimate the lost power by dipole emission as

$$P_{dip} = -\frac{2}{3c^3} m^2 \omega^4 \sin^2\alpha \tag{1.8}$$

Equating this with the spin-down luminosity and solving for the cyclic frequency derivative yields

$$\dot{\omega} = \frac{2}{3I_{NS}c^3} m^2 \sin^2\alpha\omega^3 \tag{1.9}$$

Reformulating this equation for the frequency derivative $\dot{\nu} = \dot{\omega}/(2\pi)$ we define the *braking index* n_{br}

$$\dot{\nu} = K \cdot \nu^{n_b} \tag{1.10}$$

with K being defined implicitly by equation 1.9. In principle this index may be calculated by observing the frequency and its derivative as taking the derivative of (1.10) w.r.t. time and solving for n_{br}.

$$\ddot{\nu} = K \cdot n_{br} \nu^{n_{br}} \frac{\dot{\nu}}{\nu} \overset{(1.10)}{\Rightarrow} n_{br} = \frac{\nu \ddot{\nu}}{\dot{\nu}^2} \tag{1.11}$$

In practise however, timing errors and changes of the moment of inertia due to e.g. crust quakes complicate a simple measurement. Typical values of $\dot{\nu}$ amount to $10^{-15}\mathrm{Hz/s}$. As n_{br} is expected to be on the order of unity, we can deduce that $\ddot{\nu} \approx \dot{\nu}^2/\nu$ leading to a change of $10^{-30}\mathrm{Hz/s}^2$ Following the treatment of pulsar braking indices by Johnston and Galloway [1999] this does not even amount for a whole period in 600 years of observation. Young pulsars exhibit larger values of $\dot{\nu}$ rendering the measurement possible. They however often show *glitching*, jumps in the rotational phase, caused by crust quakes of matter in the neutron star's outer layer [Shemar and Lyne, 1996]. Being unpredictable, they introduce errors to the pulsar phase and thus aggravate the problem of measuring the second derivative of the pulsar frequency. Johnston and Galloway [1999] have overcome this limit by integrating equation (1.10), however, derive braking indices that are mostly higher than expected from classical estimations.
It is clear from equation (1.9) that we expect $n_{br} = 3$. However measured values deviate from this value by orders of magnitude giving no clear picture. As we've evaluated the dipole emission only occurring for $\alpha > 0$, other terms such as the pulsar wind term may alter the theoretical prediction of n_{br}. For an example see Contopoulos and Spitkovsky [2006].
To conclude this section, let us calculate the relation between the expected surface magnetic field and the theoretical energy loss. Inserting the magnetic dipole moment $m = B_{surf} r^3$ of a sphere of constant density [Jackson, 1998] into (1.9) and solving for B_{surf} we deduce that

$$B_{surf} = \sqrt{\frac{3\dot{\omega} I_{NS} c^3}{2 \sin^2 \alpha \omega^3 r_{NS}^6}} = \sqrt{\frac{3\dot{\omega} m c^3}{5 \sin^2 \alpha \omega^3 r_{NS}^4}} \tag{1.12}$$

Inserting a period of a second, 1.5 solar masses, $\alpha = 90°$ and a relative frequency derivative of 10^{-15} we arrive at a magnetic field of about 10^{13} Gauss. At this rate of deceleration, we expect the object to drain a power of

$$P = 1.2 \cdot 10^{30} \frac{\mathrm{erg}}{\mathrm{s}} \tag{1.13}$$

when inserting these numbers in equation (1.6).
We conclude that the period and basic parameters suffice to deduce the approximate energy loss of the system. It is important to notice that, as we will see, only a small fraction of this huge energy is being converted into radio radiation.
When the dipole and rotational axis are nearly aligned, dipole radiation is not efficient any more, as can be seen by equation (1.8) for α approaching zero. In this case, the magnetospheric energy loss is dominant and magnetospheric models have to explain how the magnetosphere dissipates energy and exerts a torque onto the neutron star, slowing it down.
To explain the observed braking, the magnetosphere must be able to drain and transport energy from the pulsar. After describing the magnetosphere models under discussion and possible radio emission processes in the following sections, we will come back to this basic problem of energy conversion and transport.

1.3 Pulsar magnetosphere models

1.3.1 There is no neutron star in vacuum

Pulsar magnetospheric physics are driven by two basic assumptions. Firstly, there exists a dipole magnetic field that is possibly misaligned with the rotation vector of the pulsar (oblique rotator case) and secondly that this magnetic field is filled up with plasma. That there cannot be a pure vacuum solution is one of the main results of the well-known and often criticized paper of Goldreich and Julian [1969]. Their work however remains central to the field for the basic arguments presented. Their work has been preceded phenomenologically [Deutsch, 1955] and quantitatively [Hones and Bergeson, 1965, Pacini, 1967] but remains outstanding as the first application of more general principles to the case of a pulsar.
Let us reconsider some of their arguments for the case of an aligned rotator in our symbolic language. We assume that the stellar matter of the neutron star is an excellent conductor. As charge carriers are moved through an magnetic field by rotation, they are subject to the Lorentz force. In a force free equilibrium case, we thus conclude that

$$\vec{E} + (\vec{\omega} \times \vec{r}) \times \vec{B} = 0 \tag{1.14}$$

where $\vec{E}$ is the electric field, $\vec{B}$ the magnetic field, $\vec{\omega}$ the axis of rotation and $\vec{r}$ the position vector of the surface element. We will discuss the validity of this equation further in section 1.3.3. Consequently we may calculate the electric field induced by inserting the magnetic dipole field. If the neutron star would reside in a vacuum, Laplace's equation for the external electrostatic potential has to hold, which results in a potential that reads

$$\Phi = \frac{-B_{surf}\omega r_{ns}^5}{3cr^3} P_2(\cos\theta) \tag{1.15}$$

where r, θ, ϕ denote polar coordinates that are centred at the neutron star's centre and $\theta = 0$ is along the rotational axis and P_2 denotes the Legendre polynomial of second degree. In this case a surface charge would form in order to cancel the electric field component normal to the surface. This can happen only if the matter at the star's surface is bound gravitationally to the star. However, the arising field component $E_{||}$ parallel to the magnetic field lines takes values of

$$E_{||} = \frac{\vec{E} \cdot \vec{B}}{|B|} = -\frac{\omega B_{surf} r_{NS}}{c} \cos^3\theta \qquad \text{[Lorimer and Kramer, 2004]} \tag{1.16}$$

The minimum surface electric field that surmounts gravity at θ can be calculated equating the gravitational and electric force and solving for the magnetic field:

$$\frac{GM_{NS}m}{r_{NS}^2} = e\frac{\omega B_{surf} r_{NS}}{c} \tag{1.17}$$

$$B_{min} = \frac{G \cdot M_{NS} mc}{r_{NS}^3 \omega} \tag{1.18}$$

The minimum magnetic field to generate a surface electric field surmounting gravity takes values of a few ten of Gauss for an electron and about 10^4 Gauss for a proton, inserting the exemplary values of $M_{NS} = 1.5M_\odot$ and $P = 1$s. The values of magnetic fields calculated in section 1.2 supersede these values by factors of 10^6 to 10^{12} rendering binding of this surface charge to the neutron star's surface unlikely.
If we now assume the co-rotating magnetosphere to be filled with plasma, the charge density

that arises in the co-rotating part of the magnetosphere has been calculated by Goldreich and Julian as

$$-e \cdot n_{GJ} := \frac{-B}{Pc} \frac{1}{1 - (\frac{2\pi r}{Pc})^2 \sin^2 \theta} \tag{1.19}$$

This density is a net charge density and as such the local number density consisting possibly of charge carriers of both signs may be a multiple of this exemplary density. The minimum number density that is associated with this charge density above the magnetic pole consequently amounts to

$$n_{GJ} = \frac{B}{ecP} = 6.9 \cdot 10^{-2} \cdot \frac{B_0}{P_0} \left(\frac{r_{NS}}{r}\right)^3 \frac{1}{\mathrm{cm}^3} \tag{1.20}$$

where r is the distance to the neutron star's centre and this formula omits the angular dependence in equation (1.19). There exists a more general solution discussed by Hones and Bergeson [1965] extending to the oblique rotator, the more common case in the pulsar population for which magnetic and rotation axis are misaligned.
At that densities we may still treat the plasma as effectively collision-less. Having negligible resistance, the plasma can be treated with ideal magnetohydrodynamics(MHD). The very high magnetic field strengths let the plasma only flow along the field lines, as movement in other directions would instantaneously be stopped by synchrotron radiation. The synchrotron cooling time may be estimated as $8 \cdot 10^8/((B/\mathrm{Gauss})^2\gamma)$ [Ghisellini, 2013] leading to values of picoseconds and well below in our case.
As we will see in the following sections, there exist areas exhibiting finite conductivity, rendering the MHD approximation invalid.

1.3.2 The light cylinder

Up to now we have only considered a rigidly co-rotating magnetosphere extending into space. However, due to inertia the plasma is not able to follow this rotation for large radii from the axis of rotation as this would imply the plasma to have speeds superseding the speed of light. As a lump of plasma will pass a distance of $2\pi r_L$ in P, the limit imposed by staying subluminal amounts to

$$r_L = \frac{cP}{2\pi} \tag{1.21}$$

which limits rigid rotation to about 5 to 5000 pulsar radii depending on the period. At that radius the inertia of the plasma exerts a drag on the magnetic field that lets the field lines bend against rotation. The lines passing this cylinder are expected to be closed at infinity. This divides the magnetic field lines of the pulsar in two classes, open field lines that extend beyond the cylinder and dipole like closed field lines. As $\sin^2\theta/r$ is constant along a field line for a dipole magnetic field, following Lorimer and Kramer [2004] we estimate the angle of open field lines tracking the last closed field line that touches the light cylinder. For this field line we can deduce its opening angle at one pulsar radius r_{NS}:

$$\frac{\sin^2\theta_L}{r_L} = \frac{1}{R_L} = \frac{\sin^2\theta_P}{r_{NS}} \Rightarrow r_P \approx r_{NS} \sin\theta_P = \sqrt{\frac{2\pi r_{NS}^3}{cP}} \tag{1.22}$$

We depicted qualitatively the magnetic field in Figure 1.1. As the field lines adapt a toroidal component by being swept back, Ampere's law tells us that either an electric field in form of a displacement current arises or an electric current is driven through the magnetic field lines:

$$\vec{\nabla} \times \vec{B} = \frac{1}{c}(\partial_t \vec{E} + 4\pi \vec{j}) \tag{1.23}$$

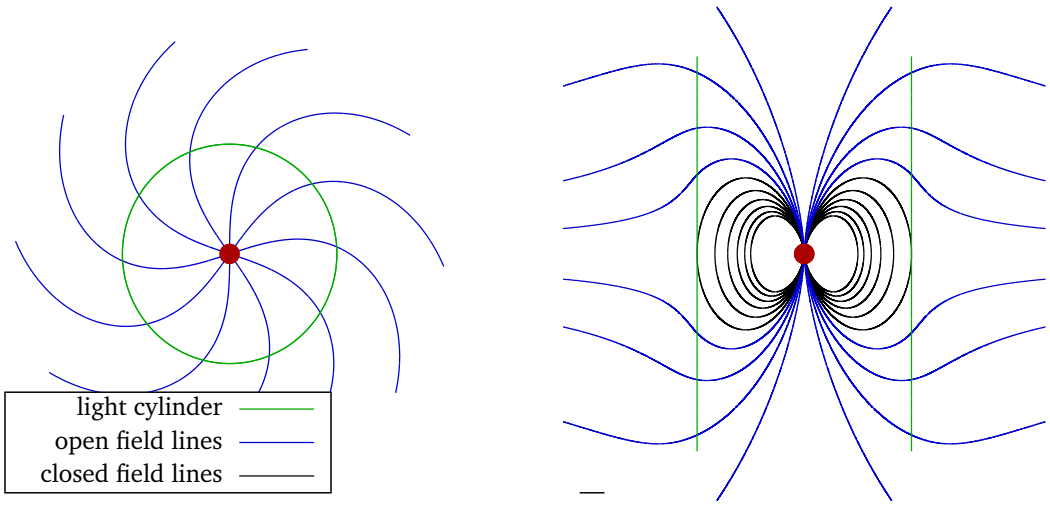


Figure 1.1: **Schematic features of the impact of the light cylinder:** We depict the idea how the light cylinder affects the geometry of the pulsar magnetosphere in the aligned rotator case. As the inertia of the plasma prohibits co-rotation with the pulsar, the field lines are dragged behind the counter-clockwise rotation of the pulsar. This implies that the field lines crossing the light cylinder are not closed but open to the surroundings. The toroidal component at the light cylinder implies a current to be drawn from the surface or a high electric field will be built up by the displacement current.

As is illustrated in Figure 1.1, integrating the rotation of the magnetic field over a slice of the light cylinder perpendicular to the rotation axis can by Stoke's theorem be reformed to a line integral over the light cylinder line. The line integral and thus the left side of Ampere's law(1.23) does not vanish, giving rise to a toroidal magnetic field component. The r.h.s. of Ampere's law states that there will be a rising electric field or a current is driven through the area.
While the exact impact of this drag or twisting and the associated currents on the magnetosphere remains an issue to be solved, the fact that there is such an inevitable current is central to the understanding of the pulsar magnetosphere problem.
As the neutron star will not change spin direction, the displacement current's derivative is non-oscillating and thus has to be zero on the long term evolution. Short term deviations, however, could produce a static electric field that could possibly remain. As this electric field has a large component parallel to the magnetic field lines, it is in conflict with the assumption of infinite conductivity in ideal magnetohydrodynamics. As the inevitable directed current flows along these possibly existing fields, we arrive treating the pulsar like an electric circuit, an idea that was followed by e.g. Shibata [1991] and that is being revived supported by numerical simulations.
Before turning to these more recent ideas, let us close with some critical remarks that the light-cylinder argument introduces to the Goldreich-Julian model. As a current is drawn from the pulsar, the pulsar will effectively be charged if there is no back-flow onto the surface, an effect that indeed leaves us with an electric field but no current flowing any more. A large electric field accelerates charge carriers to high energies. There exist several mechanisms such as e.g. electron-photon interaction that can lead to creation of electron positron pairs, an effect we will detail in the next section. These pairs act as a source of plasma that could sustain the

current, even if all charge carriers are drained from the neutron star. If the electric field is not large enough, electron-positron generation stops and the pulsar indeed acquires a rigidly rotating magnetosphere, that, however, is contained mostly within the light cylinder. This has been shown in simulations [Chen and Beloborodov, 2014] and confirmed the expectations of numerous authors [see e.g. Michel, 2004, and references therein]. As Sturrock [1971] already suggested, this could explain the so called pulsar death line, after which a pulsar cannot produce enough pairs to sustain an electron-positron plasma. Then its current ceases and the radio emission vanishes.

1.3.3 Spark Gap Models and Ideal MHD

Consequently there must necessarily exist non force-free regions of $E_{||} > 0$, so called gaps, in the pulsar magnetosphere where particles are accelerated to high energies [Chen and Beloborodov, 2014]. The main idea of the mechanism is that once the particles reach an energy of $2m_e c^2$, curvature radiation photons may trigger pair production. Secondary pairs arise by acceleration of the created pairs, triggering a cascade of creation events. Following Chen and Beloborodov [2014] there have been proposed three gap regions: The polar-cap or inner gap [Ruderman and Sutherland, 1975, Sturrock, 1971], the slot gap above the polar cap [Arons, 1983, Muslimov and Harding, 2004] and the outer gap [Cheng et al., 1986] to produce pairs.
The exact pair creation location is still an open issue [Grenier and Harding, 2015] and the multiplicity of the following pair production cascade[1] seems to be lower than expected [Timokhin and Harding, 2015]. There are by now numerous authors that disfavour the inner gap as the source of pair creation, such as Chen and Beloborodov [2014], Jessner et al. [2001].
Let us motivate the interplay between gaps and current. In vacuo there can exist an electric field. The presence of plasma, however, comes with consequences. In magnetohydrodynamics the current flowing is approximated as being dependent on the electric field and the electric conductivity σ:

$$\vec{j} \propto \sigma \vec{E} \tag{1.24}$$

Multiplying this equation with $\vec{E}$ tells us that a current flow having finite conductivity and thus resistance, will dissipate an energy density of $\vec{j} \cdot \vec{E}$. Integrating this power loss over volume leads to the well-known formula for electric work in a circuit $P = U \cdot I$. This intrinsically relates the electric field parallel to the current with dissipating or gaining energy. In ideal magnetohydrodynamics however we assume the conductivity to be infinite. As the current is finite, we conclude from (1.24) that the electric field parallel to the current must vanish, $\vec{E}_{||} \to 0$, leading to a force free plasma. There still can be perpendicular fields that balances out the magnetic force on the flowing plasma. This is the deeper reason why equation (1.14) holds, as the Lorentz force on a particle is balanced out and thus effectively equals zero.
If the parallel component of the electric field is not zero, energy is being converted from electric to kinetic energy or vice versa. In both cases, we may deduce by proportionality of the current to the field that

$$\mathcal{P} = jE_{||} \propto \sigma E_{||}^2 \tag{1.25}$$

which means that at high conductivity, already a small electric field will inject a lot of power into the plasma, quickly driving it to high values of γ. This power has to be supplied electrically or magnetically and converted via a displacement current from the outside to maintain the electric field. As the force free magnetosphere approximation breaks down at the light cylinder, it is not clear why the associated power should free itself near the polar cap, as the plasma is in rigid motion with the magnetic field lines there and locally there can be reached a force

[1] Number of particles gained by one elementary input electron

free equilibrium. A naive guess in the sense of the presented arguments would render the slot and outer gap as more likely candidates, as typical magnetohydrodynamical resistivity is much more easily to be produced there. Both inertial effects of the plasma and mixing with an outer field become important. While reconnection is deemed impossible by the high fields above the polar cap, it is expected to play a major role in generating the pulsar wind as magnetic fields drop to moderate values far above the polar cap.
If, however the rotational energy has been converted to kinetic energy somewhere, there must exist regions where $E_{||} \neq 0$ to make the conversion possible. Let us assume for a moment that all energy that dissipates (equation (1.13)) would have been converted to kinetic energy somewhere on the way and particle number is conserved. We could argue that if an electron positron plasma should carry this energy outwards, its kinetic energy flow has to match the total output:

$$\mathcal{P} \approx mc^2\gamma \int_{PC} nc \, \mathrm{d}A = \frac{2\pi^2 mc}{e} r_{NS}^3 \frac{B_{surf}\gamma}{P^2} \tag{1.26}$$

where we inserted the area of the polar cap radius according to equation (1.22) and the Goldreich-Julian density as of equation (1.20). Solving for γ we find

$$\gamma = \frac{e}{2\pi^2 mc r_{NS}^3} \frac{P^2\mathcal{P}}{B_{surf}} \tag{1.27}$$

Which evaluates to γ factors of 10^5 for our exemplary case of a pulsar with $P = 1$s and $B_{surf} = 10^{13}$ Gauss. Indeed McCann [2011] find photons with energies pointing to such high γ factors.[2] However the authors exclude emission from a radius nearer than 2% of the light cylinder radius and leave open whether the hard photons have been up-scattered to such high energies by the inverse Compton effect. We will discuss such high gamma factors w.r.t. radio emission in the final section of this chapter.
While such high energies are being reached, there exists no straightforward argument why a pure magnetic coupling should power a stream of plasma (in contrast to a current), except that magnetic energy is completely dissipated, since higher γ values do not lead to stronger currents. As charge flow is invariant of energy and only dependent on the number density and velocity of a current of charges, and finally it is current and not flow that couples to magnetic fields.
A current of low γ may transport energy from a source to a sink magnetically. As e.g. Spicer [1982] argue, energy transport in ideal MHD conditions is free of dissipation (if that were not the case, electric fields would arise due to resistance). Never the less, energy can be transported along a current by the associated magnetic field. This is simply an expression of Ampere's law. Imagine a current flowing through a flux tube. Clearly it carries a magnetic contribution of $\vec{\nabla} \times \vec{B}$ that is coupled to the whole magnetic energy density available $\epsilon_B = \frac{|\vec{B}^2|}{8\pi}$. If this current decays, a displacement current arises due to magnetic induction and the energy in the associated magnetic field is released again.
The rotation of a pulsar w.r.t. the surrounding space and the light cylinder argument demand this rotation to cease with distance from the star. Also in the case of an oblique rotator, the magnetic field is at least "wound up" around itself as the ideal MHD current freezes in the magnetic field lines. This process inevitably pumps magnetic energy into the system that has to be released e.g. by pulsar wind in zones where conductivity takes low enough values to play a role. We will discuss some of the coupling of the current to the magnetic energy in chapter 6.

[2] While their work is concerned with the Crab pulsar, a very young pulsar with a 33ms period, we just want to point out that such high photon energies exist and should be explained by a proper magnetospheric model

1.3.4 Current models

Let us sum up the arguments and counter-working mechanisms presented graphically in Figure 1.2. Solving the pulsar magnetosphere problem amounts to resolve the contradictions depicted. While this work will treat aspects of the problem of coherent radio emission, the background and framework in which radio emission takes place is the pulsar magnetosphere. Even though there is a strong separation in both time scale and length scale , the assumptions on the magnetosphere set the stage for, and will be the background and boundary conditions of, the problem of coherent radio emission.
Radio emission as well as energy dissipation (measured as slow-down of rotation frequency and in terms of γ-radiation) constitute the well-established observables of the problem. We have investigated the **two main theoretical concepts**, the vacuum / rotating dipole model and the light cylinder problem.
The rotating dipole in vacuo would generate high fields that could prevail if no net charges arise or no currents are flowing. In the case of current flow, deviations from ideal MHD, that have to be explained by microphysics have to exist to still yield high field values. To maintain an electron positron plasma, pairs lost due to annihilation and possibly outflow have to be constantly replaced. This is most easily done using the free energy provided by acceleration by an electric field. Plasma can also be drawn from the surface of the neutron star by high field values. Thus while globally a near to ideal MHD treatment is appropriate in presence of a plasma, deviations have to arise to ensure long-term existence of the magnetosphere.
The light cylinder argument supported by the well established theory of relativity dominates the large scale shape of the magnetosphere. It motivates that by Ampere's law there will be a current flowing and points to another source of free energy; the differential rotation of the pulsar w.r.t. the surrounding medium. How this essentially magnetic energy will be converted and transported is not yet totally explained, but, as we have argued, recent observations and models suppose the energy to be converted forming the pulsar wind and accelerating particles at the outer gap where the strength of the magnetic field may not be dominant over plasma dynamics any more. Plasma inertia leads to the opening of some of the magnetic field lines to the outside, establishing the idea of a constant current flowing to or from the outside.
Thus, while the concept of the rotating dipole implies large fields, the light cylinder implies large currents to be essential. A correct model will settle somewhere in the middle of both extremes, as both currents and fields are required by the fundamental processes expected to take place.
The problem of an open current charging the neutron star has been early on proposed to be solved by a back-flow along an equatorial current sheet, taking a path along the separatrix of the magnetic field lines at the intersection of the closed field lines and the light cylinder and back over the first open field lines. Along a thereby closed circuit, we expect power to be dissipated and induced magnetically for the following reason: As a static potential does not lead to a perpetual flow but charge separation, a static potential may be present, but is of minor importance to the current. If however energy is dissipated over kinetic channels, then $\int \mathrm{d}\vec{l} \cdot \vec{E} \neq 0$ along the current and thus the line integral along the current does not vanish. This however means, that $\mathrm{curl}\ \vec{E}$ does not vanish by virtue of Stoke's theorem. Maxwell's equations however tell us that there needs to be a term $\partial_t \vec{B}$ associated with the rotation of the electric field. This points to a magnetic dynamo coupling. As the rotation of the pulsar always acts in the same direction, we would by this naive arguments expect the magnetic field strength to rise indefinitely.
However, we assumed the line of the circuit to be constant in shape and closed, an assumption ultimately driven by the ideal MHD scenario that currents only flow along field lines. As this

Figure 1.2: **Overview of the magnetospheric problem:** The right magnetospheric model has to solve the contradicting consequences of the vacuum dipole model and the implications of the existence of the light cylinder, both depicted in light blue. The derived counterparts in green interact in a non-trivial way to shape the actual magnetic and electric field configuration of the magnetosphere. They are connected to the main observational sources of information depicted in red. See also discussion in section 1.3.4

cannot be the case for the previously presented arguments of the light cylinder and energy dissipation, we conclude that the shape of the circuit geometry has to be time variant. We conclude that there exists magnetic reconnection and complex magnetic dynamo like processes that complicate the analytical treatment of the pulsar magnetosphere.
Fortunately there has been large progress in numerical simulation of the pulsar magnetosphere [Chen and Beloborodov, 2014, Contopoulos et al., 1999, Tchekhovskoy et al., 2013, Timokhin, 2006]. In virtually all recent treatments the authors argue that there exists a back-current on the star, minor electric fields in the interior, but strong fields in the wind zone and at the light cylinder radius. Both currents and densities are found to agree with the Goldreich Julian value, however there is some evidence of possibly larger number densities than Goldreich-Julian. If this is the case we can immediately conclude the presence of a low γ flow, as for relativistic energies the current is dominated by number density as $\beta \to 1$. There is also analytical evidence for these low gamma factors to be more realistic [Kunzl et al., 1998, Petrova, 2015]. Numerically this is also a result of [Tchekhovskoy et al., 2013, figure 2].
As we have already noticed, when transporting the energy from spin-down kinetically, particles should be accelerated to high relativistic speeds. As we are only in a minor relativistic regime, this might not be the case. We consequently expect that magnetic energy coupling to transport the rotational energy outwards. Therefore the magnetic energy density inside the light cylinder should exceed the kinetic energy density, a result that can also be seen in Chen and Beloborodov [2014, Figure 3b] who also show that in case of decelerating rotation and falling below the limit for pair creation the magnetosphere starves, the current ceases and high static fields arise. The plasma in this case however is confined within the light cylinder. This last result resolves the long known contradictions of the proposed existence of only statically charged pulsar magnetospheres based on the work of Goldreich Julian, a critique raised by e.g. Michel [2004]. The solution is that of a pulsar with only static charge and no currents and simply is the solution of a seemingly “dead” pulsar.

1.4 Coherent radio emission

From their discovery onwards, the most striking feature about the radio emission was its brightness. Let us consider Figure 3 of Allen and Horvath [2000] and assume that the luminosities reported have been emitted thermally by black body emission. In this case we may assign the so called brightness temperature, the temperature at which a perfect absorber would emit such a high intensity in the radio range:

$$T_{bb} = \frac{L_\nu c^2}{2A\delta\nu\nu^2 k_B} \tag{1.28}$$

where L_ν is the received radiation over the whole bandwidth, A is the area of the black body, ν the frequency at which the radiation has been received, $\delta\nu$ the receiver bandwidth and k_B the Boltzmann constant. This formula has been taken from Imgrund [2012] and is valid for a bandwidth that is small w.r.t. the whole spectrum. Inserting the values from Allen and Horvath [2000], we see that $L_\nu/\delta\nu \approx 2 \cdot 10^{19} \frac{\mathrm{erg}}{\mathrm{sHz}}$. We may estimate the radiating area A by the found variability time scale, assuming a coupling with the speed of light. As pulsar radio radiation shows variations in single pulses of milliseconds and below, we may estimate an area that is below $4\pi(c \cdot 1\mathrm{ms})^2 = 1.1 \cdot 10^{16}\mathrm{cm}^2$. Assuming an observation frequency of 100 MHz and inserting Boltzmann's constant, we find $T_{bb} = 3 \cdot 10^{23} K$, a value far too high to correspond to a real temperature. pointing to a non-thermal and coherent origin of the emission. Total radio luminosities are assumed to fall in the range of about 10^{21} to 10^{28} erg per second, amounting

to a small fraction of the total luminosity calculated by slow-down of the period.
While this never the less enormous power output can be supplied by rotational energy, the time scale at which this energy is being released was believed to be microseconds. However, recent observations both of spectral form [Krzeszowski et al., 2014] and intensity [Hankins et al., 2003] set new limits and give hints that the radio emission may consist of nanosecond shots of intense radiation.
The problem of coherent emission is two fold. On the one hand there needs to be explained how a tiny fraction of a large energy flow is being radiated without possibly more energy loss and on the other hand, that this process is happening very fast and seems to be depleting the energy available in the region of coherent emission.
Let us estimate the energy of a single shot of radiation from [Hankins et al., 2003] following their argumentation. If a $L_\nu = 1000\mathrm{Jy}$ pulse has been emitted in t $= 2$ ns, we would address a volume of $(ct)^3 \approx (60\mathrm{cm})^3$ to the radiation process. This process is emitting an energy density of $\mathcal{E}_{tot} = L_\nu \pi r_{Crab}^2 \delta\nu t/(ct)^3 \approx 5 \cdot 10^{18} \frac{\mathrm{erg}}{\mathrm{cm}^3}$ given a receiver bandwidth of $\delta\nu = 0.5$Ghz and the distance to the examined Crab pulsar, $r_{crab} \approx 2$kpc. Assuming the pulses have been emitted directly above the polar cap, we estimate a gamma factor per particle of $\gamma = \mathcal{E}_{tot}/(n_{GJ} m_e c^2) \approx 3 \cdot 10^{11}$. This energy supersedes kinetic supply and wavelike interactions are hard to imagine at such relativistic masses. Such high energies lead to a plasma frequency below a Hz for Goldreich-Julian densities, as plasma frequency is proportional to $\gamma^{-3/2}$. As virtually all plasma instabilities couple to the plasma frequency as a scaling factor, this makes emission from a counter-streaming plasma which's centre of mass frame is at rest w.r.t. us unlikely. The emission cannot easily reach GHz. Conversely, the observation of radio emission at frequencies of about 100MHz can place a limit on charge density and γ factors [Kunzl et al., 1998]. Hence we expect the electric or magnetic coupling to supply the energy to the process, as local kinetic energy densities does not suffice. If e.g. magnetically supplied, a magnetic field energy density of $B = \sqrt{8\pi\mathcal{E}} = 1.1 \cdot 10^{10}$ Gauss decaying locally suffices to power the enormous intensities. The pioneering work of Hankins et al. [2003] has been investigating giant pulses from the Crab pulsar and it is not clear whether the same mechanism governs ordinary pulsar radiation that averages to mJy intensities. Recent work on the observed spectrum however Krzeszowski et al. [2014] expect the radio emission to be omnipresent in the pulsar population to also consist of nanoshots of radiation. However the intensities received fall under the threshold where the single shots can be resolved individually. In analogy to the above but with a factor of about 10^6 smaller intensities, one can conclude that the magnetic fields required to decay are on the order of 10^7 Gauss. Estimating the energy density difference by magnetic field decay

$$\Delta\mathcal{E} \approx \Delta B \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}B} = \frac{B^2}{4\pi}\frac{\Delta B}{B} \tag{1.29}$$

shows that also a small fraction of a large total magnetic field may suffice to drive such powers if a process can be identified to convert the magnetic field energy into radio emission. This fact again points into mechanisms based on current decay, as in solar flare physics [Spicer, 1982]. The current is intimately related to the displacement current via Ampere's law (equation 1.23) that provides a way of short-term generation of high field values. Furthermore the current is balanced by the global magnetic field configuration and thus field energy.
Having posed the challenge of highly coherent emission at large intensities, let us review candidate processes that have been proposed to actually produce the radio radiation observed.

1.5 Overview of candidate processes

1.5.1 Curvature radiation

Taking the Goldreich-Julian model literally means that, if all plasma has escaped along open field lines, we are left with the (now disfavoured) inner gap, starting to undergo pair creation in a very large electric field. We will follow the arguments presented by Ruderman and Sutherland [1975]. As the authors calculate, the potential drop at the polar cap will be on the order of $\Delta U = 10^{14}$V directly over the polar cap. This consequently leads to acceleration of the charge carriers to Lorentz factors of $3{\cdot}10^6$, which is on the order of the energy estimated to be necessary to carry all spin-down energy kinetically. With synchrotron losses being tremendous, the charge carriers follow strictly the form of the magnetic field lines. As the curvature of these lines is growing with distance from the star's surface, the leptons emit most radiation after having passed and left the inner gap. The positrons then reach a maximum gamma factor and photon energy of

$$\gamma_{max} = 1.2 \cdot 10^7 \frac{B_{12} h_4}{P_0} \qquad \hbar\omega_c = \frac{3}{2}\gamma_{max}^3 \frac{\hbar c}{\rho} = 5.4 \cdot 10^4 \frac{B_{12}^3 h_4^6}{\rho P_0^3} 2 m_e c^2 \tag{1.30}$$

where h denotes the size of the inner gap and ρ the curvature radius of the field lines. While the most energetic particles decelerate along the distance of one neutron star radius due to synchrotron emission, the lower energetic ones prevail and give away their energy at dipole-dominated parts of the magnetosphere, leading to a curvature radius of $\rho \geq 2\sqrt{r \cdot cP} = 1.4 \cdot 10^9 \sqrt{r_8 P}\mathrm{cm}$. With now mild Lorentz factors of around 800, the frequency of the curvature radiation falls into the radio band at GHz frequency. The fraction of total energy radiated by a single particle is around 10^{-13} and thus too low to explain coherent radiation. What is missing here is a mechanism to bunch the charge carriers before they undergo radio radiation, as for particles bunched below wavelength of the emitted radiation one can treat them as one coherently radiating particle.
Ruderman and Sutherland however argue that the faster passing population of positrons at high gamma factors triggered the two stream instability to kick in, an argument that we will highly discourage in section 1.5.2 and that has been criticized by many authors [see e.g. Melrose, 1992]. This is a well-known problem of theories of radio emission in pulsars: While a viable radiation process may be quickly found by assuming the energy to be present kinetically, the extreme coherence can only be explained if the charge carriers exhibit collective behaviour. The other extreme also occurs in theories, where a priori there is a way of bunching the particles, but the energy source driving the radiation is missing, too small, or the process is happening yet too fast to explain the extreme focussing of energy.
Ironically, curvature radiation was found a possibly viable explanation for nearly all but the coherent radio emission [Lesch et al., 1998] and is widely used in numerical models to explain part of the energy losses, but not the energetically negligible coherent radio emission. As discussed in section 1.3.4, the suspected emission region has been conceptually shifted outward due to missing accelerating fields in the inner magnetosphere.

1.5.2 Maser & plasma instabilities

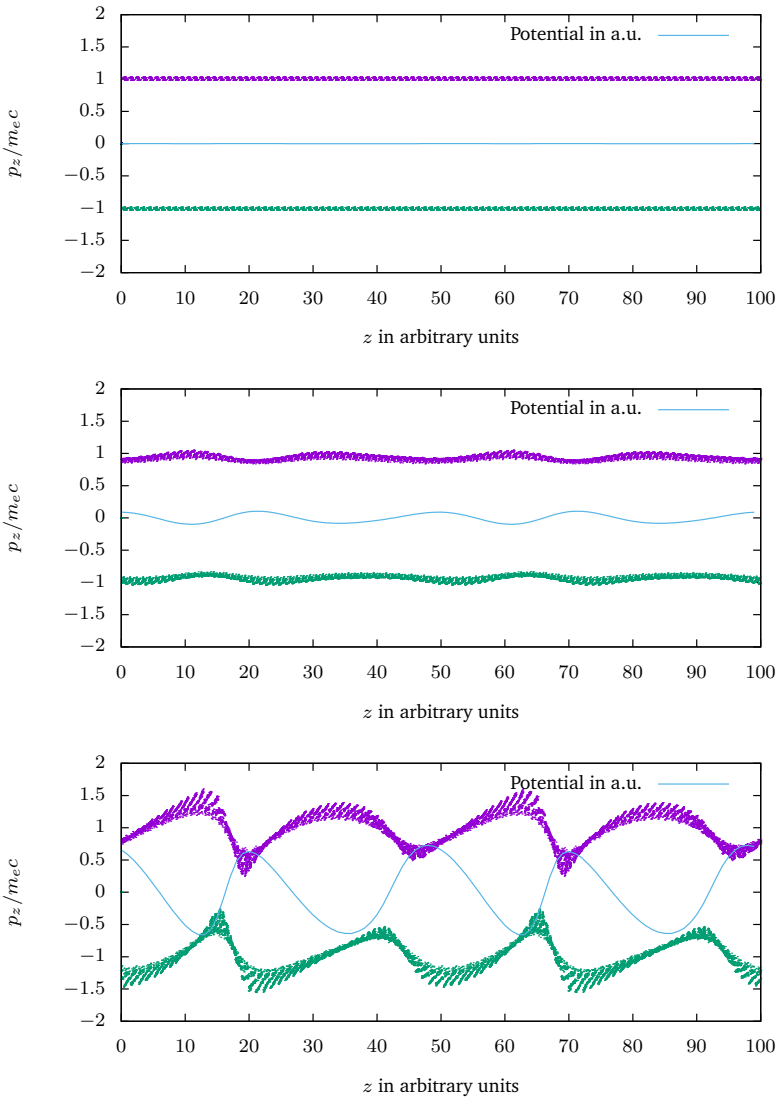
We will discuss emission from plasma instabilities exemplary for the two-stream instability, a by now textbook example of a plasma instability which has been named by numerous authors as the root of more complex processes or the reason for bunching. From both observational and theoretical constraints it has been clear that a current is flowing through the open field lines

and that this current most likely possesses charge carriers of different signs. In this case there exists a frame of reference where both constituents of the current are counter-streaming
As the distribution function in phase-space is two-peaked (one peak for every sign of charge carriers), there exists free energy w.r.t. the thermal equilibrium case in which the whole particle distribution is Maxwellian. This free energy will be converted into electrostatic energy by the instability.

Assume a small disturbance in the particle density that presents w.l.o.g. a slightly negatively charged region. By Poisson's law, this charge density will create an electric field with opposite signs to each side of it. As the current passes, electrons slow down encountering this potential and the stream of electrons is getting more dense in that region, while positrons speed up leading to a decrease of their density in that region. The potential is evidently growing by this process, the disturbance is amplified and the instability grows. For a cold beam, where the temperature of both electrons and positrons are assumed to be zero, this instability grows until all particles are trapped in potentials and the fields are dragged into a triangle like shape, as depicted in Figure 1.3. We will sum up the results presented by Imgrund [2012].
The linear effect of the relativistic two-stream instability is the amplification of Langmuir waves. The linear growth rate $\Gamma_{\max}$, at wave vector $k_{\max}$ for a relativistic positron-electron plasma with no net charge are

$$\Gamma_{\max} = \frac{\omega_{p,\mathrm{rel}}}{2}; \qquad k_{\max} = \sqrt{\frac{3}{4}}\frac{\omega_{p,\mathrm{rel}}}{v_0} = \sqrt{\frac{3\pi e^2 n_0}{m}}\sqrt{\frac{1}{c^2\gamma_0(\gamma_0^2-1)}} \tag{1.31}$$

where γ_0 denotes the Lorentz-factor of the electron-positron stream in the frame where both have the same velocity. We emphasize that the two stream instability is happening very fast, exhibiting a growth rate that is half the plasma frequency. In fact, the system has virtually no time to equilibrate to a global phase in position space, leading to only small wave packets or even just noise and turbulence. If the two stream instability is slightly suppressed however - an effect that may be caused e.g. by a small external electric field - the time to develop phase-coherent disturbances grows and with that the area of synchronous oscillation, and thus perhaps coherent radiation.
The Two Stream instability features complex physics in the regime of growing temperature or relaxing the assumption of a symmetric electron and positron density distribution. Thermal effects [e.g. White, 1985] for example let the instability break down when the thermal energies become comparable to the relative kinetic energy. This can be understood noticing that the resonant wavelength is coupled to the relative speed of two parts of the population. If no frequency dominates, the instability will not form coherent bunches because amplifications at different wavelengths interact only over one global potential. If now the distribution parts are interacting at a continuum of available relative speeds, no global wavelength will be dominantly resonant any more.
There are also intrinsic difficulties at high Lorentz factors: As the growth rate falls down like $\gamma^{-3/2}$ the time to full development of the instability can get very long. Even in a mild electric field, this may inhibit the instability completely as the resonant wavelength is constantly shifted by current acceleration by the electric field. As the source of free energy is kinetic, we expect the radio emission only to be explainable by a population of particles with high Lorentz-factors, because energy would not suffice otherwise to yield the energy output (see section1.4). This evidently complicates arguing for the then highly suppressed two stream instability at wrong wavelengths to act as a source of coherent emission.
As some electric field however should be present to get the most ordered structures and thus highest coherence, we are left with the need to justify occurrence of the two stream instability in an unclear scenario, as radio emission by such structures would cause an electric field to rise



timestep 0:
Simulation is initialized with $p_\pm = \pm 0.8 m_e c$

timestep 100-250:
TSI starts to build up, linear regime

timestep 300:
Instability enters strongly non-linear regime

Figure 1.3: **Particle in Cell simulation of the Two Stream instability:** Phasespace plot of positron / electron quasiparticles in violet resp. blue for different timesteps of the simulation Imgrund [2012]. The nearly cold beam is interacting non-linear with the potential generated by itself. If the charge carriers would have higher temperature, there would be a full build-up of "islands" in phasespace that are depleted of particles.

that effectively blocks the Two Stream instability until the fields have relaxed again. While this sketch of a radiation process is attractive, realistic calculations are cumbersome.
The basic mechanism of the two stream instability has been extended both in the linear and non-linear regime leading to maser instabilities and Langmuir solitons. There exists also rich literature (e.g. [Bernstein et al., 1957, Schamel, 1971, Schamel and Luque, 2005]) on electron cavities and phase space holes driven by two stream instability or similar processes. Basically, there exist stationary potential solutions that do not exhibit the two stream instability.
Extending this idea of static potentials, one finds that the electric current attainable is limited by space-charge effects, a result that can be derived in a very similar framework to the works on static potential solutions. Beloborodov [2008] for example argue that such an effect known as pierce diode [Pierce, 1944] is not only able to chop the current into bunches but also may be responsible for decreasing conductivity and thus enabling electric fields to build up. Explaining bunching or at least motivating a high γ background population is vital to provide coherence and having available an energy reservoir large enough to explain radio emission.

1.5.3 Langmuir solitons

All presented instabilities rely on the basic mechanism of self-amplifying potential differences. The most sophisticated treatment of this mechanism is based on work done by Zakharov [1972] and exemplifies the idea of a collapse of Langmuir waves to soliton solutions that present a stable spatial grid for the high γ background population to emit coherent radio radiation. An extensive treatment of the idea can be found in Pelletier et al. [1988] and references therein. We will shortly present their findings in the following.
When a relativistic beam is interacting with a slower plasma population, the aforementioned mechanisms may take place, or soliton-like solutions will be formed. "Either the unstable waves react to the beam particle motions, and quasilinear effects arise due to wave particle interactions, or a physical state of weak turbulence develops due to wave-particle or to wave-wave-interactions, or else wave saturation occurs through wave-wave interactions and leads to a state of strong Langmuir turbulence" [Asseo et al., 1990, paraphrasing Goldman [1984]]. There exists a dimensionless parameter w defined as the ratio of Langmuir wave energy density to thermal energy parting the treatment into three cases, the weak regime $w \ll 1$, the strong regime $w < 1$ and the ultra strong regime $w \geq 1$. In case of the strong regime, the wave packets' envelope is unstable leading to the case of the modulational instability. Strong turbulence is an expression of dominant non-linear effects. As such, the characteristic non-linear time scale $\tau_{NL} = (\omega_p w)^{-1}$ has to fall below the dispersion time scale $\tau_L = (\omega p \Delta k^2 \Delta k_{||} \lambda_D^2)^{-1}$ for the non-linearities to have an effect, where Δk and $\Delta k_{||}$ are the total respectively parallel component of the Langmuir wave vector and λ_D the Debye length. The authors derive that $w > (\lambda_D^2 \Delta k_{||}^2)$. There also exist transverse dispersive effects that have to be overcome and thus $w > (\lambda_0^2 \Delta k_\perp^2 / 8\pi^2)$ where $k_\perp$ is the perpendicular wave vector of the Langmuir wave and λ_0 the resonant parallel wavelength.
Having fulfilled theses conditions, the ponderomotive effect leads to non-linear contributions to the dispersion relation of the plasma. This non-linear term can be treated approximatively around the non-disturbed linear solution which transforms into the non-linear Schrödinger equation. The one dimensional Schrödinger equation adapts a soliton-like solution in this case. While the linear evolution of the system would have dispersed an existing wave packet, the non-linearity balances dispersion by focussing the wave packet. In one dimension there exists a stable static envelope solution for such a Langmuir wave packet that is similar to a soliton. Shape and amplitude of this quasi-soliton are coupled. We find the system effectively to be confined to the one-dimensonal case by a strong magnetic field. When the field weakens, the

quantity	value
parallel length in Debye lengths	$5\lambda_D$
parallel length in plasma wavelengths	$44\omega_P$
perpendicular length in plasma wavelengths	$1.34 \cdot 10^3\omega_P$
parallel temporal width for observer	$\approx 10^{-7}$s
perpendicular temporal width for observer	$\approx 5 \cdot 10^{-8}$s
number of solitons	$2.6 \cdot 10^3$

Table 1.1: **Characteristics of Langmuir microwave substructure:** This table from Asseo et al. [1990, Table 1] shows the characteristic length and time scales of the long-lived soliton substructures that serve as a background lattice to generate radio emission from an assumed high-γ stream passing over them or by itself.

soliton is about to decay as dispersive effects in the other spatial dimensions occur.
The length of the wave packet also influences its stability. For long lengths, the soliton is about to decay as it becomes vulnerable to modulational instability. When the wave packet is too short and the system has two or more spatial dimensions, the packet will be prone to collapse as the non-linear term is dominating and will not be balanced by dispersion any more. As the system may only be treated as effectively one dimensional when the frequencies of the plasma system are small compared to the cyclotron frequency, there exists effectively a minimum and maximum packet length in which the plasma may form long-lived solitons in an ordered lattice also referred to as *Langmuir microstructures* by Asseo et al. [1990]. The characteristics of such a lattice can be found in tabular 1.1. The temporal width in all directions nearly match due to relativistic motion of the structures w.r.t. the observer and settles at tens of nanoseconds.
Pelletier et al. [1988] argue that two types of radiation sources arise; one source is the highly relativistic background striving over the Langmuir microstructures, the other one is emission by the structures themselves. In the original paper, the first source was thought to have a negligible power output w.r.t. the second type, a statement revised in Asseo [1994] by noting that there might be more phase-coupling in the background plasma present as previously considered. The Langmuir microstructure itself radiates as the three dimensional solution has also wave-like components that are assumed to be able to penetrate the surrounding magnetosphere. In this case, however, it is not clear how the enormous energies can be reached and whether the electric field built up contains enough energy from the relativistic pump beam to explain the enormous energy drain. It is important to note that irrespective of the actual channel realized in reality, the proposed power is ultimately assumed to be fed by the relativistic kinetic flow.

1.5.4 Modulational instability

It has been clear from the beginning on [Goldman, 1984] that the Langmuir microstructures could be prone to the modulational instability and decay faster than expected. Weatherall [1998] directly opposes the view presented by Asseo et al. [1990] and explains the radio emission by conversion of the energy from the Two Stream driven pump wave to radio emission.
The crucial argument criticized is the quasi one dimensional treatment of a still existing electromagnetic coupling in transverse direction despite the plasma being kinetically trapped to the field lines. Weatherall [1997] calculates that the non linear plasma dispersion relation leads to the four-wave interaction typical for a modulational instability. The relation that is to be fulfilled is $\omega(\vec{k}_0 + \vec{k}) - \omega(\vec{k}_0 + \vec{k}) - \omega(\vec{k}_0 - \vec{k}) + \omega(\vec{k}_0) = 0$ where the pump mode is effectively the Two Stream instability / Langmuir soliton collapse mode $\vec{k}_0$ and a transverse perturbation

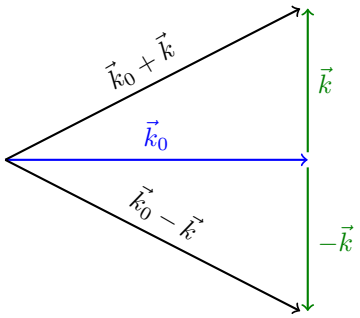


Figure 1.4: **Modulational instability - Momentum conservation:** A strong pump wave vector $\vec{k}_0$ decays into two $\vec{k}_0 \pm \vec{k}$ components that give rise to transversal growing envelope modulations with wave vector$\pm\vec{k}$ [Weatherall, 1997].

$\vec{k}$. The momentum conservation scheme is depicted in figure 1.4. There exist non linear terms that interact at much lower frequency than the plasma frequency that then drive the plasma into transversal instability. Weatherall [1997] describes this process as follows:

1. The wave energy increases exponentially during the growth of a single monochromatic wave (pump wave) because of an external instability.
2. As the wave amplitude grows, it eventually stimulates the growth of modulational instabilities. Because these transverse modulations grow at the expense of wave energy in the pump wave, the modulational instability can saturate the pump wave growth.
3. The system reaches a fully developed turbulent state characterized by the rapid formation of intense localized wave packets [...]. These collapsed wave structures are damped by the escape of radiation.
4. The system relaxes to a state of lower energy density. In this weakly turbulent state, when the instability growth time is long, one-dimensional soliton-like structures sometimes emerge.

As with the Langmuir microstructure, the modulational instability also relies on relativistic transformations to explain the power of the received radiation. Thus, while these plasma physically very sophisticated and interesting topics can even be measured in the laboratory [e.g. Antipov et al., 1981], the exact application to the pulsar case in an not yet settled theoretical model of the magnetosphere is both promising but not sufficiently motivated.

1.6 The radio emission problem

As we have motivated in this chapter, the problem of radio emission connects to the problem of the magnetospheric background in which emission takes place. The true magnetosphere is situated somewhere between the extremes of a completely current driven, ideal MHD system and the strong gap solutions based on the pure Goldreich-Julian model. While this fact has been known for decades recent numerical simulations of the magnetosphere point more into the current-dominated regime. This in retrospective weakens the ground-breaking arguments of Ruderman and Sutherland [1975] that lead to a high γ flow through a low γ region.
All viable candidates for radio emission however highly depend on the existence of a free kinetic energy source by the high gamma beam. Simple arguments presented in section 1.4

corroborate the need for high Lorentz factors to be present in case of a kinetic energy source. There are however two strong arguments against highly relativistic flow in the radio emission zone, connected to the high electric fields needed to accelerate the particles to high energies.
Electric fields accelerate particle, but particles with opposite charges will be pulled in opposite directions. This counter-streaming charge carriers then could undergo plasma instabilities. However, in their inertial frame they exhibit large γ factors leading to a very low plasma frequency, as $\omega_{p,rel} \propto \gamma^{-3/2}$. Leaving out the exact details of interaction, this frequency never the less sets the length scale of any kinetic plasma process. As Goldreich Julian densities above the polar cap reach the observed radio frequencies for low γ currents, relativistically counter-streaming plasma exhibits too low frequencies.
Boosting by Lorentz transformation and assuming a high relative speed of the radiating frame towards the observer may both focus radiative output by beaming effects and convert the observed frequencies to the proper range by the relativistic optical Doppler effect. This also weakens the problem of relative gamma, as a relativistic plasma flow carrying a small current exhibits low relative particle energy in the plasma inertial frame. A fast moving plasma inertial frame (in which the total mass flow is zero) however constitutes a relativistic stream of particles (and not necessarily high current) in the observer frame. Acceleration by an electric field can only supply a current but not a relativistic flow, if neutral plasma is accelerated. Thus the need for an additional mechanism generating flow or converting current into flow arises.
The main candidate process for converting a high γ current into a flow is generation of positron electron pairs by photon scattering on the stream of electrons, where the momentum of the initial particle is converted to flow, not current of the generated pairs. This however binds initially kinetic energy in the form of rest mass and the newly generated particles will be accelerated themselves by the electric field and are able to carry a current. Having more charge carriers available, it will be harder to explain that a static accelerating field is maintained.
Numerical simulations both of global models and micro- and meso-physics of cascade processes (see section1.3.4) indeed support that near the polar cap, only small Lorentz factors occur. Consequently we find the system in a current dominated regime that however is not able to supply the local kinetic energy density needed to explain radio emission.
Being in a mild relativistic current regime a priori sets plasma length and time scales that fall into the radio emission's range of length and time scales, which boils down to the simplest assumptions possible. The main open question of this regime boils down on pointing out an energy source for radio emission, as kinetic energy will not suffice.
Physically, the free energy in the magnetosphere clearly stems from magnetic coupling to plasma currents and overall topology of the plasma being in an dynamical equilibrium due to the differential rotation of the pulsar w.r.t. the outside. The most striking feature in the pulsar case is that the light cylinder argument clearly demands a current to be flowing through the magnetosphere. This fact has been expected by all theoretical models that predict radio emission and is supported by observations [Kramer et al., 2006a]. As charge and electric current is not a function of the Lorentz factor but just the velocity β, increasing the flow instead of the current will not assist any process aimed at transporting field energy outward. In this light, a process that leads to tremendous increase of kinetic flow on small scales to reduce magnetic fields induced by differential rotation seems counter-intuitive. Increasing current instead of flow would pave a much more balanced way to reach equilibrium as this amounts to straightforward coupling in a process demanding a current.
Thus, current and magnetic field are strongly coupled by Ampere's law. Slowing down or stopping this current leads to high displacement currents due to induction. As outlined in section 1.3.4 the arising electric field will free energy stored in the magnetic field. We argue that this process could possibly pose as an alternative energy source to the electric fields by global mag-

netospheric configuration. We will investigate this relation further in chapter 6.
Even though there exist many proposed radiation mechanisms, neither the magnetospheric boundary conditions of the process are clear nor the exact flow of energy. Furthermore, for most processes the back-reaction of the built up field (as radio emission induces a resistivity) has not been examined properly, even though this may change the underlying physics process drastically, especially for high relativistic flow. The observational and numerical results of the recent years point to an only mildly relativistic current flowing and furthermore elementary emission time scales that fall below a nanosecond. Both findings question the basic assumptions made for the most successful attempts to describe radio emission. While these developments clearly justify investigating the radio emission in more detail, new candidate processes are hard to find facing the overwhelming plethora of observational constraints to be fulfilled. We will describe some of the most important observational phenomena in the following chapter.

CHAPTER 2

The appearance of pulsar radiation

In this chapter we will introduce only the most striking and important observational features of pulsar radio emission. First observed at MHz frequencies and on variability of seconds, the frequency range, time resolution and total time span of observations has been increased tremendously. An overview over the observed phenomena is given in Figure 2.1.
There have been extracted **Times of Arrival(ToAs)** from the observations of rotation period, time points of observation where the pulsar has a constant rotation phase w.r.t the line of sight. Why this is possible will be explained in the following section.
Tracking the received rotation phase enables determination of several orbital parameters, fitting for peculiar motion, for slow-down by dissipation of rotational energy (see chapter1) and many more. The influence of the **interstellar medium (ISM)** is imprinted in the signal. It manifests itself as dispersion measure, **DM**), turbulence or density fluctuations that cause short time-scale energy fluctuations of the signal, called **scintillations**, and apparent movements of the star on the sky to name just a few.
The variability of the radio signal has been examined on smaller and smaller time-scales, finding **nulling, moding** and **microstructure** in the variability of single pulses and ultimately **nanoshots** of radiation taking place in so called giant pulses. Recent observations suggest that all radio emission might consist of those nanoshots. Facing the enormous number of results in the area of pulsar radio emission, we will focus on processes and phenomena relevant to the observation and physics of the emission process.

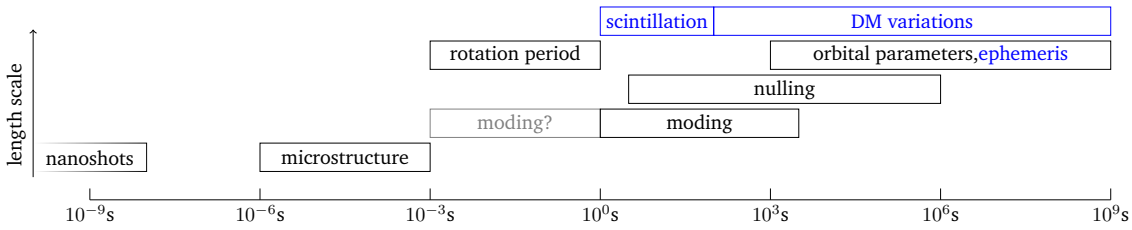


Figure 2.1: **Various pulsar and radio time-scales:** Pulsar radio emission is affected by all of the named processes. To correctly interpret and model the pulsars radiation, one in principle has to master subsequent stages of the displayed time- and length scale hierarchy. External influences are depicted in blue.

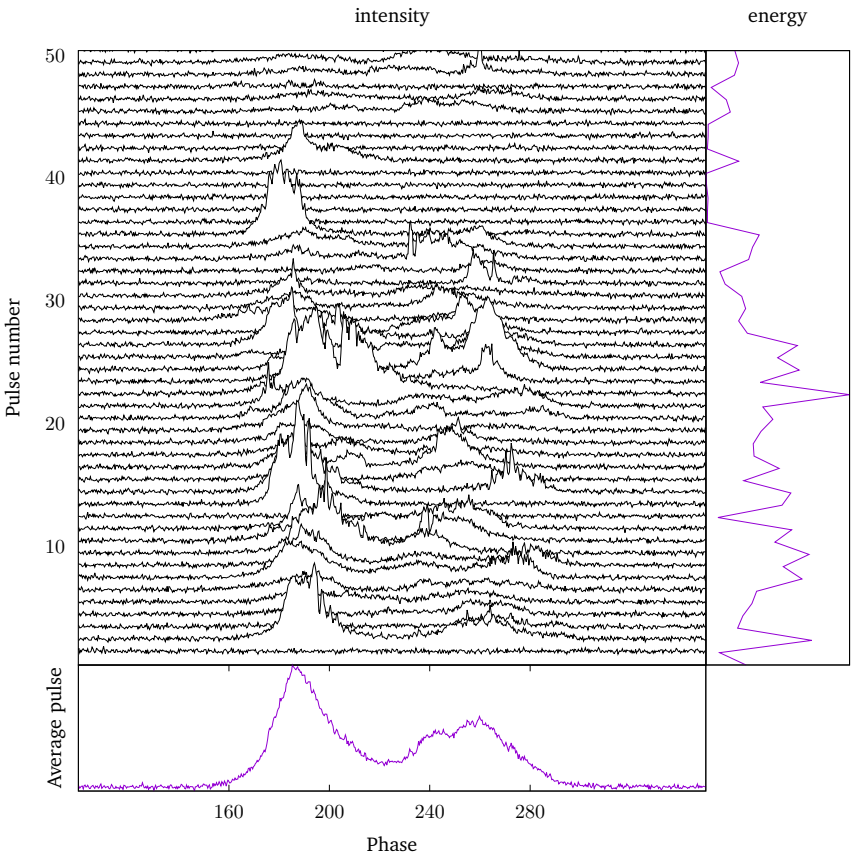


Figure 2.2: **Single pulses** from PSR1133+16 that exhibit typical features of pulsar radio emission: While the energy of a single pulse as well the form is varying on a pulse to pulse basis, there exists a characteristic average envelope when integrating over several pulses.

2.1 Single pulses

We have encountered the periodicity of the emission up to this point multiple times in this thesis. Investigating single pulses as depicted in Figure 2.2 however makes clear, that the proclaimed exact measurement of e.g. the period down to more than ten digits cannot be accounted to observations of single pulses. Single pulses from a pulsar show high variability in form and intensity. It is only the integration over many of those pulses, that subsequently leads to a consistent average pulse named **profile**. This average pulse exhibits less noise both in the sense of averaging out receiver noise and fluctuations of the individual pulses and can be compared to a reference profile called a **template**. By analysing the phase shift of the observation to the reference profile, the pulsar rotation phase can be tracked over many distinct observations called **epochs**. These phase-over-time measurements can be related to each other with a rotation model describing the physics of the neutron star itself (angle between rotation- and magnetic axis, period, period derivatives) and the surrounding (possibly orbit for binary and n-ary systems, relative position and movement to earth, ISM influences, relativistic contributions). Fitting the parameters to the ToAs constitutes the process of **pulsar timing**, in the best case leading to a **phase connected solution**, a situation where not only the pulse phase of

each pulse is known, but also its number, counting every physically happening rotation over all observations. It is this model and its parameters that can be determined to very high accuracy. A single pulse however clearly cannot provide very exact phase information. One of the puzzles this thesis seeks out to contribute to amounts to finding ways of extracting statistically accurate information out of these highly noisy and fluctuating elementary observations. This was done to develop new methods that may be able to fit parameters of emission models.
As an example, extracting ToAs and their significance out of single pulses or few pulse integrations will be examined with a novel method in Chapter 4. There are many pulsars that are not bright enough to resolve their single pulses. In this case one can still integrate over a long observation time and find an average profile, but the single pulse stays hidden in receiver and antenna noise. With growing sensitivity of instrumentation and for very luminous pulsars however, we are able to receive the single pulse fluctuations.
The more bright these fluctuations are, the more they cause an effect called 'phase jitter'[Keane, 2013], a random movement of total observation phase. This jitter can dominate the believed accuracy of the observation by ToA generation. We will discuss the mathematical details of this jitter in Chapter 4. Its origin is the modulation of fluctuations that constitute the actual radio emission by an envelope function that represents the rigidly corotating magnetospheric conditions. Demodulating this signal gives insights into the actually happening physical process on mesoscopic scale, enlightening local physics of the emission region and not the magnetospheric model. The mathematical model presented in this thesis is only a first step into this direction due to the difficulties in both analytical treatment and numerical evaluation of the data at hand. Future generation radio telescopes will have a higher sensitivity and thus be more likely limited by phase jitter than receiver noise [Liu et al., 2012]. While classical techniques of integration sufficed to mitigate the jitter effect, understanding the meso-scale fluctuations in the radio emission process will most probably play a large role in understanding and minimizing timing errors. Let us investigate slightly larger integration times in the following.

2.2 Polarisation and beam geometry

Integrating the pulsar's single pulses over sufficient times to yield a stable profile naturally leads to the question of profile stability. Profile stability is the main assumption underlying the confidence of pulsar timing, as two different profiles cannot be compared in phase [Liu et al., 2012]. Most profiles seem to be stable enough or at least exhibit only a small set of profile variations. Exceptions are young pulsars which sometimes undergo jumps in their rotation frequency (**glitching**) due to changes in the neutron star's crust. These events have also been associated with profile shape changes[Lyne et al., 2010].
As the charge carriers movement is confined to be parallel to the magnetic field, pulsar radio emission is highly polarized. Recent results on polarisation can be found in Noutsos et al. [2015] and references therein. We will loosely follow some of their arguments. Ruderman and Sutherland [1975] proposed that if radio emission is connected to the plasma frequency, the frequency of observation corresponds with the radius of emission from the pulsar, as frequency scales with plasma density. This idea has been coined **radius to frequency mapping (RFM)**. As higher densities lead to higher plasma frequencies, the assumed emission height of high frequency components is at smaller heights compared to the lower frequencies. Thus, for higher frequencies we expect propagation effects through the magnetosphere to play a larger role, possibly leading to the observed depolarization at high frequencies. There are however works opposing RFM [Weltevrede and Johnston, 2008, and references therein] as height estimations by independent methods yield no consistent results.

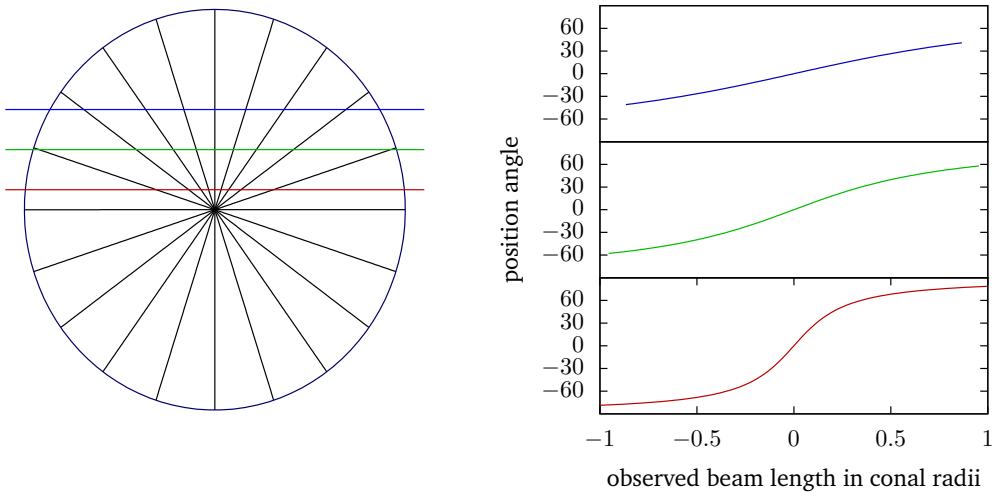


Figure 2.3: **Toy model for position angle:** Drawn is an ideal very small cut through the beam cone. The coloured lines are the paths that the line of sight takes through the emission region when the pulsar's magnetic axis is perpendicular to the rotation axis. The projection of the magnetic field lines onto the plane perpendicular to the line of sight are depicted in black. The right panel shows the angle of polarisation over position. Note that the arguments follow by rotational symmetry of the emission region.

The angle of polarisation of the linear emission is dependent on magnetic field line direction. Hence the projection of the magnetic field lines w.r.t. the plane perpendicular to the observer's line of sight is the relevant quantity to determine polarisation direction. As can be seen from Figure 2.3, a naive model of the beam cone and emission region shows that the angle of polarisation is expected to rotate when the line of sight sweeps over the emission cone. While qualitatively this picture is correct, quantitative observations will be dependent on further parameters like the angle between rotational and magnetic field axis and detect further components and deviations from the idealization chosen [see e.g. Noutsos et al., 2015, Figure 5].
From these qualitative arguments it becomes clear that a proper mapping of the rotation phase data along the line of sight gives insights into polarisation and intensity data of the emission region. However these data must always be taken with care as propagation effects may change the picture drastically [Arons and Barnard, 1986, Barnard and Arons, 1986, Blaskiewicz et al., 1991, Wang et al., 2010].
There are cases in which the line of sight slowly moves perpendicular to its projected path over the emission zone. This in principle makes a two-dimensional examination of the emission zone possible, assuming stability of the emission zone over longer times than the movement of the line of sight takes. One such example are binary systems or free precession [Stairs et al., 2000]. In this case, the change of the line of sight has a systematic effect on the measured profile shapes and polarisations.
Weisberg and Taylor [2002] for example use this effect to map the changing cuts of the radiation cone in two dimensions to generate an image of the cone. They deduce an hour-glass shaped beam to be present in the examined system. Comparing the expectancies of the theory of propagation, polarization and intensity with observational results not yet leads to a consistent picture but supports various geometrical possibilities and locations of the emission region.

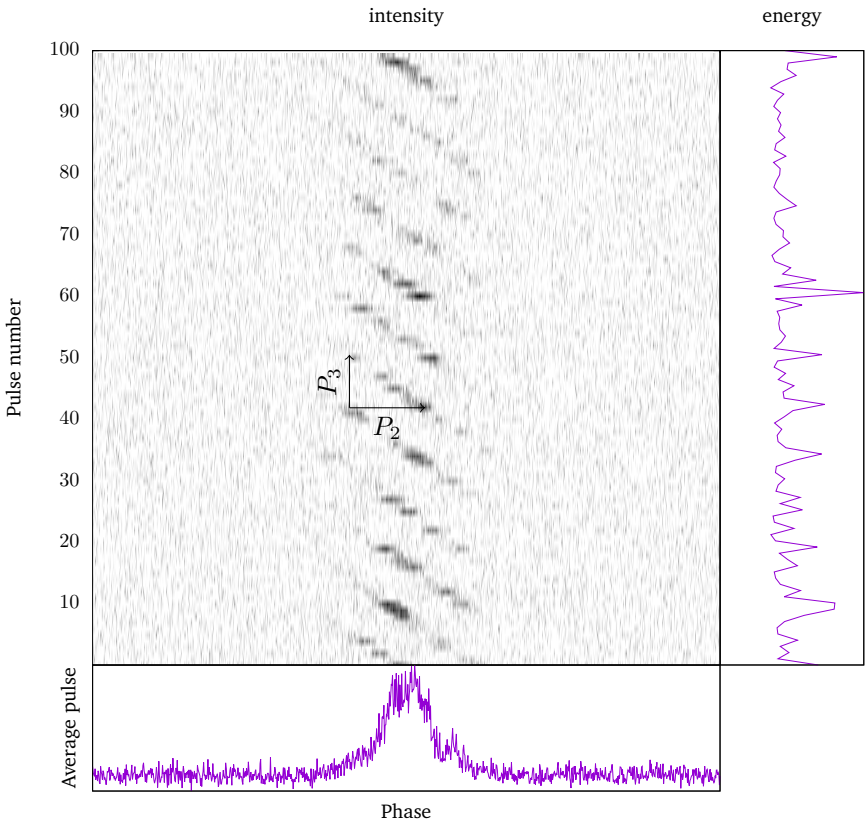


Figure 2.4: **Simulation of drifting subpulses:** Many pulsars show subpulse intensity variations that move over the envelope in the course of subsequent pulsar pulses. Note that there can be more than one subpulse in a single pulse. The time P_2 characterizes the distance between single subbeams, while P_3 is related to the whole time it takes for one beam to rotate around the emission zone (see section 2.3).

2.3 Subpulses and subpulse drift

While we have treated the single pulse to be completely random up to this point, there exist some clearly systematic phenomena in the variations of single pulses. Subpulse drift as shown in Figure 2.4 for example is observed in about half the pulsar population [Weltevrede et al., 2006b] and was also observed in millisecond pulsars [Edwards and Stappers, 2003]. There exist maxima in single pulses called subpulses that drift slowly over the whole envelope in subsequent pulses. As is evident from the average pulse below the figure, this phenomenon can only be observed when resolving single- or few pulses or folding the single pulses with the correct period. Consequently we expect new results on the phenomenon since many pulsars could not be observed on the single pulse level as their intensity is too faint. There are authors[e.g. Weltevrede et al., 2006b] who suspect the phenomenon to be present in virtually all pulsars, but may not be observable in the single pulses as the different subpulses may overlap. The phenomenon has been early described observationally by [Backer, 1973, Drake and Craft, 1968] and a possible explanation has been given in the framework of spark-gap models by Ruderman and Sutherland [1975]. They describe the inner gap giving rise to sparks that drive coherent

emission: Before assuming an inner gap, the strength of the electric field tangent to the polar cap has been given by equation (1.14). Due to the inner gap however, there is an additional term arising, that can be deduced by noticing that near the polar cap, closed line integrals over the electric field amounts to zero. Additionally to the vacuum terms, there arises a term caused by the inner gap being present. Imagine a line integral taking a path starting at half the polar cap radius, continuing along the magnetic field lines over the gap, then perpendicular to the border of the open field line region, back to the surface of the star and then closing over the polar cap again [compare Ruderman and Sutherland, 1975, Figure 4]. Besides the vacuum electric field, only two of four sides of the path amount to non-zero. As the path in the stellar surface is trivially zero and $\vec{E} \cdot \vec{B} = 0$ at the open field line's boundary, the only non-trivial line integral parts left are the one from the surface over the gap region and the second segment to the border of the open field lines.
The line integral over the gap yields the potential drop as $\int \mathrm{d}\vec{s} \cdot \vec{E} = \Delta V$. Since the sum however amounts to zero, the part between upper spark border and closed field lines must deviate from the vacuum value. This is achieved by the rotation deviating from the vacuum value. As the authors conclude this change in velocity amounts to

$$\delta v = \frac{\Delta V}{B_{surf} r_P} c \tag{2.1}$$

which can be evaluated in the inner spark gap model to $\frac{\hat{P}_3}{P} \approx 5.6 B_{12}/P_0^2$s and comparing this velocity with the circumference of the polar cap leads to the period

$$\hat{P}_3 = 2\pi \frac{B_{surf} r_p^2}{\Delta V c} \tag{2.2}$$

If we have n such sparks present, P_3 from Figure 2.4 relates to $\hat{P}_3$ as $P_3 \approx n\hat{P}_3$. P_2 is then simply the distance from spark to spark when compared to the whole rotation period P. These sparks are arranged in a beam carousel that slowly precesses around the magnetic field axis and thus periodically wander through the line of sight. While there is also contrary evidence like subpulse drift in both directions [see e.g. Weltevrede et al., 2006a, for a recent review], this model is still very popular.
Careful analysis of the single pulses can parametrize a projection from the time series to the beam carousel rest frame that is rotating under the line of sight by precession. Deshpande and Rankin [1999, Figure 5] integrate the projections of the lines of sight to the co-rotating beam carousel to generate an impressive map of the radio emission zone. This of course assumes that the emission regions themselves do not undergo drastic changes during the time of integration.

2.4 Moding and nulling

A dramatic change however can be detected when the profile shape is varying between different discrete shapes, a process called **moding** or the radio emission ceases completely, called **nulling**. The phenomenon has been described first by Backer [1970]. While nulling can be ascribed to single pulses, moding and statistics of it are only evident over several pulses as mode changes might be minuscule. The question of when to assign a pulse to a certain mode is still not clear. While the time-scales for moding are evident for many pulsars, the exact time point of change from one mode to the other is not. A novel mechanism for detecting nulling and assigning single pulses to certain modes is proposed in Chapter 4.
The question is of some importance to the understanding of the emission mechanism. On one

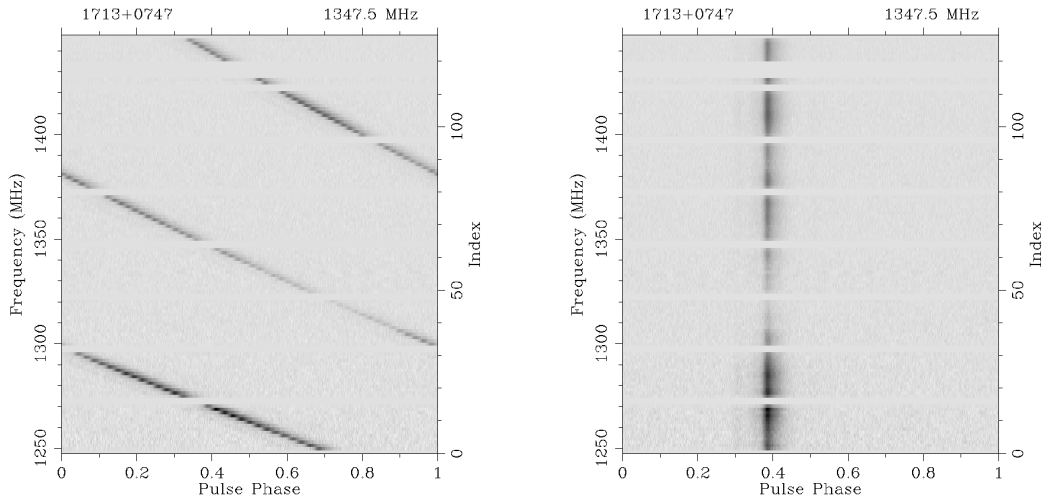


Figure 2.5: **Exemplary frequency over phase plot:** This plot shows a heatmap of the intensity received from PSR 1713+0747, that has been time integrated. While the right panel has been corrected for dispersion measure, the left panel shows the data before dispersion. The white horizontal stripes are produced by the receiver while the pulsars' signal can be seen as black curved line expressing the non-linear dependence of dispersion on frequency in the left panel. When corrected for the dispersion measure, the signal is synchronous over all frequencies, and can be seen as the vertical line in the right panel.

hand, sub-pulse drift is shown to be systematically affected by mode changes at least for some pulsars [Fowler et al., 1981, Taylor et al., 1975]. On the other hand nulling has been identified to correlate with changes in the spin-down rate[Kramer et al., 2006a]. Both facts point to the mechanism being related to global states of the magnetosphere. If mode switching is in fact related to changes in the magnetosphere, rapidity and fluctuation spectrum could prove valuable determining the associated length and time-scales and finally the mechanisms generating this complex interplay. Cordes [2013] discuss the statistics and systematics of random mode changes and also stress that assigning correct modes to pulses at hand is vital to these analyses.

2.5 Interstellar medium effects

2.5.1 Refractive index, Dispersion Measure and Rotation Measure

Up to now we have assumed the signal from a pulsar to arrive at earth in the actual condition it has been emitted at the pulsar. This is not the case due to atmospheric effects and the ISM. In the following we will visit the most important effects for pulsar radio astronomy and discuss how to correct for them.
ISM effects usually are strongly frequency dependent and thus can be examined using frequency and time resolved intensity maps. The main cause why the ISM affects radio radiation is the change of the refractive index induced by the presence of free charge carriers. As the electron is the dominating light particle, it is the (in this regime) non-relativistic plasma frequency associated with the electron density that dominates the calculation of the refractive index. The interaction of the charge carriers with the electromagnetic waves changes the electric perme-

ability $\eta(\omega)$ for waves of frequency ω as [Jackson, 1998]

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_p^2}{\omega^2}, \qquad \omega_p = \sqrt{\frac{4\pi n e^2}{m}} \tag{2.3}$$

The refractive index η then amounts to

$$\eta = \Re\left(\sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}}\right) = \sqrt{1 - \frac{\omega_p^2}{\omega^2}}, \qquad \omega > \omega_p \tag{2.4}$$

where we neglected magnetic permeability and ω must exceed the plasma density to propagate. At typical ISM densities of $n = 0.01\mathrm{cm}^{-3}$ to $0.2\mathrm{cm}^{-3}$[see e.g. Taylor and Cordes, 1993] this amounts to frequencies of $1 - 10$kHz that are easily surmounted by the frequencies at which pulsars are observed. It is the dependence of the refractive index on local plasma density that causes most of the ISM effects. As the refractive index relates to the group velocity v_G of a plasma wave as $v_G = c\eta$, the propagation time of the pulsars signal to earth at distance d will be delayed frequency dependent defining the dispersion measure

$$\Delta t = \int_0^d \mathrm{d}l \left(\frac{1}{v_G} - \frac{1}{c}\right) \overset{\omega \gg \omega_p}{=} \int_0^d \mathrm{d}l \left(\frac{1}{c}\left(1 + \frac{\omega_p^2}{2\omega^2}\right) - \frac{1}{c}\right) = \underbrace{\frac{e^2}{2\pi mc}}_{=:\mathcal{D}} \frac{1}{f^2} \int_0^d \mathrm{d}ln(l) \tag{2.5}$$

where f is the observation frequency and $\mathcal{D} \approx 4.1 \cdot 10^3 \mathrm{MHz}^2\mathrm{pc}^{-1}\mathrm{cm}^3\mathrm{s}$. We now may define $\int_0^d \mathrm{d}ln(l) =: DM$, the dispersion measure. As the delay is proportional to the inverse square of the frequency, instantaneously emitted signals will be delayed frequency-dependent. Clear evidence for this can be seen in Figure 2.5 in the left panel, as the pulsar's pulse wraps around the pulsar period due to the missing dispersion measure correction. The dispersion measure can be used to estimate the distance to a given pulsar by assuming a constant density to be present or using more sophisticated models of ISM densities [e.g. Greiner et al., 2015, Taylor and Cordes, 1993].
When treating a signal $s(\omega)$ in Fourier space, we can also represent the dispersion measure as linear operator $\mathcal{T}(\omega', \omega) = \delta(\omega - \omega')\exp\left[\mathrm{i}\frac{2\pi\omega_\mathrm{p}^2}{\omega}\mathcal{D} \cdot \mathrm{DM}\right]$ working on the signal to yield $s'(\omega)$, the received signal:

$$s'(\omega') = \int \mathrm{d}\omega \mathcal{T}(\omega', \omega) s(\omega) \tag{2.6}$$

Most ISM influences may be translated into this operator language that in principle ad hoc gives an inversion of the effect by inverting the associated operator, if possible. Another example of this is the Faraday effect leading to a quantity called rotation measure. In the formalism described above this can be formulated introducing another term into equation (2.3) that distinguishes between different circular polarisation states:

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_p^2}{\omega^2} \mp \frac{2\pi\omega_p^2\omega_B}{\omega^3}, \qquad \omega_B = \frac{eB_{||}}{mc} \tag{2.7}$$

where ω_B is the cyclotron angular frequency. Following [Lorimer and Kramer, 2004] the absolute time corrections implied by this term are on nanosecond time-scales and below. While absolute delays between polarisations are minuscule, the relative phase delay between the two circular polarisations are not negligible since the delay causes a phase difference of

$$\Delta\Phi = 2\pi\omega\Delta t_B = \frac{e^3}{\pi m^2 c^2 f^2} \int_0^d \mathrm{d}ln(l) B_{||}(l) = 2\lambda^2 \cdot \underbrace{\frac{e^3}{2\pi m^2 c^4} \int_0^d \mathrm{d}ln(l) B_{||}(l)}_{=:\mathrm{RM}} \tag{2.8}$$

where λ is the observed wavelength and Δt_B is derived analogous to Δt. The rotation measure (RM) then again is the measure of the influence of the ISM without the frequency/wavelength dependence. Because $RM \propto \int \mathrm{d}B_{||} n$ and $DM \propto \int \mathrm{d}n$ we may divide the rotation measure by the dispersion measure to find the density weighted average magnetic field component along the line of sight. Evaluating the numbers yield

$$\langle B_{||} \rangle = 1.23 \mu\mathrm{G} \left(\frac{\mathrm{RM}}{\mathrm{rad\,m^{-2}}} \right) \left(\frac{\mathrm{DM}}{\mathrm{cm^{-3}pc}} \right)^{-1} \tag{2.9}$$

Whereas for pulsar observation, dispersion measure and rotation measure pose an inevitable source of noise and disinformation, it becomes clear that the radio signals received may be of value investigating ISM structures. Fitting for example for dispersion measure changes is on one hand vital for precision timing, on the other hand it can give insights on ISM turbulence.

2.5.2 Scintillation and scattering

While first observed by Lyne and Rickett [1968], Scheuer [1968] introduced the well known model for scintillation and scattering based on a scattering screen half-way to the pulsar. This approximation may be made as the total scattering angle is "less than 10^{-5} radians for interplanetary and interstellar scintillation" [Rickett, 1977] and thus very small. We may approximate the refractive index:

$$\eta \approx 1 - \frac{\omega_p^2}{2\omega^2} = 1 - \frac{e^2}{2\pi m_e} \frac{n}{f^2}, \qquad \omega \gg \omega_p \tag{2.10}$$

This leads to a linear dependence on local density fluctuations. Assuming fluctuations on the way to earth, the wave vector k is fluctuating:

$$\Delta k = \Delta \left(\frac{2\pi}{c} \eta f \right) = \frac{e^2}{mc} \frac{\Delta n}{f} \tag{2.11}$$

Assuming the spatial extent of the fluctuation to be a, we find the relative phase shift

$$\delta\Phi \approx \Delta k a = \frac{e^2}{mc} \frac{\Delta n a}{f} \tag{2.12}$$

under the assumption that the fluctuation has equal extend in both dimensions, as a in the numerator is the extension along the line of sight and a in the denominator the transverse extension. Along the line of sight, the ray is passing through d/a such random fluctuations, amounting to a total phase fluctuation of

$$\Phi \approx \sqrt{\frac{d}{a}} \delta\Phi = \frac{e^2}{mc} \frac{\sqrt{da}\Delta n}{f} \tag{2.13}$$

We may now assume this fluctuation to happen at a screen midway between pulsar and earth that has an extend of a. In this case, the angle of diffraction α can be calculated as

$$\alpha \approx \frac{\Phi}{ka} = \frac{\Phi c}{2\pi a f} = \frac{e^2}{2\pi m} \sqrt{\frac{d}{a}} \frac{\Delta n}{f^2} \tag{2.14}$$

The cone opening angle seen by the observer however is only the half of α as the light rays falling onto the screen have to be bent twice this angle to form a symmetric path to the observer.

We thus conclude that the observer sees the light coming from the pulsar as if it came from a diffusely radiating source of $\Phi_{obs} = \alpha/2$ extent. Since the distance the light travels is longer for rays diffracted under larger angles, an initially sharp pulse will have late arriving components. The time a ray virtually scattered midway at an angle Φ'_{obs} takes to reach earth amounts to

$$t = \frac{d}{c}\sqrt{1+\Phi'^2_{obs}} \overset{\Phi'_{obs}\ll 1}{\approx} \frac{d}{c}\left(1+\frac{\Phi'^2_{obs}}{2}\right) \tag{2.15}$$

Thus the time delay scales with the observing angle to the square. Due to rotation symmetry Φ'_{obs} is distributed as

$$\mathcal{I}(\Phi'_{obs}) \propto 2\pi \exp\left[-\frac{1}{2\Phi^2 obs}\Phi'^2_{obs}\right]\Phi'_{obs}\mathrm{d}\Phi'_{obs} \tag{2.16}$$

It follows from $\mathrm{d}t \propto 2\Phi\mathrm{d}\Phi$ that

$$\mathcal{I}(\Delta t) \propto \mathcal{I}(\Phi'_{obs}(t)) = \pi \exp\Big[-\underbrace{\frac{c}{2d\Phi^2_{obs}}}_{:=1/\tau_s} t\Big] \tag{2.17}$$

Where the time-scale of scattering delay is

$$\tau_s = \frac{2d\Phi^2_{obs}}{c} = \frac{e^4}{8\pi^2 m^2 c}\frac{d^2}{2a}\frac{\delta n}{f^4} \tag{2.18}$$

This result is in accordance with Cordes [2002] and Spangler [1988][1]. As scattering does not occur at one length scale a only, the true delay seen is the total effect of the whole turbulent spectrum. Filling out the details is non-trivial. We for example have not distinguish between weak and strong regimes of turbulence and we assumed the scattering disk to be large w.r.t. the fluctuation scale. This describes the **Diffractive Interstellar Scintillation (DISS)** regime. There also exists a theory for the **Refractive Interstellar Scintillation (RISS)** regime, in which fluctuations larger than the scattering disk parameter are dominant. As length scales are bigger, RISS scintillations happen on larger time-scales than DISS scintillations and slowly blend over into DM variations. For an overview on the effect of fluctuations on pulsar timing, see Stinebring [2013] and references therein.

As the scattering disk's radius is proportional to f^{-2} and the temporal broadening scales as f^{-4} it is immediately evident that higher frequencies will arrive way more unaffected from scintillation effects. The dependency possibly could also be used to investigate the ISM as the volume probed by the scattering delay varies with frequency and thus can probe for density fluctuations on different scales. This additional information however complicates corrections of scintillation effects.

As all ISM effects lead to additional correlations of the signal in time and frequency, analysis of the time-correlation function has been a valuable measure for correlation strength and structure. There has been made promising progress in understanding scintillation effects due to new measurement techniques such as **cyclic spectroscopy** [Demorest, 2011], a technique that exploits the quasi-periodicity of pulsar signals to measure ISM effects directly. The key point of the method is to Fourier-transform the correlations from time domain to frequency domain dependent on phase lag and by doing that to de-modulate ISM effects and pulsar envelope function.

While the interstellar medium inevitably plays an obfuscating role in pulsar observation, it also poses a rich source of additional information that is both not yet fully understood and used.

[1] The derivation loosely follows Lorimer and Kramer [2004] correcting some minor inconsistencies

2.6 From single pulses to nanoshots

Reminding ourself of Figure 2.1, we see that ISM effects, pulse jitter, subpulse drift, moding and nulling have been the effects dominating the temporal modulation of the single pulses besides the pulse envelope. While all of these originate from external factors and not the actual radio emission process, their presence partly masks or smears underlying effects.
As understanding of primary effects and disturbances as well as sensitivity and bandwidth increased, observations unravelled smaller and smaller elementary time-scales for the radiation to occur. Observations like Hankins and Eilek [2007], Hankins et al. [2003], Jessner et al. [2010] show that the elementary process may happen below the nanosecond level. While these extreme outbursts on such time-scales have only been resolved in giant pulses, the noise present and temporal resolution possible may hide that perhaps all emission takes place on that time-scales. Being individually unresolved, nanoshots of radiation still expose a certain spectrum that has been suggested to fit observational constraints [Krzeszowski et al., 2014, Löhmer et al., 2008].
Following just their simplest argument, nano-shots can be described by a potential that decays as $\phi(t) = \phi_0 \exp[-t/\tau_e]$. Superposing those elementary interactions randomly leads to a spectrum given by

$$S(\omega) = \frac{S_0}{1 + \omega^2 \tau_e^2} \tag{2.19}$$

While the whole form fitted is a bit more complex, the characteristic plateau that decays to higher frequencies is preserved. The potential function can also be changed slightly, adding a factor linear to time, without the spectrum being altered dramatically. This renders discerning between different models difficult, but may be possible collecting more accurate data in the future. The reduced χ^2 values also still leave room for model variation while showing clear evidence for a sub-nanosecond decay.
Hankins et al. [2003] argues that of all known emission process candidates only the decay of plasma turbulence [Weatherall, 1998] is capable of explaining an emission time-scale of $\approx$ 2ns. In light of more recent results suggesting an emission in $< 0.4\mathrm{ns}$ [Hankins and Eilek, 2007, Löhmer et al., 2008] even this extreme model is at its limit. We agree with Hankins et al. [2009] that even though there exists one candidate mechanism that reproduces the time-scale, there has to be put more effort in explaining the coupling of the emission process to the magnetosphere and uncover the details of emission.
Hence in light of observations the gap between theory and the system becomes clear as two main questions:

- What is the physics on the scales between nanoshots and microsecond variability?
- How is energy compressed to such small temporal and spatial dimensions before being converted to radio emission?

Facing this spread of theory and observation, we decided to work on statistical behaviour of the receiver equation (see Chapter 4) and algorithms for transparently accessing large datasets (see Chapter 5). This was done to ameliorate analysis of single pulse data and develop a framework in which the envelope and other modulations of single pulses can possibly be removed to reveal a more unbiased view on the micro-structure and its variability. The second question was addressed from the theoretical side, proposing magnetic field energy to play a larger role in generation of radio emission (see Chapter 6).

CHAPTER 3

Problem treatment

In the previous chapters we have described several basic problems of pulsar radio radiation. It becomes evident that, despite many promising attempts, key issues are not solved to satisfaction. Facing half a century worth of papers, theses and books, it is hard to contribute new ideas [Michel, 2004]. To maximise possible usefulness to the whole field instead of the emission process problem alone, we decided to contribute mainly by methodological progress in the intersection of three problems (See also Chapters 4 and 5):

- Statistical properties of the receiver equation
- Noise in pulsar timing
- Handling large datasets in Bayesian context

The treatment of these related problems seeks out to enable analysis of single- and subpulse data to find more signatures of the emission process barely visible and buried in noise. From a more general perspective, we investigated a possible energy source for radio emission (see Chapter 6) by investigating electron-positron annihilation in the context of a large current and magnetic field.
Before the "big picture" will be dissolved into the details of the following chapters, let us lay out a few general remarks about the problem and methods that could be useful to solve it.
For decades, integrating and thus averaging the data received has been proven necessary and useful to generate stable profiles. The procedure is fast, easy to understand, reliable and data-reducing. Hence, why would one, instead of integrating, develop a much more complex method that is based on single pulse data and takes a long time to compute?
We will answer this legitimate question in the following three sections.

3.1 Integration equals losing information

Over the years, many of the observational features (see chapter 2) were discovered by a better temporal or frequency resolution. Before having that resolution, the phenomena were covered by integration of the data. This becomes evident e.g. by looking at the energy fluctuations and integrated profile of the drifting subpulses in Figure 2.4. There is no trace of the subpulse drift left in the summation of the pulses while it cannot be ignored from the intensity plot of single pulses. Thus, integration is a great tool to mitigate random influences such as noise from an observation, but should be handled with care as we possibly lose valuable information when

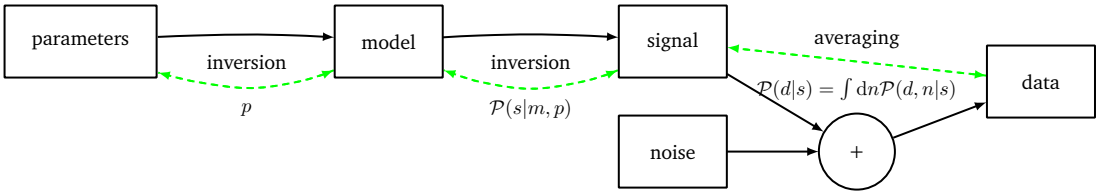


Figure 3.1: **Bayesian vs. classical parameter estimation:** A model with parameters imprints its physical content in a signal that is then received as data with noise. Classical estimators try to invert the relationships to the parameter and propagate the noise as uncertainty (error bars). Bayesian methods do not invert the relationship but rely on determining the remaining plausibility that a certain parameter takes a certain value, given the data. See also section 3.2.

blindly doing that.
By integrating data, we assume a certain way to think about the data to be true. It is merely a psychological reason, that the seemingly smooth and noise-free profile we get by integration gives us the best estimation on pulsar timing, a fact we will highly question in the next chapter. The logical fallacy made in this case is that a property inherent to the phenomenon is exactly related to the integrated data. In this case it is the envelope's phase of every single pulse that is assumed to be bound to the average pulse profile's phase. There can be many reasons why the average profile is not exactly settling to the single pulse's envelope and thus there may be more information contained in the single pulses than in the average pulse.
An exaggerated but accurate example of this fallacy can be seen in the following statement: On average, humans have less than two legs. In fact, the average of human leg count settles between one and two and it becomes clear that we should not mistake the average for something that is straight-forward realized in reality. It is evident in this case, that we lost essential information about human leg count by averaging. But then - what is the meaning of the average?

3.2 Integration may be the answer to a different question

Of course an average quantity can have a meaning in a physical system. Depending on the setting, however, it is likely that we settle on the wrong interpretation of the average. Let us go a step back and reconsider which goal we tried to reach when calculating the average, namely finding out about a physical parameter in our model of pulsar radiation, the envelope.
In the following we will present ideas that loosely follow some arguments of Jaynes and Bretthorst [2003] and introduce standard Bayesian theory. The interested reader may be referred to Jaynes and Bretthorst [2003, and references therein]. For a more general introduction on probability theory in the context of information theory see for example Cover and Thomas [2006].
Observations in general seek to increase knowledge about which physical model is not wrong and which parameter ranges of the model are compatible with the data aggregated. Given a certain model is true, the information about a model manifests itself in a signal from the physics derived that then will be measured. This data include undesirable external influences, noise, that obfuscates undisturbed observation of the signal. Figure 3.1 shows the schematic of the measurement problem. The aim of an observation is to invert the functional relationship between the signal and the data, as well as the parameters and the signal. This problem of signal reconstruction thus has been oversimplified to inversion of a probably only injective relationship. Historically this has lead to development of the theory of error propagation to separate

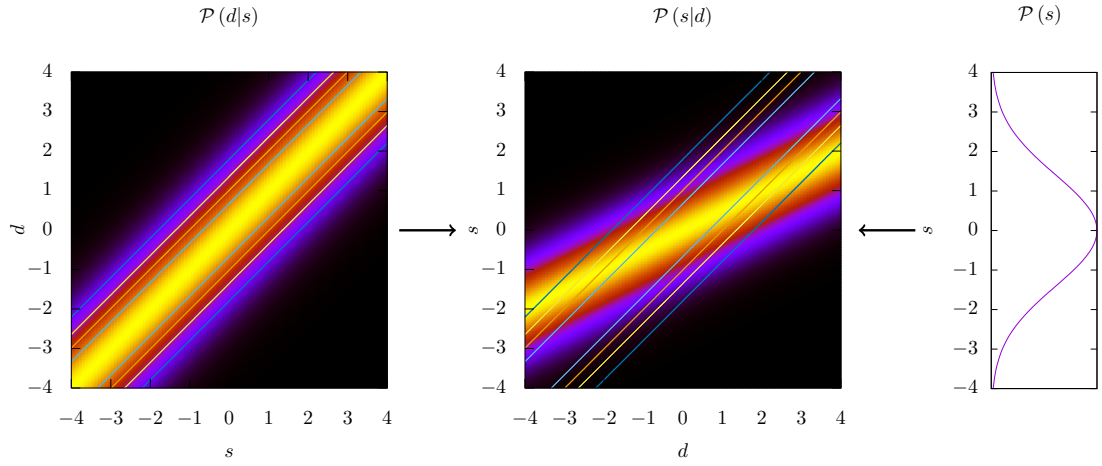


Figure 3.2: **Bayesian knowledge updating:** On the left hand panel, the conditional probability of the data given a certain signal value of s is shown (known as the likelihood), while on the right panel, the prior probability is drawn. The middle panel shows the posterior distribution conditional to the data measured. We overlayed the contour lines of the likelihood to make evident that the prior changes the likelihood's estimate of the signal. See the example in section 3.2.

noise and signal from each other. For additive, uncorrelated noise with zero mean and constant signal, this justifies the average as being an estimator on the signal:

$$d_i = s + n_i, \qquad \mathcal{P}(n_i) \propto \exp\left[-\frac{1}{2\sigma_n^2} n_i^2\right] \tag{3.1}$$

$$s_c := \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{1}{N}\sum_{i=1}^{N}(s + n_i) = s + \frac{1}{N}\sum_{i=1}^{N} n_i \overset{N\to\infty}{=} s + \langle n_i \rangle = s \tag{3.2}$$

where d_i are the measured data values, n_i is the additional noise term,$\mathcal{P}\left(()\, x\right)$ denotes the probability density function of x, σ_n^2 its variance and s the signal from the model assumed constant in this calculation. As is evident, the behaviour of the sum term of the noise determines accuracy of the observation on s. While in this case, the noise can clearly be filtered out, in non-linear systems this must not be true. As most of the radio observations for timing pulsars record d_i^2, the square of the electric field amplitude measured at the receiver, the simple arguments from above have to be modified, as we will derive in chapter 4.
Finding a good estimator for a certain parameter is a highly non-trivial endeavour and, most likely, it will not finally be the average. Bayesian statistics however provide at least a framework in which a certain question of reconstruction can be asked in a mathematical consistent way without trying to invert a given relation. The key is inverting the statistical way from model to data. For a certain **given** model parameter, there usually can be formulated a **conditional probability** on the arising signal. The same bridge exists between the signal and the measured data. Let us state without proof that

$$\mathcal{P}(a|b) \cdot \mathcal{P}(b) = P(a, b) \tag{3.3}$$

The joint probability of a and b is given by the conditional probability of a given b times the probability that b. It holds furthermore

$$\mathcal{P}(a) = \int \mathrm{d}b \mathcal{P}(a,b) \tag{3.4}$$

We can marginalize out b the joint probability of a and b to yield the probability on a[1].
The first equation in 3.1 states that $\mathcal{P}(d_i|s,n_i) = \delta(d_i-(s+n_i))$. The quantity we are interested in is $\mathcal{P}(d_i|s)$, but using equations 3.4 and 3.3 we can reformulate that statement

$$\mathcal{P}(d_i|s) = \int \mathrm{d}n_i \mathcal{P}(d_i,n_i|s) = \int \mathrm{d}n_i \mathcal{P}(d_i|n_i,s)\cdot\mathcal{P}(n_i) \tag{3.5}$$

Where we now may insert the model of 3.1. This yields

$$\begin{aligned}\mathcal{P}(d_i|s) &= \int \mathrm{d}n_i \cdot \delta(d_i-(s+n_i)) \cdot \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{1}{2\sigma_n^2}n_i^2\right] \\ &= \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{1}{2\sigma_n^2}(s-d_i)^2\right]\end{aligned} \tag{3.6}$$

We now have calculated the conditional probability of finding data d_i when the signal s is given. Reconstructing s out of observation is done using Bayes' theorem for inverting conditional probabilities:

$$\underbrace{\mathcal{P}(s|d_i)}_{\text{Posterior}} = \frac{\overbrace{\mathcal{P}(d_i|s)}^{\text{Likelihood}} \cdot \overbrace{\mathcal{P}(s)}^{\text{Prior}}}{\underbrace{P(d_i)}_{\text{Evidence}}} \tag{3.7}$$

The posterior probability distribution describing the new probability on the signal, having measured the data equals the likelihood, how likely it is to find a certain data value multiplicated by the prior probability on the signal to be present and divided through the probability of finding the data, called evidence. As the evidence is independent of the signal, it becomes clear that this is merely a normalization factor. Assuming that the prior knowledge of the signal is a flat distribution $\mathcal{P}(s) = \mathrm{const}$, and inserting equation 3.6 into the formula we find

$$\mathcal{P}(s|d_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{1}{2\sigma_n^2}(s-d_i)^2\right] \tag{3.8}$$

Let us give a pictorial example(see Figure3.2) for the updating rule and, for the moment, assume a Gaussian prior centred around zero. Let us further assume that the noise and the prior have the same variance of unity. We have graphically represented the update of the knowledge (the posterior, middle panel) compared with prior knowledge (on the right panel) and the likelihood of measuring a certain data d given a signal s being present (left panel). We see that the final reconstruction takes into account both prior knowledge and the new findings represented by data d and will give a reconstructed signal $s|d$ given a certain datum. The prior has been chosen to give a large influence in order to be clearly visible in the figure. In practise, one would prefer to choose a more ignorant prior when there is no previous knowledge on the signal.

[1] This is a further example of a process where information is lost by integration, as correlations between a and b will be dropped.

Returning to the flat prior and assuming we have consecutively measured various values, we may apply Bayes' theorem repeatedly for the new data[2] to find the posterior given all data $\{d_0, ..., d_{N-1}\}$:

$$\begin{aligned}\mathcal{P}(s|d_0, ..., d_{N-1}) &\propto \prod_{i=0}^{N-1} \exp\left[-\frac{1}{2\sigma_n^2}\left(s - d_i\right)^2\right] \\ &\propto \exp\left[-\frac{1}{2\sigma_n^2/N}(s - \frac{1}{N}\sum_{i=0}^{N-1} d_i)^2\right] \end{aligned} \tag{3.9}$$

which is a Gaussian centred at the average data value with a variance of σ_n^2/N. In the following we will write d without an index when we mean the whole data set. Thus, the simple example described above leads to the same estimation like the classical procedure. This however is the case as the

- problem is linear
- prior knowledge was assumed to have a flat probability distribution
- noise has Gaussian statistics with zero mean and is independent
- signal was only single-valued

Of course there might be other deviations that deem "averaging" or a more complex form of it non-optimal. If one of the above points is violated, there may arise a more complex estimator, for example a more differentiated weighting of the data.
Bayesian statistics do not only give a framework for the accurate reconstruction of the data, but also make some more subtle nuances of the involved statistics evident. It has been a common critique that by specifying a prior distribution, Bayes' theorem would give way to introduce arbitrary influences to the result. It is true that a prior distribution can introduce prejudice into the problem. If we e.g. a priori exclude a certain signal region A to be not possible, setting a subset of $\mathcal{P}(s \in A) = 0$, a reconstruction of a signal that nevertheless falls into this region is excluded by the theorem. Thus, one has to choose a prior wisely to reflect the most ignorant state about the signal if starting from a point of no knowledge at all. However, in this framework we have to choose a prior and name it. In classical reconstructions formulae, the prior has been chosen by construction of the estimation formula. We see this by the following argument. Let us transform the signal's prior with a function $s' = f(s)$. As $\mathcal{P}(s)$ is a distribution, it transforms differently than a pure value:

$$\mathcal{P}\left(s\right)\mathrm{d}s = \mathcal{P}\left(s'\right)\left(\frac{\partial f}{\partial s}\right)^{-1}\mathrm{d}s' = \mathcal{P}\left(s'\right)\frac{\partial f^{-1}}{\partial s'}\mathrm{d}s' \tag{3.10}$$

We conclude that an initially flat prior distribution can take arbitrary forms by transforming the signal's coordinate since $\mathcal{P}\left(s'\right) = \frac{\partial f}{\partial s}\mathcal{P}\left(s\right)$. But this means, when transforming the signal in (3.9), the prior assumed on the signal is not a uniform distribution any more since equation (3.10) holds.
The straight-forward classical estimator for s_c' is $f(s_c)$ with s_c being the classical estimate of s given by equation (3.2) and it is tempting to conclude that

$$\left.\frac{\partial\mathcal{P}\left(\left(s|d\right)\right)}{\partial s}\right|_{s_c} = 0 \Rightarrow \left.\frac{\mathcal{P}\left(\left(f^{-1}(s')|d\right)\right)}{\partial s'}\right|_{s_c'} = \underbrace{\frac{\mathcal{P}\left(\left(f^{-1}(s')|d\right)\right)}{\partial f^{-1}}}_{=0}\left.\frac{\partial f^{-1}}{\partial s'}\right|_{s_c'} \tag{3.11}$$

[2]We implicitly assume statistical and logical independence of the single measurement here

And in fact, this is the case. However, the right hand side is not the correct posterior distribution. The correct posterior distribution is given by

$$\mathcal{P}\left(s'|d\right)\mathrm{d}s' = \underbrace{\mathcal{P}\left(f^{-1}(s')|d\right)\frac{\partial f^{-1}}{\partial s'}}_{:=g(s')}\mathrm{d}s' \tag{3.12}$$

Notice that the $\mathcal{P}$ on the left hand side takes not the same functional form than that on the right hand side. The argument the function on the left hand side takes is an element of the space of s'. The argument the integrand of the distribution on the right hand side takes is an element of the original signal space. This is a major caveat when dealing with the typical physicists' notation of probabilities. It is hiding the fact that these functions differ as they are integrands in two different spaces. Deriving the new integrand $g(s')$ w.r.t. s' at s'_c will only yield a local extremum too if[3]

$$\left.\frac{\partial^2 f^{-1}}{\partial s'^2}\right|_{s'_c} = 0 \tag{3.13}$$

as we deduce by carrying out the product rule and noticing that the first term will be zero by equation (3.11). As the point s'_c is effectively determined by the data, equation (3.13) should be fulfilled at arbitrary points in the space of s'. Assuming smoothness of the transformation function f and their first and second derivative, this means we will restrict ourselves to linear transformations only.

It is a well known statement that inverting non-linear relations with classical parameter estimation will be biased (which is effectively the case if the maximum and thus the whole distribution is shifted). We want to make a more subtle point here. There is a way of ensuring that $s'_c = f(s_c)$ is the 'correct' classical estimator. If we took not a uniform prior but the prior

$$\mathcal{P}'(s) \propto \frac{\partial f(s)}{\partial s} \tag{3.14}$$

the transformation had put the functional form of the integrand of $\mathcal{P}\left(s'|d\right)$ into the form given by equation (3.11) as the derivative of f^{-1} and f would cancel out. Thus, the estimator for the transformed signal can then possibly again be justified by statistics under assumption of a prior differing from the uniform distribution.

Thus, given classical and Bayesian estimation agree in a most likely value of the signal, we have shown that under non-linear transformation, agreement may only be reached by assuming a prior being present in the classical case. This means that classical methods may have implicit prior assumptions on their signal that are not evident by the actual form of estimation. We thus conclude that, against commonly raised concerns, the seemingly arbitrary choice of a prior in Bayesian statistics is also inherent in classical parameter estimation theory. However in the latter case, the prior assumptions taken are not explicitly specified and thus, may be arbitrary too, or, even worse: clearly wrong. We deem it a much better way of doing science to speak about the prior assumptions made, than simply swiping them under the rug of apparently simple math.

[3] Excluding exotic cases where $\mathcal{P}\left(f^{-1}(s')|d\right)\big|_{s'_c} = 0$ for example

3.3 Ignoring a problem does not help solving it

Having understood that the result of parameter estimation relies on

- well specified prior knowledge (or at least a wisely chosen non-informative prior)
- a well-fitting physical model of the signals in the data
- exact knowledge of sources of noise
- the data taken

we realize that taking the data does by far not suffice to interpret it correctly. The exact path from a solid model to the signal expected to the data gathered and back has to be calculated in order to infer a parameter's probability distribution. We agree that there are "short-cuts" like the law of large numbers and the central limit theorem that finally may equalize the findings of a more complex statistical model with "taking the average" of some form that leads to the desired quantity in a simple way. However, there may be subtle influences by non-linearities or simply wrong model assumptions that may flaw secondary parameters' estimations such as ToAs in a way not noticeable in the final stages of the analysis.
Chapter 2 named various effects that modify the expected signal and the observation of this signal drastically. In light of knowing this, one should at least check and make sure, that there are no subtle effects left in the parameter's estimation. As we will see in the following chapter, there is indeed noise present in pulsar timing data which origin is not completely understood.
If one could model the emission process and the propagation effects in more detail, we argue that at least some part of this noise may be understood and be part of a signal instead of a simple noise term.
The foremost problem with Bayesian methods however, is the computational feasibility on large datasets. While for simple models analytical solutions can be found, more complex models will rely on approximative calculations or resource-intensive numerical evaluation.
There have been developed frameworks like Information Field Theory [for an introduction see e.g. Enßlin, 2013, Lemm, 1999, and references therein] that can simplify calculations while still taking the most significant corrections into account. These theories may provide operators that amount to calculating and weighting simple averages or higher statistical moments, to ameliorate numerical evaluation.
As the reduction may depend on the actual parameters' values, preprocessing the dataset by "blindly" integrating will most likely lose data. If we do not want to lose information on the way, it must be possible to reprocess the data as a whole, in an automated, transparent and reproducible way. Finally we will be facing huge datasets that are to be reduced dynamically, given the parameter's probability density function.
Even at the stage of first steps towards Bayesian analysis of pulsar data, we face a numerical limit when memory shortage on the local system, not computational power is the bottleneck due to the massive data sets. To pass this limit we decided to write Rambrain (see chapter 5), a library that abstracts away the details of data storage and that cares for swapping out the data in a memory-efficient way. The advantage of having such a library is simple: When writing complex algorithms, the burden of memory management now is handled by this library. It is much easier to write algorithms accessing large datasets when the algorithm can be put in simple statements that do not have to deal with physical limitations of the machine the code will be run on. Abstractions of this form may prove essential to pave the way to handle big datasets in science.

Having such a long history, the field of pulsar and radio emission science has produced numerous models and a plethora of observational data to draw from. It is very likely that the solution to the problem of radio emission lies before our very eyes, but we lack the methods to extract it from both data and theory. It is this ignorance of actually known facts that we seek out to reduce in this thesis.

CHAPTER 4

Paper : A Bayesian method for pulsar template generation

M. Imgrund, D.J. Champion, M. Kramer, H. Lesch
This chapter has been published in slightly modified form in Monthly Notes of the Royal Astronomical Society, June 1, 2015 449 (4): 4162-4183

Abstract: Extracting Times of Arrival (ToAs) from pulsar radio signals depends on the knowledge of the pulsar's pulse profile and the method being used to compare such profiles. We contrast classical generation mechanism of averaging received intensity to generate these so called templates with a new approach based on Bayesian inference. Contrary to relying on purely integrated data we explore the statistical properties of the receiver equation. The novel "statistical templates" enrich the classically static information about the profile shape by assigning statistically reasonable uncertainties. We explore the benefit of this statistical information on both simulated and real measurement data from PSR B1133+16, PSR B0329+54 and PSR J1713+0747. After thoroughly analysing and testing the algorithm we apply the novel approach to classical problems such as Nulling detection, Moding and ToA generation.
We show that statements of classical analysis like "pulsar showed nulling" or "pulsar radiates in mode A" are enhanced with probabilities that can be used as weights in subsequent analysis. Implementing ToA generation we find this algorithm to yield comparable timing errors to the classical method but giving a more accurate estimate of the remaining uncertainty in the ToAs. Consequently we propose using such a method when being dominated by fluctuations, for example on bright observations, pulsars showing moding or observations with the increasing sensitivity of future telescopes like the square kilometer array (SKA).

4.1 Introduction

Pulsars are used as clocks of nanosecond precision at galactic distances, particularly in areas of fundamental physics such as general relativity [Cordes et al., 2004, Kramer et al., 2006b]. These measurements make use of the exceptional rotational stability of the pulsar observed as electromagnetic pulses with radio telescopes. Most of these analyses rely on determining the arrival times of the pulses out of tens of thousands of integrated single pulses.
The individual (or single-) pulses observed from a pulsar are extremely variable, not only in flux but also in the shape of the pulse. It is only when 10s of 1000s of pulses are averaged in an integrated pulse profile that it becomes stable [Helfand et al., 1975, Liu et al., 2012]. Some pulsars exhibit temporal variations in this pulse profile; showing no emission for a period of

time - nulling (see Sec. 4.3.5), or switching between one or more additional stable profiles (see Sec. 4.3.7). For the majority of pulsars however, the pulse profile remains stable, in some cases over at least years.
There exist ready to use tool sets like PSRCHIVE[Hotan et al., 2004] and TEMPO2 [Edwards et al., 2006, Hobbs et al., 2006] to analyse raw observation data. They are used to assign times of arrival to integrated observations and constrain the parameters of the physical model under consideration. The remaining differences between the model's prediction of an arrival time and the actual time measured are called timing residuals. Timing residuals can be as low as a few tens or hundreds nanoseconds, while the periods of observational data span several years. This amounts to tracking the rotational phase of the pulsar at a certain observational epoch to an accuracy of 10^{-4} and well below. A single pulse however is a highly variable object that apparently cannot be tracked down to that accuracy (see e.g. fig. 4.11). It is immediately evident that this very exact phase information is imprinted in, and has to be recovered out of, tens of thousands of very individual pulses.
The classical way of generating times of arrival (ToAs) is based on techniques comparing the shift of an epoch under consideration to a template. The template is often an integrated pulse profile of the entire data set for that pulsar at the appropriate observing frequency and bandwidth. Usually it is then fit with an analytic function to produce a noiseless template. These templates are occasionally updated and the ToAs reproduced.
Observations integrated over shorter times are then tested for a phase shift w.r.t. the template. The user of PSRCHIVE e.g. has the choice among five algorithms all comparing the data of an observation to the reference. These methods state a relative shift and an error estimate on it. They yield sub-bin accuracy by e.g. discrete cross-correlation and interpolation Hotan et al. [2005] or examining the phase gradient in the fourier space representation as in Taylor [1992]. The phase shift is then converted into a ToA using the timestamp and folding period of the observation. In a next step, the actual pulsar timing is carried out.
Pulsar timing is the process by which a rotational model of the pulsar is produced and the value of its rotation phase is mapped to the ToAs at the observatory of every pulse. This model is fit to the ToAs from observations to produce a phase coherent solution that accounts for every rotation of the pulsar in the range of observations. It is the coherence of this solution that gives pulsar timing its extraordinary precision.
All data input to this procedure, the template and the observed profiles, relies on integrated and thus averaged data before correlating. Single pulses cannot be compared directly for they are often not bright enough to be distinguished from noise introduced on the signal's way to the telescope and by the antenna-receiver-amplifier system. Furthermore their distinct and individual shape make it difficult to correlate them to the average profile as simply matching the shape of the pulse with the template is not possible. Thus timing individual pulses by comparing them to a template lacks the necessary precision.
The classical way of integrating pulses to form a stable profile justifies itself to mitigate the outlined problems. However in actual observations, the reduced χ^2 value, a measure comparing how well a given model fits the data taking into consideration the reported uncertainties in the data, is much larger than unity [see e.g. Bailes, 2010]. As the global timing models of pulsars stand on very solid grounds, the deviance from unity points to an underestimation of the remaining error in the data by the classical procedure and algorithms. This additional error is believed to have its root in additional noise sources like e.g. scattering by the interstellar medium. A comprehensive overview of the noise budgets can be found e.g. in Cordes and Shannon [2010].
Mitigating errors after determining ToAs has been intensively investigated by assessing and fitting the possibly introduced effects on the ToAs [Coles et al., 2011, Cordes and Shannon,

2010]. As Bayesian methods already have been successfully used for correlating ToA-data from different pulsars, extending these to mitigate the timing errors by supposing a Bayesian timing model on the ToAs as described by Lee et al. [2014], Lentati et al. [2014], Vigeland and Vallisneri [2014] and others suggests itself.
The shape of single pulses is affected mainly by the short term noise contributions coming from current observational conditions. One can aptly paraphrase these imponderables as interstellar weather and pulse jitter caused by the fluctuating state of the magnetosphere or radiation process. Both influences affect mainly the single pulse statistics or influence the profile over short timescales, but can systematically affect timing and even dominate timing precision [Osłowski et al., 2011].
Increasing the gain of the observations with future systems like the Square Kilometre Array (SKA) aggravates this problem. These fluctuations demand for longer integration times if no solution is found to incorporate them in the determination of ToAs rather than trying to average them out [Liu et al., 2011].
Finding a generic way of dealing with single pulse individuality can ameliorate the situation and fill the statistical gap between single pulses and an integrated profile without the need to take further modelling assumptions[1].
In this paper we will show a way to address the problem of generating pulsar ToAs with single pulse statistics and the benefits of having a more accurate statistical representation for single pulses' behaviour. By using more information than the classical method (which integrates this information) we hope to improve the accuracy of pulsar ToA generation and to allow ToAs from shorter observations to be unbiased by pulse jitter.
We introduce a model for statistics of the single pulses and substitute the classical template by a "statistical template" representing the single pulse statistics. We will then work out the steps of statistical template generation and phase shift detection replacing the classical counterparts. Hereto we will derive a measurement model for squared and amplified receiver voltage in Sec. 4.2. We argue how to handle the signals from pulsars best, splitting them in a radiation process and a phase coupled template part. Using the Bayesian formalism, we deduce the posterior distribution for both the amplitude and phase model introduced and thus show a way to infere a statistical template from input data. Though focusing on the important application of generating ToAs, we will also demonstrate the benefit of using statistical templates on other applications such as determining nulling and analysing a moding pulsar. Sec. 4.3 will evaluate the proposed algorithm and methods using both simulated and real pulsar data. We show how a statistical template can enhance the view on the pulsars profile from a statistical perspective. Comparing the generation of ToAs to the classical results we will highlight improvements in the χ^2 value of the fit. We summarize our findings in Sec. 4.4.

4.2 Modelling pulsar measurements

4.2.1 Terminology of the used statistical model for single pulses

In order to describe single pulse statistics we use the notion of a fluctuating envelope besides to the classical set of terms of single pulses and integrated profiles.
Instead of modeling an average profile, we assume a single pulse to be the result of multiplying a stationary Gaussian random noise process, which amounts to the physical process that generates the radiation in the pulsar magnetosphere, with a possibly unstable envelope (see

[1] Incorporating assumptions on e.g. scattering of the ISM or pulse jitter behaviour however may further improve the results presented

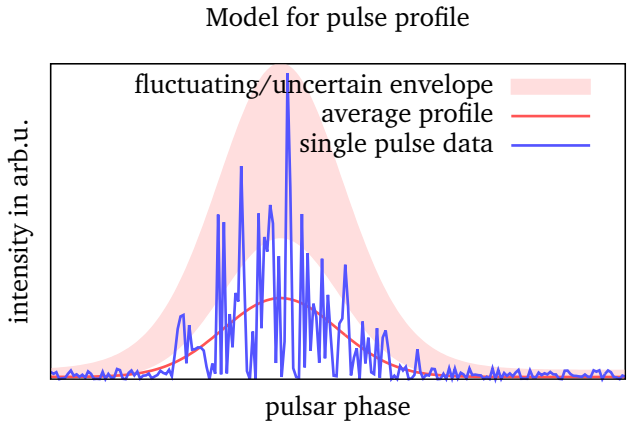


Figure 4.1: Scheme for single pulse statistics as described in Sec. 4.2.1

fig.4.1). This envelope models the periodic influence of the magnetosphere. The uncertainty in the envelope may reduce with observation time when the envelope is not fluctuating. However if the magnetosphere conditions for some region mapping to the pulse phase are fluctuating, the envelope will stay fluctuating in that region.
Only in the limit of no uncertainty the envelope would equal the (scaled) average of the pulses and amount to the same thing as a classical integrated profile. As averaging single pulses generates a pulse profile, we emphasize that there are many possible fluctuating envelopes which are integrated to the same profile.
The empiric or analytic template that is used for generating times of arrival will consequently be substituted by a Bayesian statistical template describing the fluctuating envelope.

4.2.2 Amplitude model

Every Bayesian analysis is based on formulating a signal and noise model to describe the statistical imprint of a certain signal in the data. Thus, before inserting the outlined statistical model of the signal , we have to understand the influence of an arbitrary signal. Therefore we will introduce the basic measurement model, insert our signal model and then derive the formulae for signal reconstruction using Bayes' theorem.
Typical pulsar measurements at radio frequencies give the intensity of received radiation by suitable sampling of $\mathbf{d}$, the quadrature of the received voltage. The problem is to infer a unknown signal $\mathbf{s}$ (which in our case amounts to the noise-free series of single pulses as emitted from the pulsar) from the noisy data $\mathbf{d}$. The noise - leaving aside radio interference - is dominated by thermal fluctuations of both receiver and antenna and is accurately described by white noise of a usually known variance σ_{n}^2 entering before quadrature of the signal. Thus, the baseline corrected data $\mathbf{d}$ measured, given a signal $\mathbf{s}$ and a random noise $\mathbf{n}$ amounts to

$$\mathbf{d} = (\mathbf{n} + \mathbf{s})^2 \tag{4.1}$$

$$\mathbf{d} - \mathbf{n}^2 = \mathbf{s}^2 + 2\mathbf{s}\mathbf{n} \tag{4.2}$$

where

$$\mathcal{P}(\mathbf{d}|\mathbf{s},\mathbf{n}) = \delta(\mathbf{d}-(\mathbf{n}+\mathbf{s})^2) \tag{4.3}$$
$$\mathcal{P}(\mathbf{n}) = \mathcal{G}_{\mathbf{n}}(0,\mathbf{N}) \tag{4.4}$$

where $\mathcal{P}(x|y)$ is the short notation for the probability density of $X = x$ given $Y = y$, where X, Y would be the corresponding random variables. We use boldface lower case letters to emphasize that data, signal and noise consist of many single values $\mathbf{d} = \{d_1, d_2, \cdots d_t, \cdots\}$ for which the equations stated are valid component-wise. Thus operations like taking a root of a vector or multiplying two vectors are carried out only on the components of the vector. For the scalar product of vectors $\mathbf{a}$ and $\mathbf{b}$ we use the notations $\mathbf{a}^\dagger\mathbf{b}$ and $\sqrt{\mathbf{a}^\dagger\mathbf{a}} = ||\mathbf{a}||$. We abbreviate $\det(\cdot)$ as $|\cdot|$ and define $\delta(\mathbf{a}) := \prod_k \delta(a_k)$. Putting a hat over a vector $\hat{\mathbf{v}} = \mathrm{diagmat}(\mathbf{v})$ denotes constructing a diagonal matrix with a diagonal with the elements of$\mathbf{v}$. When dealing with sets, $\mathbf{A} \setminus \mathbf{g}$ denotes the set $\mathbf{A}$ without the elements $\mathbf{g}$.
$\mathcal{G}_{\mathbf{x}}(\mathbf{m},\mathbf{V})$ denotes that $\mathbf{x}$ is distributed as the multivariate Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{V}$. It is defined as

$$\mathcal{G}_{\mathbf{x}}(\mathbf{m},\mathbf{V}) := \frac{1}{\sqrt{2\pi|\mathbf{V}|}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^\dagger\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

Integrating over a vector $\mathbf{a}$ is denoted by $\mathcal{P}(\mathbf{b}) = \iint \mathrm{D}\mathbf{a}\mathcal{P}(\mathbf{a},\mathbf{b}) = \iint \mathrm{d}a_1 \cdots \mathrm{d}a_k \mathcal{P}(\mathbf{a},\mathbf{b})$.
The classical way of determining the template intensity amounts to averaging eq.(4.2) over several pulsar periods, noticing that the term mixing signal and noise vanishes leaving us with the desired quantity of $\langle\mathbf{s}^2\rangle$ when subtracting σ_{n} from the averaged data. To uses Bayes' theorem, we will instead calculate the exact probability distribution to measure $\mathbf{d}$ at a certain time, given a certain signal value $\mathbf{s}$ at that time:

$$\mathcal{P}(\mathbf{d}|\mathbf{s}) = \iint \mathrm{D}\mathbf{n}\mathcal{P}(\mathbf{n},\mathbf{d}|\mathbf{s}) = \iint \mathrm{D}\mathbf{n}\mathcal{P}(\mathbf{d}|\mathbf{n},\cdots)\mathcal{P}(\mathbf{n}) \tag{4.5}$$
$$= \iint \mathrm{D}\mathbf{n}\delta(\mathbf{d}-(\mathbf{n}+\mathbf{s})^2)\frac{1}{\sqrt{|2\pi\mathbf{N}|}}\exp\left[-\frac{1}{2}\frac{||\mathbf{n}||^2}{\sigma_{\mathrm{n}}^2}\right] \tag{4.6}$$

We substitute $\widetilde{\mathbf{n}} = (\mathbf{n}+\mathbf{s})^2, \mathbf{n} = \pm\sqrt{\widetilde{\mathbf{n}}} - \mathbf{s}, \mathrm{d}\mathbf{n} = \frac{\mathrm{d}\widetilde{\mathbf{n}}}{2\sqrt{\widetilde{\mathbf{n}}}}$ and notice that $\sqrt{|\mathbf{N}|} = \sigma_n^{N_{\mathrm{tot}}}$, where N_{tot} is the total number of samples in all bins.

$$= \int_0^\infty \mathrm{D}\frac{\widetilde{\mathbf{n}}}{\sqrt{\widetilde{\mathbf{n}}}} \quad \delta(\mathbf{d}-\widetilde{\mathbf{n}})\frac{1}{|\sqrt{2\pi}\sigma_{\mathrm{n}}^{N_{\mathrm{tot}}}|}\exp\left[-\frac{1}{2}\frac{||\sqrt{\widetilde{\mathbf{n}}}\pm\mathbf{s}||^2}{\sigma_{\mathbf{n}}^2}\right] \tag{4.7}$$
$$= \frac{1}{\sqrt{|2\pi\mathbf{d}|}\sigma_{\mathrm{n}}^{N_{\mathrm{tot}}}}\exp\left[-\frac{1}{2}\frac{||\sqrt{\mathbf{d}}\pm\mathbf{s}||^2}{\sigma_{\mathrm{n}}^2}\right] \tag{4.8}$$

where we now have to assign a signal model for s. We assume

$$\mathbf{s} = \exp[\mathbf{f}]\cdot\mathbf{g}\sigma_{\mathrm{n}} \tag{4.9}$$

Where we have decided for the signal to be measured in units of noise temperature of the receiver-antenna system. The $\pm$-operator means that the whole density would be a sum of the densities with the respective sign. This operator vanishes later on, when integrating over g and using $\mathcal{P}(g) = \mathcal{P}(-g)$. $\mathbf{f}$ and $\mathbf{g}$ are Gaussian random variables with also to be determined

covariances $\mathbf{F}$ and $\mathbf{G}$ which describe the statistical properties of the envelope respectively the radiation process. The exponential form for the envelope was chosen since on one hand it is able to grasp the occurring fluctuations to high radiation densities of single pulses and on the other we may easily assume a scale-free prior for the amplitude, as we will see below.
We assume the radiation process to have stationary statistics given by

$$\mathbf{G} = \mathbf{G}(t - t')$$

and the part that is generating the observed envelope to be periodic in time. Even though the covariance matrix of $\exp[f]$ alone has non-diagonal components, the product's covariance matrix $\langle g(t)\exp[f](t)g'(t)\exp[f]'(t)\rangle$ has only diagonal components if we marginalize over g, because of the stationarity of g as assumed in (4.2.2). Thus we describe the covariance matrix solely by it's Fourier-coefficients at $\omega_k = k \cdot \frac{2\pi}{T}$. Since the expectation of the signal's covariance matrix for a certain g integrated over a specific pulse certainly does not vanish, the non-diagonal components of the signal's matrix carry information about g which we will discard in this paper. T denotes the real pulsar period, which we will discriminate from the assumed period τ later on.

Joint Probability and Information Hamiltonian

For a moment, let us assume the covariance matrices $\mathbf{F}$ and $\mathbf{G}$ are known. Then we may write for the joint probability

$$P(\underbrace{\mathbf{f}, \mathbf{g}, \mathbf{s}, \tau}_{=:A\cup\{\mathbf{s}\}} \,|\, \underbrace{\mathbf{F}, \mathbf{G}}_{=:B}) = \delta(\exp[\mathbf{f}]\,\mathbf{g}\sigma_\mathrm{n} - \mathbf{s})\mathcal{G}_\mathbf{f}(0, \mathbf{F})\mathcal{G}_\mathbf{g}(0, \mathbf{G})\mathcal{P}(\tau)$$

and we define the parameter sets A and B where $\mathcal{P}(\tau)$ is the probability density over the assumed period τ. In order to define the Information Hamiltonian $H_B[A, \mathbf{d}] := -\log P(A, \mathbf{d}|B)$ we notice that $\mathcal{P}(A, \mathbf{d}|B) = \iint \mathrm{D}\mathbf{s}\mathcal{P}(\mathbf{d}|\mathbf{s})P(A \cup \{\mathbf{s}\}|B)$ and use the measurement model given by (4.8):

$$\begin{aligned} -\log P(A, \mathbf{d}|B) = H_B[A] = \Bigg[&(\sqrt{\mathbf{d}} - \exp[\mathbf{f}]\cdot\mathbf{g}\sigma_\mathrm{n})^\dagger \frac{1}{\sigma_\mathrm{n}^2}(\sqrt{\mathbf{d}} - \exp[\mathbf{f}]\cdot\mathbf{g}\sigma_\mathrm{n}) + \\ &\mathbf{f}^\dagger\mathbf{F}^{-1}\mathbf{f} + \mathbf{g}^\dagger\mathbf{G}^{-1}\mathbf{g}\Bigg] - \log\mathcal{P}(\tau) + \frac{1}{2}\log(|2\pi\sigma_\mathrm{n}^2|^{N_\mathrm{tot}}|2\pi\mathbf{F}||2\pi\mathbf{G}||\mathbf{d}|) \end{aligned} \tag{4.10}$$

Since we are interested in the pulsar's statistical template given by the envelope characteristics we marginalize over the radiation process g by integrating it out. This calculation is outlined in Sec. 4.6.1. We are left with

$$\begin{aligned} H_B[A] = \frac{1}{2}\Bigg[&\sqrt{\mathbf{d}}^\dagger \frac{1}{\sigma_\mathrm{n}^2}[1 + \widehat{\exp[\mathbf{f}]}\mathbf{G}\widehat{\exp[\mathbf{f}]}]^{-1}\sqrt{\mathbf{d}} + \mathbf{f}^\dagger\mathbf{F}^{-1}\mathbf{f} + \\ &+ \log(|2\pi\mathbf{D}_\mathbf{f}^{-1}||2\pi\sigma_\mathrm{n}^2|^{N_\mathrm{tot}}|2\pi\mathbf{F}||2\pi\mathbf{G}||\mathbf{d}|)\Bigg] - \log\mathcal{P}(\tau) \end{aligned} \tag{4.11}$$

where the operator $[1 + \widehat{\exp[\mathbf{f}]}\mathbf{G}\widehat{\exp[\mathbf{f}]}]^{-1}$ together with the noise variance is a variation of the operator leading to the well-known case of the Wiener filter [Wiener, 1949]. This similarity arises due to the integration over the Gaussian random field $\mathbf{g}$ added upon the signal field $\mathbf{s}$. However, the normalisation w.r.t. the data differs and we further distinguish between the periodic part described by $(\mathbf{f}, \mathbf{F})$ and the radiation part now showing up only as $\mathbf{G}$. The functioning

of the operator can be understood by considering its interplay with given data. Maximizing probability equals minimizing the Hamiltonian. Thus high data values will be compensated by large mean values of f. Conversely, for too high values of f the exact value of the data becomes irrelevant, leading to a sub-optimal solution independent of the data. The optimal solution resembles the actual probability distribution the data would be drawn from. As a function of $\mathbf{d}$, the joint probability is a special case of a gamma distribution $\propto \exp\left[-||\frac{\mathbf{d}}{\mathbf{m}}||\right] ||d||^{-.5}$ where $\mathbf{m}$ is the mean value of the distribution. In this light the operator resembles the classical formula $\langle \mathbf{d} \rangle = \mathbf{m} = \sigma_{\mathrm{n}}^2 + \sigma_s^2$ since $\mathbf{m} \triangleq \sigma_{\mathrm{n}}^2[1 + \frac{\mathbf{s}^\dagger \mathbf{s}}{\sigma_{\mathrm{n}}}]$ for our case.
$\mathbf{G}$ becomes diagonal in Fourier space since $\mathbf{g}$ is assumed to be stationary. As $\exp[\mathbf{f}]$ is a periodic signal, its Fourier space is limited to discrete frequencies that are a multiple of the assumed periodicity. We may absorb the values of G_ω, where $\omega_k = k \cdot \frac{2\pi}{\tau}$ into $\mathbf{F}$. In the following we will discard information about G from single pulses. Since our main interest lies on the statistics of the envelope, we may take this loss.
Under these assumptions, and neglecting non-diagonal terms arising when Fourier transforming f (see Sec. 4.7 and 4.2.2), the $G_\omega =: \sigma_{g_k}^2$ may be absorbed into an effective mean $\mathbf{f}$ as is evident in eq. (4.11):

$$\exp[2f_k] \cdot \sigma_{g_k}^2 = \exp\left[2(f_k + \overline{f}_k)\right]$$

where $\overline{f}_k = \log \sigma_{g_k}$. We redefine $\mathbf{f}$ in (4.11) in that way and yield

$$H_B[A \setminus \mathbf{g}] = \frac{1}{2}\left[\left\| \frac{\mathbf{d}}{\sigma_{\mathrm{n}}^2} \frac{1}{1+\exp[2\mathbf{f}]} \right\| + (\mathbf{f}-\overline{\mathbf{f}})^\dagger \mathbf{F}^{-1}(\mathbf{f}-\overline{\mathbf{f}}) + \right.$$
$$\left. + \log(|2\pi \mathbf{D}_{\mathbf{f}}^{-1}||2\pi\sigma_{\mathrm{n}}^2|^{N_{\mathrm{tot}}}|2\pi\mathbf{F}||\mathbf{d}|)\right] - \log \mathcal{P}(\tau) \tag{4.12}$$

where $\mathbf{G}$ is now the unity matrix and we assume $\mathbf{F}^{-1}$ to be diagonal with components $\frac{1}{\sigma_{f_k}^2}$. We arrive at our model for a single pulse's amplitudes in Fourier space, assuming all timing parameters to be given:

$$H_{\overline{\mathbf{f}},\sigma_{\mathbf{f}}^2}[\mathbf{f},\mathbf{d}] = \frac{1}{2}\left[\left\| \frac{\mathbf{d}}{\sigma_{\mathrm{n}}^2} \frac{1}{1+\exp[2\mathbf{f}]} \right\| + (\mathbf{f}-\overline{\mathbf{f}})^\dagger \frac{1}{\sigma_{\mathbf{f}}^2}(\mathbf{f}-\overline{\mathbf{f}}) + \log C(\mathbf{F},\sigma_{\mathrm{n}},\mathbf{d})\right] \tag{4.13}$$

This is the negative logarithm of the likelihood function. Bayes' law gives us the posterior

$$P(\mathbf{f}|\mathbf{d}) = \frac{\exp\left[-H_{\overline{\mathbf{f}},\sigma_{\mathbf{f}}^2}[\mathbf{f},\mathbf{d}]\right]}{\iint \mathrm{D}\mathbf{f} \exp\left[-H_{\overline{\mathbf{f}},\sigma_{\mathbf{f}}^2}[\mathbf{f},\mathbf{d}]\right]} \tag{4.14}$$

when we assume a flat prior on $\overline{\mathbf{f}}$ which effects in a scale-free prior on the assumed amplitude, since $\mathbf{s} \propto \exp[\mathbf{f}]$ is exponentiated.

The receiver equation's pdf

Assuming G to be a diagonal of ones (meaning vanishing correlations) for this argument, the pdf of the receiver equation of each bin decouples and for a specific bin and different pulses K takes the form of

$$\mathcal{P}\left(f|\vec{d}\right) \propto \prod_K \exp\left[-\frac{1}{2}\frac{d_K}{1+f^2}\right] \sqrt{\frac{1}{(1+f^2)2\pi}} / \sqrt{d_K}$$

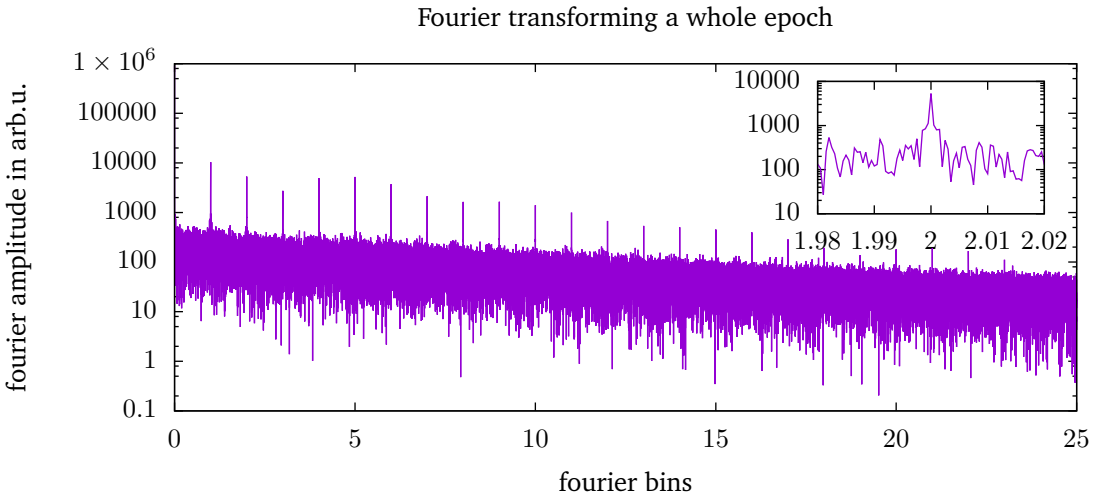


Figure 4.2: Fourier transform of a whole epoch of single pulses of pulsar B0329+54. The x axis has been scaled to the pulsar's signal harmonics. Inset: detail of one spike.

which simply states that the variance of $\sqrt{d_K}$ can be explained as sum of the variance of the noise (the one in the denominator of the exponentiated term) plus the variance of g (equals unity per definition) times the signal strength of the envelope f. The square root term of d_K after the exponential is a normalization factor w.r.t. f. We deduce that for a constant signal envelope f and vanishing correlations of g

$$\begin{aligned}\mathcal{P}\left(f|\vec{d}\right) &\propto \exp\left[-\frac{1}{2}\frac{\sum_K d_K}{1+f^2}\right]\sqrt{\frac{1}{(1+f^2)}} \\ &= \exp\left[-\frac{1}{2}\frac{\langle d\rangle}{\frac{(1+f^2)}{N_K}})\right]\sqrt{\frac{1}{(1+f^2)^{N_K}}}\end{aligned} \tag{4.15}$$

where N_K is the number of pulses involved. This probability density function attains it's maximum for $\langle d\rangle \geqq 1$ at $f = \sqrt{\langle d\rangle - 1}$ justifying the average data subtracting the noise background as the maximum likely guess. Carrying out a saddle point analysis, we conclude that

$$\sigma_f = \sqrt{\frac{(1+f^2)^2}{(2N_K f^2)}} \quad = \begin{cases}\sqrt{\frac{1}{2N_K}}\frac{1}{f} & \text{for} f \ll 1 \\ \sqrt{\frac{1}{2N_K}} f & \text{for} f \gg 1\end{cases}$$

For a signal that is constant and not the modulation of a stationary noise process, we would have expected $\sigma_c \propto \sqrt{c/N_K}$ as $c \gg 1$ such that the relative error decreases with signal strength. In our case the relative error decreases too, but only to reach 1 as a limit (see Fig. 4.3). Against our intuition we expect the shape of the profile to be most uncertain at the highest signal values while the areas of lowest uncertainty can be found where $f \approx 1$.
We emphasize that this relative uncertainty will not be integrated out by measuring more pulses, as the uncertainty in both high and low signal areas equally fall with the sqare root of N_K leaving the ratio of the remaining error untouched. This fact explains the often expressed observation that stronger pulses do exhibit larger timing errors [Osłowski et al., 2011,

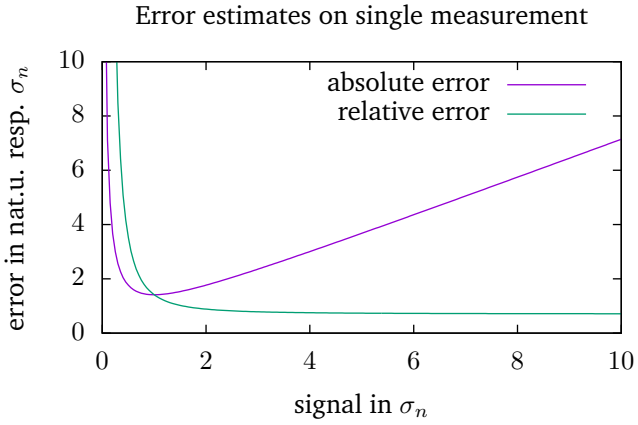


Figure 4.3: Comparison of absolute and relative error (as uncertainty over actual signal strength) for a single measurement of a signal f

Shannon et al., 2014]. The reason was found implicitly by numerous authors (e.g. Cordes [1993], Kulkarni [1989], Rickett [1975]) for the case of self-correlation of the signal. We emphasize that even without self-correlation, we expect the higher pulses to have a higher intrinsic uncertainty.
The non-diagonal terms introduced by G mark the departure from simply averaging the data. Let us review the influence of a non-white stationary process on the data's spectrum.

The influence of g on the signal's spectrum

When investigating a Fourier transform of a whole observation's time series containing several hundred pulses (see fig. 4.2) two prominent features dictate the spectrum: Spikes of the pulsar's signal dominate over receiver background noise. If the signal was strictly periodic, we would expect it to be described solely in terms of the Fourier coeffieents of harmonics of the base frequency $1/P$. As the pulsar's signal contains periodicity, we expect spikes to show up in this case too[2]. However, we are facing a stochastic process g that is modulated via a periodic envelope f. As is commonly known, multiplication in time domain translates to convolution in Fourier domain. As $g(\omega)$ is expected to be a red noise process, we expect it to cast the sharp pulsar signal to finite width in Fourier space. Indeed - as the inlay of the figure shows - the signal's power is leaked to the neighbouring bins. The width of the spectral line however is rather narrow, pointing to a fast decline of the spectrum of g in Fourier space. Consequently, the spectrum of g in time domain is spread out widely, pointing to non-vanishing correlations over a long time. These correlations diminish the independence of both adjacent lying bins and the bins of neighbouring pulses.
As the multiplication $f(t) \cdot g(t)$ amounts to a convolution in Fourier space, evaluating the posterior distribution in Fourier space exactly was found to be unfeasible. Instead, we approximated the convolution by taking only the diagonal terms, losing mathematical rigorosity and precision but keeping the problem solvable in reasonable computational time. The interested reader may refer to Sec. 4.7 of the appendix for a more detailed discussion why solving this in time domain would be favorable, but currently is not feasible.

[2] these spikes are not an artifact of folding the data as no folding was applied

The clear departure from the average comes from the off-diagonal elements of $\sqrt{\mathbf{d}}M\sqrt{\mathbf{d}^\dagger}$ where M is the operator derived. As is evident from the structure of M, it is expected to be off-diagonal only until the correlations introduced by $G(t-t')$ decay. Thus, writing this equation as a matrix equation, symmetry tells us that we can possibly store only a limited combination of $\sqrt{d}(t-\Delta t)\sqrt{d}^*(t+\Delta t)$ values and average them as they are multiplied by the same number when performing the matrix multiplication. The matrix M, subject to the exact envelope shape and correlations of the radiation process can be calculated afterwards and former data may be reprocessed with the new M in that way without losing the statistical information from the single pulse level, even though we may integrate parts of the single pulse data to yield smaller data sets. This could possibly be exploited in the future to profit from single pulse statistics without having to store all single pulses.

Maximum-a-posteriori-Filter

We are interested in finding the $\overline{\mathbf{f}}, \sigma_\mathbf{f}$ values most compatible with the data and thus set out to maximize the a-posteriori distribution[3] given the data. For a prior specified by $\sigma_{f,p}$ and $\overline{\mathbf{f}}_p$ we derive the integral over f of (4.14) w.r.t. $\overline{\mathbf{f}}$ and expect the derivative to vanish. This procedure yields an implicit formula for maximizing the posterior distribution of $\overline{\mathbf{f}}$.

$$
\begin{aligned}
0 &\stackrel{!}{=} \iint \mathrm{D}\mathbf{f} \sum_{K=0}^{K_{\max}} \sum_{k=0}^{k_{\max}} \frac{(f_{K,k}-\overline{f}_k)}{\sigma_k^2} \exp\left[-H_{\overline{f}_{k,p},\sigma^2_{f_p}}[f_{K,k},d_{K,k}]\right] \\
0 &= \sum_{K=0}^{K_{\max}} \sum_{k=0}^{k_{\max}} \int \mathrm{d}f_{K,k} \frac{(f_{K,k}-\overline{f}_k)}{\sigma_k^2} \exp\left[-H_{\overline{f}_{k,p},\sigma^2_{f_p}}[f_{K,k},d_{K,k}]\right]
\end{aligned}
$$

which may be solved for every k independently as

$$
\begin{aligned}
0 &= \sum_{K=0}^{K_{\max}} \langle f_{K,k}-\overline{f}_k\rangle_{\exp\left[-H_{\overline{f}_{k,p},\sigma^2_{f_p}}[f_{K,k},d_{K,k}]\right]} \\
\overline{f}_k &= \langle f_{K,k}\rangle_{\exp\left[-H_{\overline{f}_{k,p},\sigma^2_{f_p}}[f_{K,k},d_{K,k}]\right]}
\end{aligned}
\tag{4.16}
$$

where we organize our data in pulse numbers indexed K and have summed over single Fourier coefficients d_k of these data of pulse K.
The new variance assumed of $\sigma_\mathbf{f}$ may be calculated equally by evaluating the expectation value of $(f_{k,K}-\overline{f}_k)^2$, since deriving w.r.t. σ_k yields:

$$
\begin{aligned}
0 &\stackrel{!}{=} \int \mathrm{d}f \sum_{K=0}^{K_{\max}} \left(\frac{(f_{K,k}-\overline{f}_k)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right) \exp\left[-H_{\overline{f}_{k,p},\sigma^2_{f_p}}[f_{K,k},d_{K,k}]\right] \\
0 &= \sum_{K=0}^{K_{\max}} \sum_{k=0}^{k_{\max}} \int \mathrm{d}f_{K,k} \left(\frac{(f_{K,k}-\overline{f}_k)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right) \exp\left[\cdots\right]
\end{aligned}
$$

[3] Maximum A Posteriori or MAP refers to taking the most likely value of the posterior as best guess

which may be solved for every k independently as

$$0 = \sum_{K=0}^{K_{\max}} \langle((f_{K,k} - \overline{f}_k)^2 - \sigma_k^2)\rangle_{\exp\left[-H_{\overline{f}_{k,p},\sigma^2_{\overline{f}_p}}[f_{K,k},d_{K,k}]\right]}$$
$$\sigma_k = \sqrt{\langle(f_{K,k} - \overline{f}_k)^2\rangle_{\exp[\cdots]}} \tag{4.17}$$

Doing this once is assuming a prior of the form of the initial $\overline{\mathbf{f}}, \sigma_{\mathbf{f}}$. A very general choice may be suitable. Then taking the expectancy values of (4.16) as prior input to (4.14) and iterating is known as the expectation-maximization (EM) algorithm proposed by Dempster et al. [1977]. We may in this way in principle get rid of assumptions going into the first prior by iterating since this procedure converges on a prior fully compliant with the data, assuming gaussian distribution of f. However the EM algorithm may only find local maximum which equals a global maximum if the pulsar is really radiating with a log-normal probability distribution. The optimization problem in both $\overline{\mathbf{f}}$ and $\sigma_{\mathbf{f}}$ may be solved using a suitable iteration method (see e.g. appendix 4.8). We assumed here that the pulsar is sufficiently described by a log-normal distribution and that the dataset converges to the same (global) fix point for all initial pairs of values. In principle, a detour from this assumption, for example assuming moding (see sec.4.3.7), may open the possibility to filter out or analyse phenomena like giant pulses, subpulse drift separately. At this stage of development, we like to emphasize that ,assuming a log-normal distribution, we weight large fluctuations in a nonlinear fashion while retaining a scale-free prior.

4.2.3 Detuning model

The problem of determining ToAs is closely related to a phase shift over the analyzed single pulses. A wrongly set folding period can cause phase-drifts over the dataset's pulses and correlates different Fourier coefficients of the true signal. While this only leads to minuscule smearing of the profile when integrating, it may have a more serious impact on single pulse coefficients. As a wrong folding may go unnoticed on sub-bin scale over the whole epoch, it is crucial to study it's influence on phase measurements to determine its possible impact on timing.
Analysing Fourier transformations of finite (consecutive) pulses allow us to measure a phase-drift corresponding to the detuning of the assumed periodicity τ and the real periodicity $T = \frac{\tau}{a}$, where a is a correction factor. The equations denoted with a letter before the number are derived in appendix 4.6.2. For the signal from a single Fourier coefficient with frequency $\omega' = \frac{2\pi n}{T} = \frac{2\pi n a}{\tau}$ we will get a Fourier coefficient over the period $[K\tau - \frac{\tau}{2} : K\tau + \frac{\tau}{2}]$ of pulse K:

$$\widetilde{d}_{k,K} = s_n \exp\left[2\pi i k(\frac{n}{k}a - 1)K\right] \cdot \mathrm{sinc}[\pi k(\frac{n}{k}a - 1)] \tag{4.18}$$

where s and d with subscripts k, K denotes the k^{th} complex Fourier coefficient of the K^{th} sample. For $n = k$ eq (4.18) reads

$$\widetilde{d}_{k,K} = s_k \exp\left[2\pi i k(a-1)K\right] \cdot \mathrm{sinc}[\pi k(a-1)] \tag{4.19}$$

which means that, since $a \approx 1$, the sinc-factor is approximately 1 for $k \ll \frac{1}{(a-1)}$.Thus the relative phase of the K^{th} and the $(K+1)^{\mathrm{th}}$ pulse gives access to the model parameter a.
Let us now derive the Fourier coefficients for the full spectrum of a periodic, purely real signal.

for such a signal, $s_n = s^*_{-n}$. Inserting this relation and denoting the phase as $\arg(s_n) =: \Phi_{s_n}$ we get:

$$d_{k,K} = \sum_{n=1}^{\infty} a_n \mathrm{sinc}[\pi(na-k)] \cdot \left[\underbrace{\exp\left[i(2\pi naK + \Phi_{s_n})\right]}_{I} + \underbrace{\frac{d}{k}\exp\left[-i(2\pi naK + \Phi_{s_n})\right]}_{II} \right] \tag{4.20}$$

where $\overline{k} = \frac{an+k}{2}, d = \overline{k} - k$. This formula can be written as $d_{k,K} = \sum_{n=1}^{\infty} A_{kn} a_n$. The main phase information again is to be found in term I of (4.20) for $n = k$. However this is not the only term relevant in analysing a single coefficient $d_{k,K}$. Assuming $n = k \pm m$ the sinc function slowly converges towards zero as $\frac{1}{m}$ and determines the factor for coefficient 1. Coefficient two together with the sinc function goes like $\frac{1}{2k \pm m}$ for $m > 0$. Thus in the limit of big difference in k and m, both terms equally contribute and the phase relation deforms more and more to an elliptic one.
We emphasize that for small detuning, the sinc-term is nearly zero for $k \neq n$. In this case we may calculate only with the diagonal matrix and thus correct the data for the phase $\Delta\Phi(a) = \exp\left[i2\pi naK\right]$. Thus for $(a-1) \cdot n_{max} \cdot K_{max} \ll 1$, the detuning amounts to less than a bin over the whole epoch, and does not significantly affect phase values or leak signal into neighbouring coefficients. Thus for all practical purposes we can use the phase shift model in Sec. 4.2.6 to generate ToAs without having to worry about a possibly present sub-bin detuning in the folding period.

4.2.4 Phase model

Even for a perfectly tuned signal, we still have to face the noise introduced by the receiver and antenna system temperature on the phase. Again we will have to assign a measurement model for the data phase according to the signal. The noise passes the same linear transformation than the true Fourier coefficients and we conclude that

$$d_{k,K} = \sum_{n=1}^{\infty} A_{kn}(s_n + n_{n,K}) \tag{4.21}$$

where n is a white noise process with variance σ_n^2. Since A is dependent on the true signal phases, the inverse problem as a whole is quite complex. But if we neglect term II of (4.20) we may invert the remaining matrix $\widetilde{A}$ without considering the true signal phases and apply it to a given dataset.

$$\underbrace{\sum_{k=1}^{\infty} \widetilde{A}^{-1}_{nk} d_{k,K}}_{=: d'_{k,K}} = ||s_n|| \exp\left[i\Phi_{s_n}\right] + n_{n,K} \tag{4.22}$$

$$\widetilde{A}_{kn} := \mathrm{sinc}[\pi(na-k)] \exp\left[i2\pi naK\right] \tag{4.23}$$

This is possible since term II destroys phase information in a very smooth way and - as will be concluded below - high coefficient numbers will have higher phase errors due to the drop in signal amplitude and the relative growth of error. Consequently, higher coefficients do not give as much information about the phase and thus about a anyway. Having preprocessed the data with $\widetilde{A}^{-1}$ we arrive at the noisy signal and the phase carries information about Φ_{s_n}. Thus, given a and having processed the data vector, we may now fit for the average signal phase

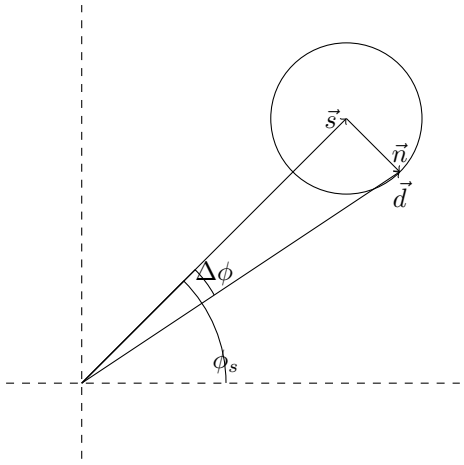


Figure 4.4: The phase error introduced to a signal by white noise scales as the expectancy value of the arctan of the signal to noise ratio

and variance by imposing a suitable model. We may estimate the error introduced in the true signal's phase by inspecting eq. (4.21) and Fig. 4.4:

$$\tan(\Delta\Phi) \approx \frac{||n||}{||s||} = \frac{\sigma_{\mathrm{n}}}{\sigma_{\mathrm{n}} \exp[2f]} \tag{4.24}$$

Rather than integrating this precisely one can use the following approximation without significant loss of accuracy

$$\sigma_{\Phi_s} := \langle \arctan(\exp[-2\mathbf{f}]) \rangle_{\mathcal{P}(\mathbf{f}|\mathbf{d})} \tag{4.25}$$

$$\mathcal{P}(\mathbf{\Phi_d}|\mathbf{\Phi_s}) = \prod_i \sum_{k=-\infty}^{\infty} \exp\left[-\frac{(\Phi_{d_i'} - \Phi_{s_i} + 2\pi k)^2}{2\sigma_{\Phi_n}^2} \right] / C \tag{4.26}$$

This approximation is valid only for a good signal to noise ratio (S/N), where we also expect the most significant evidence to be. For low S/N it underestimates the error and we must numerically integrate the complete formula to get an error estimate. We emphasize that this is the very step where the confidence of the phase data and thus the weight in later calculations is calculated using the actual amplitude data point measured together with the statistics gathered. $\mathbf{\Phi_s}$ itself is parametrized by a so-called wrapped Gaussian[4] and a mean of $\overline{\mathbf{\Phi_s}}$, and variance $\sigma_{\mathbf{\Phi_s}}$. we assume a prior of zero average and large variance.

[4] a wrapped Gaussian is literally a gaussian distribution wrapped around the unit circle by identifying all points modulo 2π

4.2.5 Joint probability for phase

Finally we arrive at the joint probability of the phase model calculating analogue to the amplitude:

$$\mathcal{P}\left(\mathbf{\Phi_d}, \mathbf{\Phi_s}|T, \sigma_{\Phi_n}, a\right) = \prod_i \sum_{k=-\infty}^{\infty} \exp\left[-\frac{(\Phi_{d_i'} - \Phi_{s_i} + 2\pi k)^2}{2\sigma_{\Phi_n}^2}\right] \prod_j \sum_{l=-\infty}^{\infty} \exp\left[-\frac{(\Phi_{s_j} - \overline{\Phi}_{s_j} + 2\pi l)^2}{2\sigma_{\Phi_s}^2}\right] /C \tag{4.27}$$

Where $T := \{\overline{\mathbf{f}}, \sigma_{\mathbf{f}}, \sigma_{\mathbf{\Phi_s}}, \overline{\mathbf{\Phi_s}}\}$ will be the parameters of the full statistical template reconstructed and $\mathbf{\Phi_{d'}} = \mathbf{\Phi_{d'}}(\mathbf{\Phi_d})$ is the phase correction according to (4.23) or other corrections for the assumed phase shift. Integrating out $\mathbf{\Phi_s}$ we find:

$$\mathcal{P}\left(\mathbf{\Phi_d}|T, \sigma_{\Phi_n}, a\right) = \prod_i \sum_{k=-\infty}^{\infty} \exp\left[-\frac{(\Phi_{d_i'}(\Phi_{d_i}) - \overline{\Phi}_{s_i} + 2\pi k)^2}{2(\sigma_{\Phi_n}^2 + \sigma_{\Phi_s}^2)}\right] /C \tag{4.28}$$

We now again invert this relation using Bayes' theorem and take a MAP-Ansatz on our model parameters $\sigma_{\mathbf{\Phi_s}}, \overline{\mathbf{\Phi_s}}, a$.

4.2.6 Phase shift model for a single epoch as reference

In order to generate ToAs we will have to determine the relative shift in the time of arrival of the pulses compared to another epoch taken as a reference. For the sake of simplicity, we assumed that the pulsar period does not to change with time (which is only a matter of interpreting the relative phase correctly or further imposing the detuning model). The pulsar's signal is assumed to be slipped by a time Δt compared to the overall period. This amounts to a change in the phase of the Fourier coefficients given by:

$$d_k = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \mathrm{d}t \quad d_t' \cdot \exp\left[-i\omega_k(t - \Delta t)\right] = d_k' \cdot \exp\left[i\omega_k \Delta t\right] \tag{4.29}$$

Thus, having reconstructed the template characteristics $T = \{\overline{f}_k, \sigma_k, \overline{\Phi}_k, \sigma_{\Phi_s}\}$ for the first time interval, the subsequent intervals may be analysed by carrying out a parameter study over Δt. This is done by shifting the measured phases $\overline{\mathbf{\Phi}}_{\mathbf{d}} \to \overline{\mathbf{\Phi}_{\mathbf{d'}}}$ according to (4.29) and taking e.g. a MAP-approach over the joint probability of the phase (4.28), now varying Δt, not a which is assumed to be 1.
We now are also able to give the approximation for $a \approx 1$. We derived in this case a phase shift of $\Delta\Phi(a) = \exp\left[i2\pi naK\right]$. Comparing with (4.29), the problem of detuning for small a may be approximated by testing all pulses for a shift of

$$\Delta t = aK \cdot T \tag{4.30}$$

dependent on the pulsar period number K. All practical cases of detuning fall in the category of the aforementioned reduction to systematic phase shift.

4.2.7 Reference independent difference phase model

The simplest way to form a reference template is to select a single observation. However this relies on the chosen observation being statistically representative of the sample as a whole and

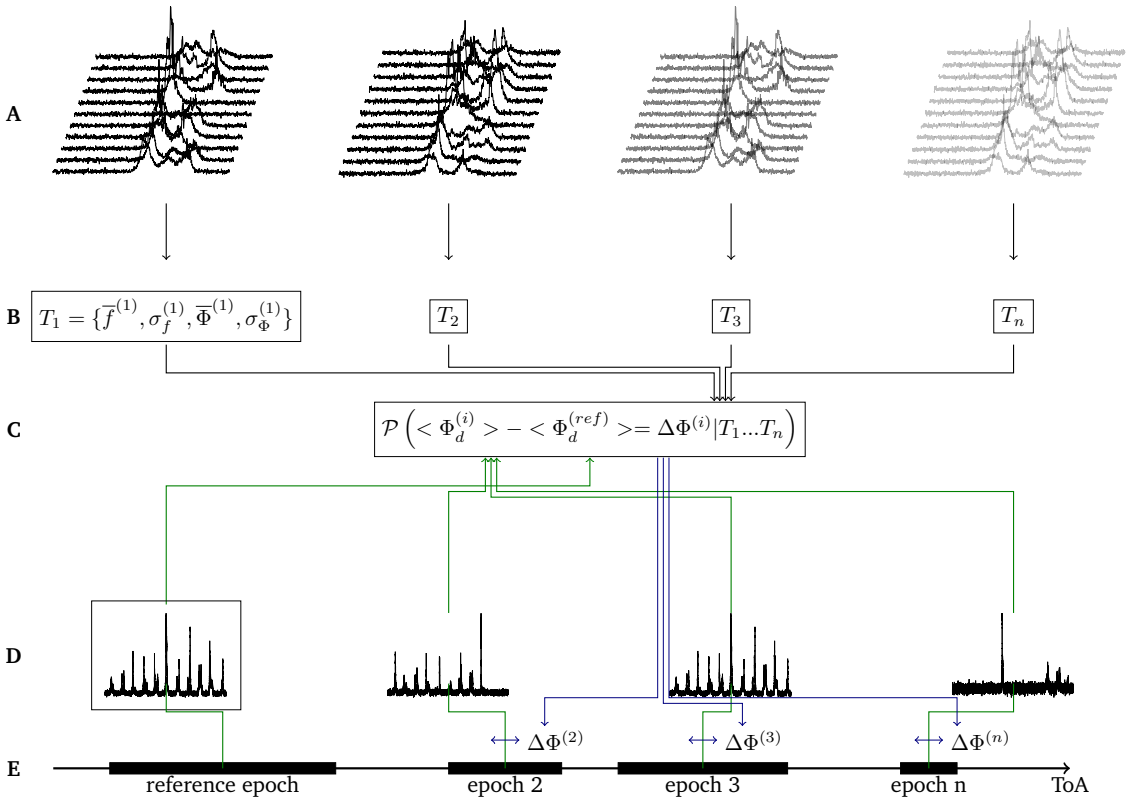


Figure 4.5: **Overview of ToA generation**: The statistical imprint of the various epochs (**A**) is used to generate statistical templates (**B**). The data of the epochs (**D**) is non-linearly filtered using the statistical templates to yield a probability distribution (**C**) for the shift of every epoch relative to a reference epoch. For every epoch, the mean phaseshift in topocentric pulsar periods is then added to epochs MJD to calculate the specific ToA (**E**). The error is calculated as the variance of (**C**) for this shift. The reference epoch's error on the ToA is determined as if it would have been timed by every other epoch.

would discard the extra information available in the whole dataset.
This in principle could be mitigated by building a total statistical template over all epochs observed. We discarded this method for two reasons. First of all it assumes that the pulsar template shape and statistical behaviour is the same in every single epoch (which is not the case for e.g. moding pulsars). Furthermore forcing the template to take the same probability density in signal amplitude for distinct epochs is too strict an assumption as for example ISM variations modulate the signal amplitude systematically. Instead, one may formulate a probability for the relative shift of two distinct epochs, given a certain epoch's statistical template is correct, and then demand that the probability distribution has to agree for all templates at once. We depicted this process schematically in Fig. 4.5. The calculated probability on a relative phase between two epochs, given all knowledge gathered, is a clear, basic statement coming without the notion of a "reference epoch". It can be incorporated in further analyses without introducing a direct bias caused by the phase information of statistical templates used since the exact phase assumed in the template drops out of the calculation. The knowledge gathered just filters the data to be compared in a clever way.
Mathematically speaking, we demand that

$$\mathcal{P}\left(\langle\Phi_d^{(i)}\rangle - \langle\Phi_d^{(ref)}\rangle = \Delta\Phi^{(i)}|T_1...T_n\right) = \prod_j \mathcal{P}\left(\langle\Phi_d^{(i)}\rangle_{T_j} - \langle\Phi_d^{(ref)}\rangle_{T_j} = \Delta\Phi^{(i)}|T_j\right) \tag{4.31}$$

The formula for one factor of this product may be deduced from (4.28) by taking the product of the equation for two epochs' shifts $\Delta\Phi_1, \Delta\Phi_2 = \Delta\Phi_1 + \Delta\Phi$ and integrating out the difference to the used template, $\Delta\Phi_1$. A factor then reads:

$$\begin{aligned}\mathcal{P}\left(\langle\Phi_d^{(i)}\rangle_{T_j} - \langle\Phi_d^{(\mathrm{ref})}\rangle_{T_j} = \Delta\Phi^{(i)}|T_j\right) = \\ = \prod_k \frac{1}{\sqrt{2\pi{\sigma_k^{(i)}}^2}} \cdot \exp\left[\frac{1}{2{\sigma_k^{(i)}}^2}(\langle\Phi_d, k^{(i)}\rangle_{T_j} - \langle\Phi_d, k^{(\mathrm{ref})}\rangle_{T_j} - k\Delta\Phi)^2\right]\end{aligned} \tag{4.32}$$

Where ${\sigma^{(i)}}^2 = {\sigma_{s',k}^{(i)}}^2 + {\sigma_{s',k}^{(\mathrm{ref})}}^2$ is the sum of the new phase variances estimated from the epochs' data, given T_j holds. The formula deduced simply states that the difference of the filtered mean phases of the two epochs under consideration is the mean phase difference times the Fourier coefficient number.
While the gathered probabilities on phase differences are reference free, in order to output a TOA, we have to define that the absolute phase of a certain epoch amounts to zero. In the classical procedure, the template was assumed to have zero phase. Additionally there exist several conventions on defining the physical point on a profile where the phase is zero, including but not limited to defining the mid of the profile at the "center of mass" of the profile or at the highest peak. To that extent we declare a "reference epoch" which's ToA is declared to be exactly the timestamp of the rising edge of the first bin of the pulse. As we assume the epoch with the highest S/N-ratio to have the lowest variances in its statistical template as seen from the others, we choose this one to be the reference epoch. While, per definitionem, this TOA's relative phase is exactly zero, the corresponding ToA has to have a variance to be fitted into TEMPO2. Thus we calculate the variance of the reference as it would have been determined by all other epoch's relative shifts as

$$\frac{1}{{\sigma^{(\mathrm{ref})}}^2} = \sum_j \frac{1}{{\sigma^{(j)}}^2} \tag{4.33}$$

For the other epochs, the probability distribution of the relative phase shift provides the correct

error estimate on the ToA automatically.
Another big advantage of parametrizing the ToAs by their difference is that (4.31) can be calculated independently for every template and added logarithmic. This can be done since the measured variable, phase difference, is the same for every pair of epochs, no matter which statistical template is assumed for the moment. This gives us the freedom to rasterize the probability distribution as a whole without preferring one template over the others and stay in the Bayesian picture until reducing the probability distribution measured into mean phase differences and variances in the very last step. This becomes essential when also short epochs with weak signal over noise ratio are to be included in the analysis. In this case multiple phases are still plausible e.g. mistaking one peak for the other (see also fig. 4.7 and Sec. 4.3.2) since the yielded probability distribution for one template may be too far from the gaussian form to be reduced to a mean and variance. The ToA t_{ToA} of epoch i with mean phase shift $\Delta\Phi^{(i)}$ and variance ${\sigma^{(i)}}^2$w.r.t. the declared reference then amounts to:

$$t_{\text{ToA}} = t_{\textit{rising edge}} + P_{0,i} \times (\Delta\Phi^{(i)} \pm \sigma^{(i)}) \tag{4.34}$$

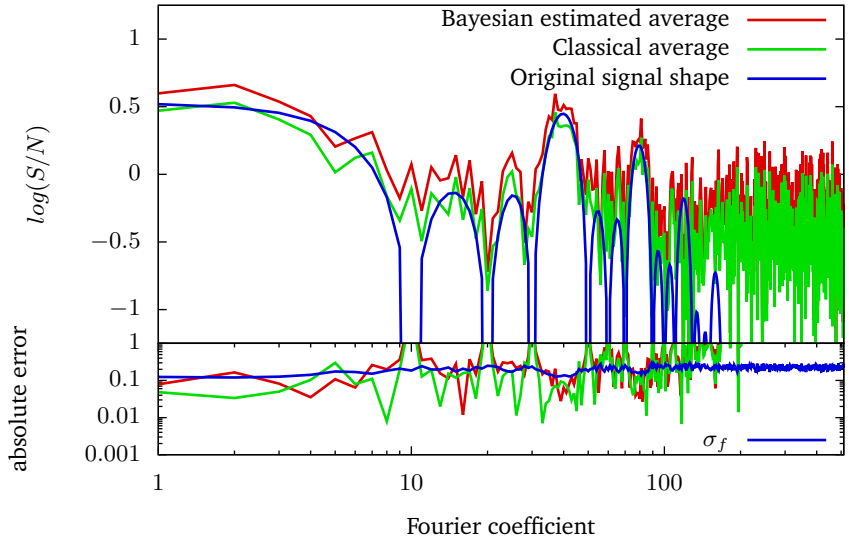
where $P_{0,i}$ is the folding period of epoch i and the phase difference takes values in the interval $[-.5 : .5]$.
Developing an interface to Bayesian extensions of TEMPO2, like TempoNest[Lentati et al., 2014], to communicate the whole probability distributions on the relative phase information to the pulsar timing code would be a further step to enhance the statistics on the ephemeris' parameter sets. While this may be desirable in the future we simply reduce the yielded probability distributions to their mean and variance and calculated ToAs directly comparable to the ones from tools like pat included in the PSRCHIVE toolset and processable by TEMPO2 without modifications.

4.3 Application to data

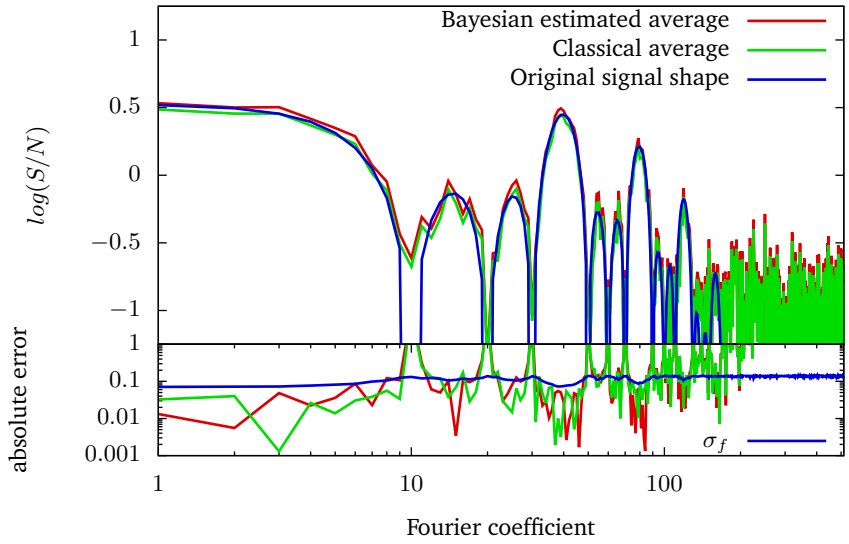
Having derived all relevant equations we may now apply the formulae upon given datasets and test the approach. We analyse (frequency and polarisation integrated) single pulse data unless otherwhise stated. Starting with simulated datasets we afterwards evaluate a dataset of PSR B1133+16 with 2041 consecutive single pulses consisting of 1024 bins each. Furthermore we determined ToAs from short time integrations of data of PSR J1713+0747 and analysed these ToAs with the TEMPO2 software package comparing with ToAs generated using "pat" from PSRCHIVE and a very accurate classical analytical template.

4.3.1 Simulated data

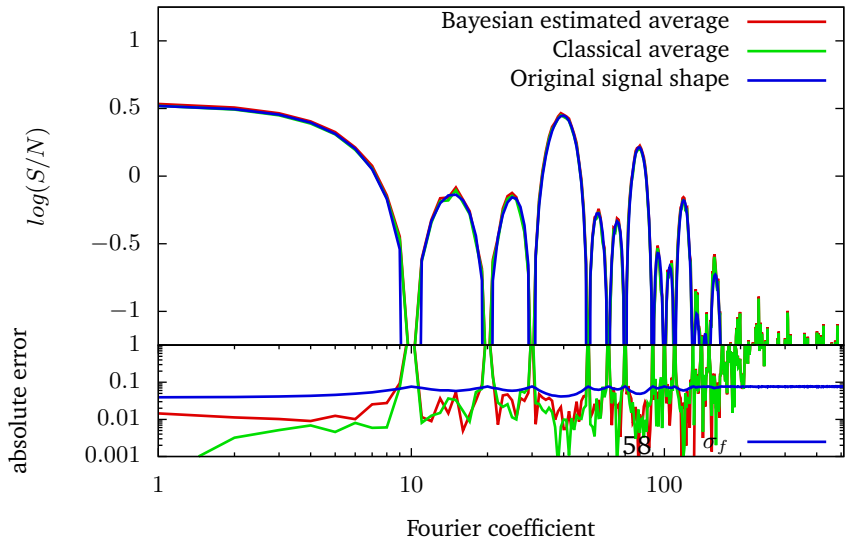
We wrote a simulated data generator serving single pulses of non-fluctuating shape $\bar{\mathbf{f}}$ but multiplied with gaussian red noise to account for a highly fluctuating radiation process and analysed the convergence of the amplitude model in Fourier space. The classical average is in this case the statistically optimal procedure to determine the pulse profile since the assumption that there is exactly one profile holds, and a quickly converging law of large numbers applies. Fig.4.6 depicts the reconstruction from ten, hundred and a thousand pulses comparing classical template integration with Bayesian reconstruction. Both methods perform equally well, as was expected. However, the Bayesian reconstruction also gives us an estimate on the profile stability and thus a better understanding of the remaining statistical uncertainty. Reconstructing a constant signal shape is a difficult task for a log-normal Bayesian model since the guess of a non-fluctuating shape is statistically only feasible by collecting lots of data. Thus, even though the test seems



After ten pulses, the Bayesian method has equally well reconstructed the signal as the average method, however it provides a handle on the signal fluctuations still possible. The seamingly systematic overestimation of the signal shape by the expected mean amplitueds comes from the log-normal exponential model used and the non vanishing $\sigma_{\mathbf{f}}$



A hundred pulses already give a profile accurate to over 90% over the essential Fourier modes of the signal. Notice how the real error on the averaged $\overline{\mathbf{f}}$ is limited to the still possible statistical fluctuations $\sigma_{\mathbf{f}}$.



A thousand pulses give a confident template up to high coefficients. The Bayesian reconstruction estimates the average shape to fluctuate below 20% of its mean. The absolute error not detected at the very high coefficients may be explained due to low assumed signal and low real signal to noise ratio. The dominant error source seems to be a systematic overestimation due to the log-normal signal model.

Figure 4.6: Reconstructions of the original signal shape by both Bayesian and classical methods and their errors after 10, 100 and 1000 pulses respectively. The lower half shows the errors of the estimation to the original signal shape and the Bayesian estimation of the remaining uncertainty $\sigma_{\mathbf{f}}$ on the expected log-normal mean $\overline{\mathbf{f}}$

.

simple, it is actually a very strict test for the Bayesian measurement model to pass as the correct solution for infinite observation time is delta peaked. The algorithm is expected to perform much better on real data for which the profile shape measured is much more fluctuating for small integrations since it has a way of incorporating the instabilities of the profile. Since in an observation of finite length the algorithm will never show zero variance on the profile shape, it estimates the profile intensity slightly but systematically too high. This gives rise to a problem if one is set out to measure the intensity of radiation naively by interpreting the statistical template as a way of describing the profile. The template gives a statistical way to determine all compatible intensities or the probability distribution of these, but not the "true" intensity directly.
The high modes exhibit a large absolute error. This is an effect of low signal to noise ratio at imperfect numerical integration. For low amplitudes, the Hamiltonian for different values of $\overline{\mathbf{f}}$ becomes rather flat and thus the outcome just reflects the prior we have been taken. In this case the prior was a log-normal Gaussian model with $2\sigma_{\mathbf{f}}$ around the average found. Notice that the $\sigma_{\mathbf{f}}$ the algorithm suggests (leaving aside the singularities of the profile) captures the true error (except for singularities and the imperfect numerical integration of high modes) giving a handle on the remaining statistical uncertainty.

4.3.2 Determining phase shift

We used a template with four peaks of Gaussian shape, sizes ranging from $1.8°$ to $3.6°$ and generated a dataset with an overall signal over noise ratio of $-17dB \approx \frac{1}{50}$. We compared two independently generated datasets of one hundred pulses with 1024 bins each, shifted by $1.8°$, which amounts to a shift of 5.12 bins. As Fig.4.7 shows, the Bayesian reconstruction yields $\Delta\Phi = -1.81° \pm 0.04°$. The other maxima in the probability distribution are of smaller size but reflect the the spacing of the templates peaks and the possibility of mistaking one peak for another. Having a local signal to noise ratio of four for every peak and a resolution of $360°/1024 \approx 0.35°$ available, we may estimate the uncertainty of every peaks location to be $1.3°$[5] in every single peak in every single pulse. Having derived the phase drift using 200×4 peaks, we classically calculate the remaining error due to the receiver noise to $\frac{1.3°}{\sqrt{2 \cdot 400}} \approx 0.046°$. In the fluctuation free case (the simulated emitted signal is exactly the four peaked shape before the receiver noise is added and the signal is squared) it can be shown that this error is the minimum reachable error. The test of our algorithm reaches this accuracy, too.

4.3.3 Real data

Generating templates from real data yields reconstructions of surprising confidence. Fig. 4.8 depicts the amplitude reconstructions from certain numbers of consecutive pulses. In each figure, the long-term convergence gathered from the whole dataset may serve as a reference. We emphasize that for data gathered by a pulsar with fluctuating shape, classical and Bayesian reconstruction are in general not expected to yield the same curve, since fluctuations are handled differently by both models. Both can only be compared to their own long-term limit. Furthermore the statistics of the profile seem to change with time for the dataset examined. We will analyse this short term behaviour in the next section.
The Bayesian reconstruction does exhibit lower noise at high fourier coefficients and converges

[5] Calculating the exact timing precision achievable is highly nontrivial. We estimated this error as follows: Assume a single peak at a S/N of one to be locateable to about its width, having a signal to noise ratio of four should increase this precision by a factor of $\sqrt{4}$. Adding the errors on the bin and resolution quadratically, we end up with an average of $1.3°$ per peak.

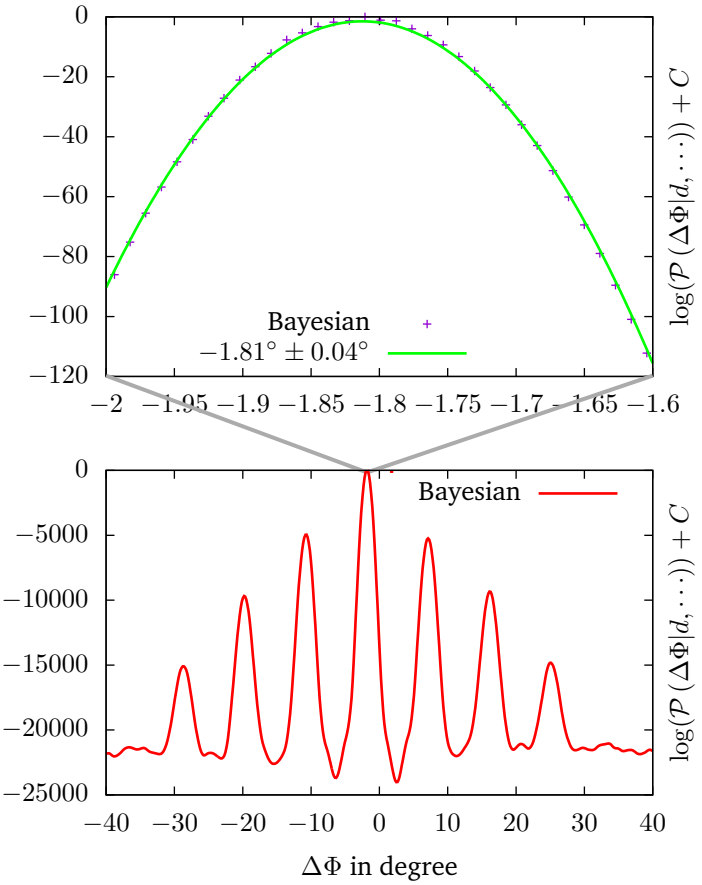


Figure 4.7: Unnormalized logarithmic probabilities for phase shifts. The comb structure comes from the four peaks present in the template.

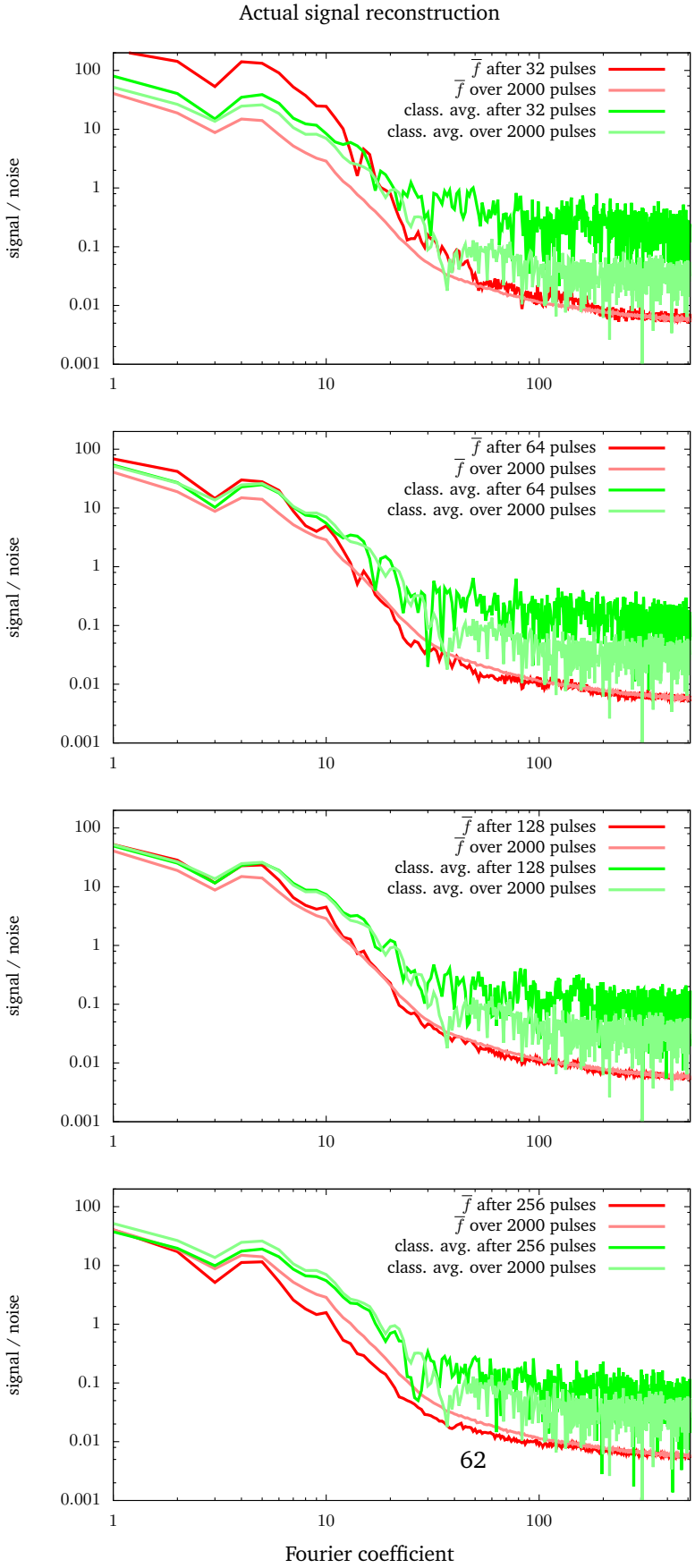
rather quickly to its ultimate form. The main features such as peaks and curvature appear already at 64 pulses while the classical average remains dominated by fluctuations occurring. However systematic offsets in the Bayesian reconstruction appear. These may largely be backtracked to the actual pulsars intensity changes and nulling periods. A numerical fluctuation of the algorithm is rendered unlikely since each point of the construction is calculated independently and thus an equal rise in all coefficients is highly improbable. The classical average also shifts upwards and downwards with varying intensity, but the effect is mostly indistinguishable from noise. For this real life example, the method derived is clearly superior to the classical method in areas of low signal to noise ratio, given that it converges very fast to its ultimate form and showing denoised systematic profile changes. It is reproducing the classical outcomes in the lower coefficients and not picking up the noise of the higher coefficients. After 128 pulses a confidence of about ten percent is already reached for all coefficients present. This paves the way to investigate changes in the radiation profiles on these scales. We could further diversify our pulses in intensity classes and perhaps in different modes of radiation given the contrast and fastness of the envelope estimate. Fig. 4.9 shows the differences of the two reconstruction methods in time domain and the single pulse probability density function as described by the Bayesian statistical template (depicted as grey shade): While the classical average just assumes the pulsar to have a certain profile shape, the Bayesian method is rather based on exclusion and describes profile shapes compatible with the data gathered. Thus, after as few as ten pulses, the Bayesian reconstruction in time domain already resembles the final form of the radiation profile. Notice that it still allows for signals out of the profile, since statistics do not exclude low signals after ten pulses.

4.3.4 Reconstruction stability to fluctuations

The workings and main differences of the two algorithms become evident when analysing the change of the classical profile and statistical template introduced by small changes of the dataset. This is examined by comparing the reconstructions of a window of fifty pulses moved along the dataset. An example of stepping one pulse further is shown in Fig. 4.10. Dropping one pulse and inserting another should not change the overall template much. If the template changes, we expect it to change smoothly with time. For the Bayesian reconstruction, this property can be easily observed whilst individual fluctuations dominate the classical average picture. This can be deduced from the right peak of the template. While the algorithm derived gives a rather smooth increase in intensity over the whole peak, the classical average is dominated by spikes of individual fluctuation. One could argue that the classical spikes might just be smoothed out of the form of the fourier transform. This is contradicted by highly localized changes in the Bayesian reconstruction that are statistically significant. The change of the left peak may be seen as an example for a localized statistical change. The classical average does not change much there, but the Bayesian reconstruction shows a significant rise of the left flank of the peak. The algorithm is found to behave in such a way over the whole dataset. Furthermore it incorporates the fluctuation seen in real envelopes. This may increase accuracy of derived parameters like the ToAs since e.g. the confidence of a peak appearing at a certain point should be, but is not, taken into consideration by classical timing procedures.

4.3.5 Nulling

Given a set of proposed models, like different statistical templates, Bayesian statistics can also assign a probability on how likely every single model is to describe the data, given that one of the proposed models is true. An example for the ability to discriminate between different



When looking at the high coefficients, the reconstruction is already quite confident.

Certain peaks and drops which are to be averaged out do not even appear in the Bayesian reconstruction

While the classical average is still not showing a clear ascend for the middle coefficients, the reconstruction clearly shows the final features.

The reconstructed curve is systematically shifted downwards exhibiting a change in average intensity while keeping the form of the curve unmodified.

Figure 4.8: Template reconstruction for a subset of pulses

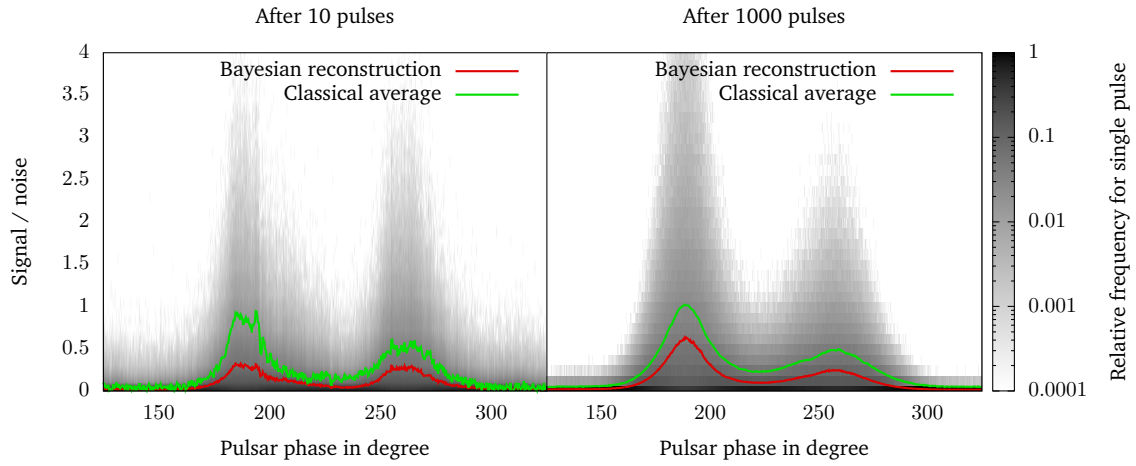


Figure 4.9: Template reconstruction after ten respectively. a thousand pulses. While the classical averaging assumes a certain fluctuating form to be true, the Bayesian template and single pulse model shows which single pulses are typically compatible with the statistics of the data gathered

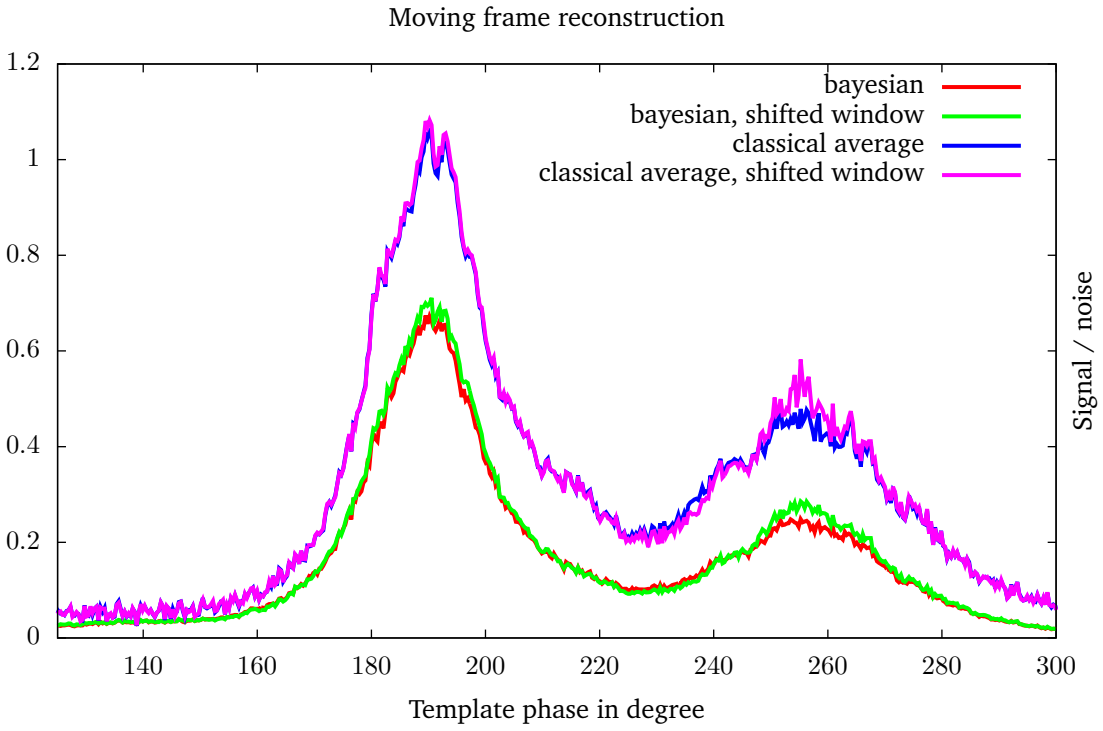


Figure 4.10: Depicted are the Bayesian and classic reconstructions from two sets of 50 pulses shifted by a single pulse. The change of the classical template is dominated by the spiky fluctuations as is evident from the change in the right peak.

models is nulling detection. Two sets of parameters are analysed within the framework: One set of $\bar{\mathbf{f}}, \sigma_{\mathbf{f}}$is generated from a training set of nulling periods taken from the first 50 pulses, the other one is taken from the overall reconstruction as above. The likelihood of a model i described by a set of parameters m_i is compared as

$$\mathcal{P}\left(m_i|\mathbf{d}\right) = \frac{\iint \mathrm{D}\mathbf{s}\mathcal{P}\left(m_i, \mathbf{s}, \mathbf{d}\right)}{\sum_i \iint \mathrm{D}\mathbf{s}\mathcal{P}\left(m_i, \mathbf{s}, \mathbf{d}\right)} \tag{4.35}$$

which may be derived straight-forwardly from Bayes' theorem. Deciding between the white noise model and the model derived from all thousand pulses, this formula gives the probabilities given in Fig. 4.11 for a certain single data frame to be nulling or the pulsar's signal. These probabilities could be used as weights when determining further parameters without biasing the statistics by deciding whether there was a pulse or nulling. Furthermore when looking at the plot, the Bayesian probabilities resemble our state of uncertainty for pulses which are in the 50% region when a by eye decision should be made. Fortunately, the method is able to quantify this uncertainty for us automatically. Nulling information could in principle further improve the accuracy of detuning and time of arrival analysis.

4.3.6 Determining Times Of Arrival

ToAs from simulated data

Timing simulated data sets is a precise way to test the algorithm for theoretical performance since the true signal to be found out of noisy data are known. We implemented the phase shift model (as described in Sec. 4.2.6). We generated a statistical template out of 10,000 simulated pulses and then measured the accuracy on timing $N = 10, \cdots, 50,000$ pulses with that fixed reference, where we used a MAP approach. Statistics was gathered over 20 randomly shifted datasets for each value of N recording the absolute error on the ToA and deriving the mean error and standard deviation to quantify the accuracy of the algorithm. The algorithm performed as depicted in Fig. 4.12. As expected, the systematic phase error was well bounded by the fluctuations given by the test datasets. The deviation from the true value followed a $1/\sqrt{N}$ law for large N as expected and reached the accuracy that is information theoretical possible. When there were more than 10,000 pulses in the dataset to time, a systematic error appeared. This was expected, since the reference statistical template itself (generated from 10,000 pulses) does not contain more accurate phase information. Since we were measuring an absolute, not a relative shift here, the error contained in the template becomes systematic. The sudden drop of the systematic error between 10 and hundred pulses is a random sampling artifact of the simulation involved. The average over the discrete maximum values taken was in this case accidentally zero. We also evaluate the reached accuracy for different S/N-ratios in Fig. 4.13. It is interesting to notice that the algorithm returns a rather flat probability distribution rendering a lot of phase values likely. This can be deduced from the high errors encountered leading to a plateau for low values of N. At a certain point the probability distribution begins to peak and accuracy increases with a jump, locking onto the signal. This may be explained by passing the point where different peaks of the profile shape can be distinguished statistically. This is not a drawback of the method but reflects the fact that, also in the ideal case, the error follows the $1/\sqrt{N}$ slope only for large N and a unique maximum. After that point the performance follows a stable $1/\sqrt{N}$ curve.

We also simulated the whole process from measuring epochs to determine the ToAs therein. To this end we simulated 20 epochs with 100 pulses each. Then we generated statistical templates from each epoch and used that to time the relative shifts of every other epoch. Then

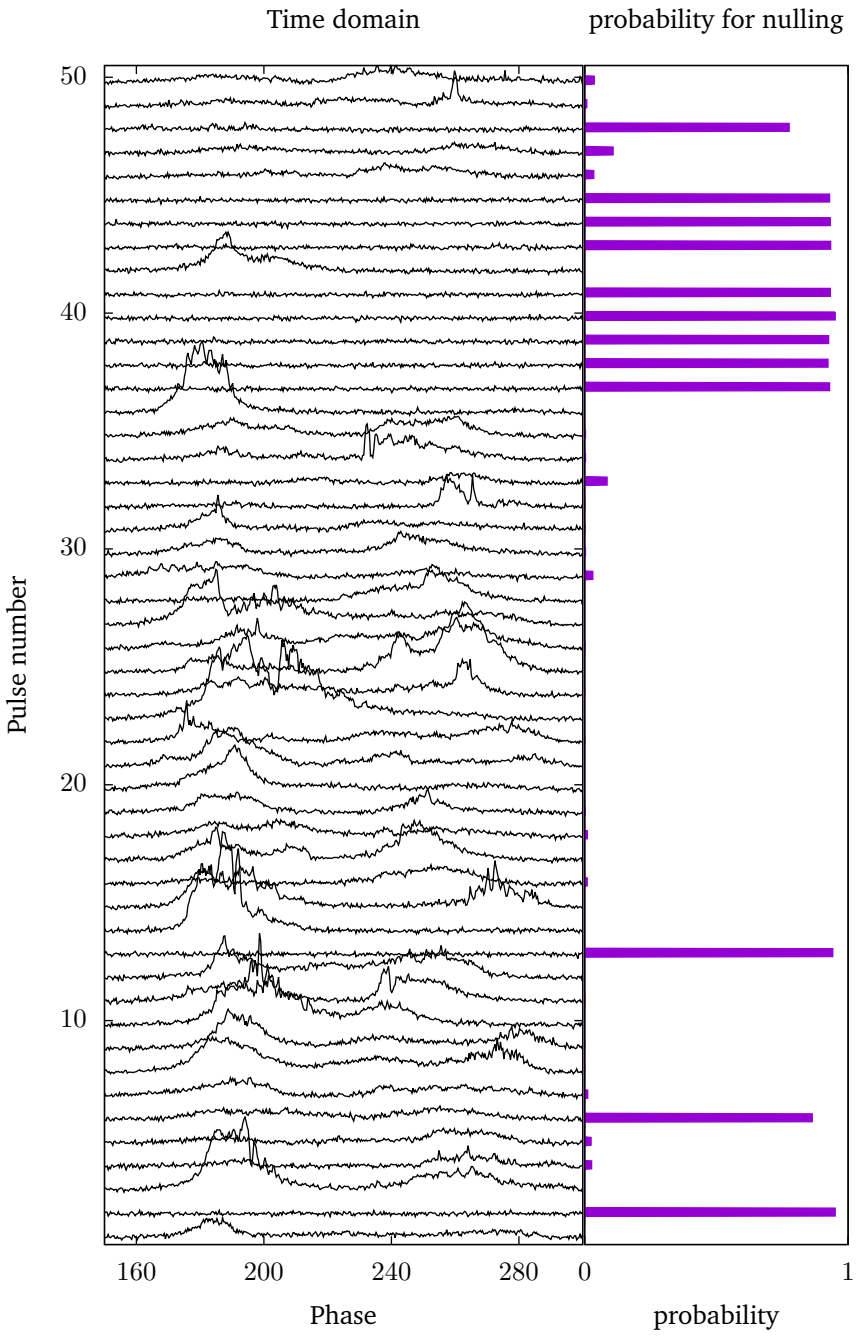


Figure 4.11: Discrimination-less decision: Bayesian analysis gives a probability for a certain pulse candidate of PSR B1133+16 to be explained by white noise only. In subsequent data-processing, this probability may be used to weight signals that are only present in non-nulling phase instead of a threshold algorithm producing false positives or negatives.

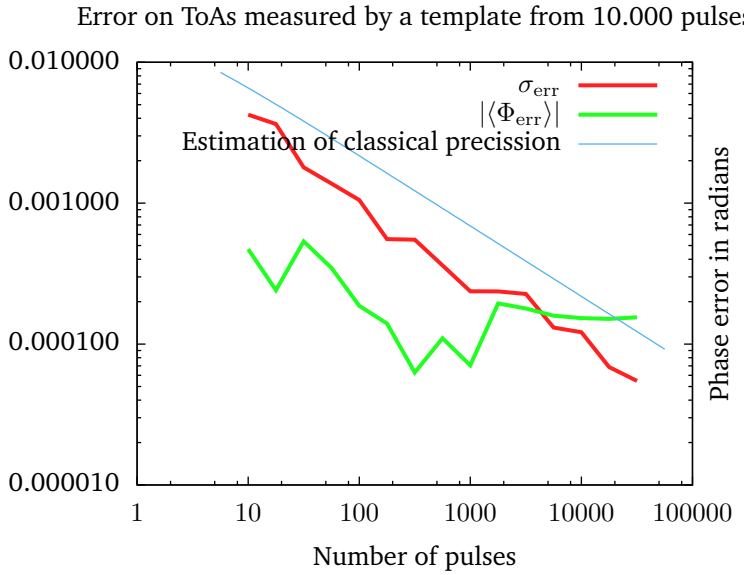


Figure 4.12: Test of accuracy over epoch size of the phase shift model as described in sec. 4.2.6. a variable number of pulses is used to generate ToAs and the algorithm is examined for statistical (red) and systematical (green) errors.

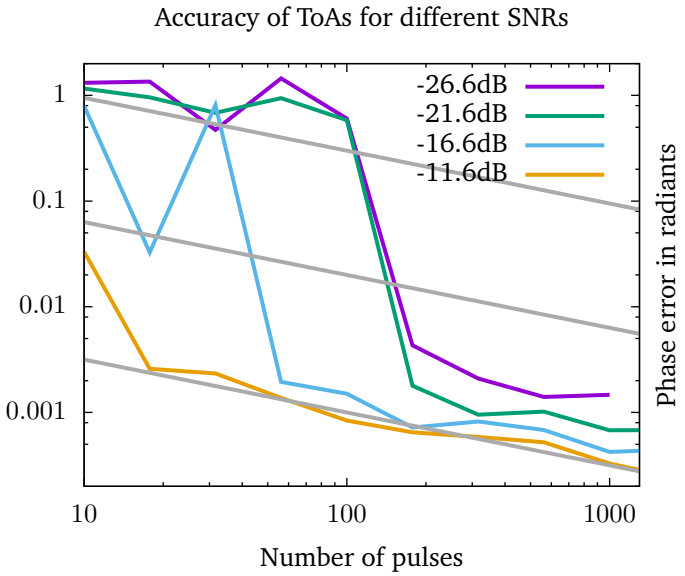


Figure 4.13: Test of accuracy for different S/N-ratios of the phase shift model as described in sec.4.2.6 the grey dashed lines all follow a $1/\sqrt{N}$ slope.

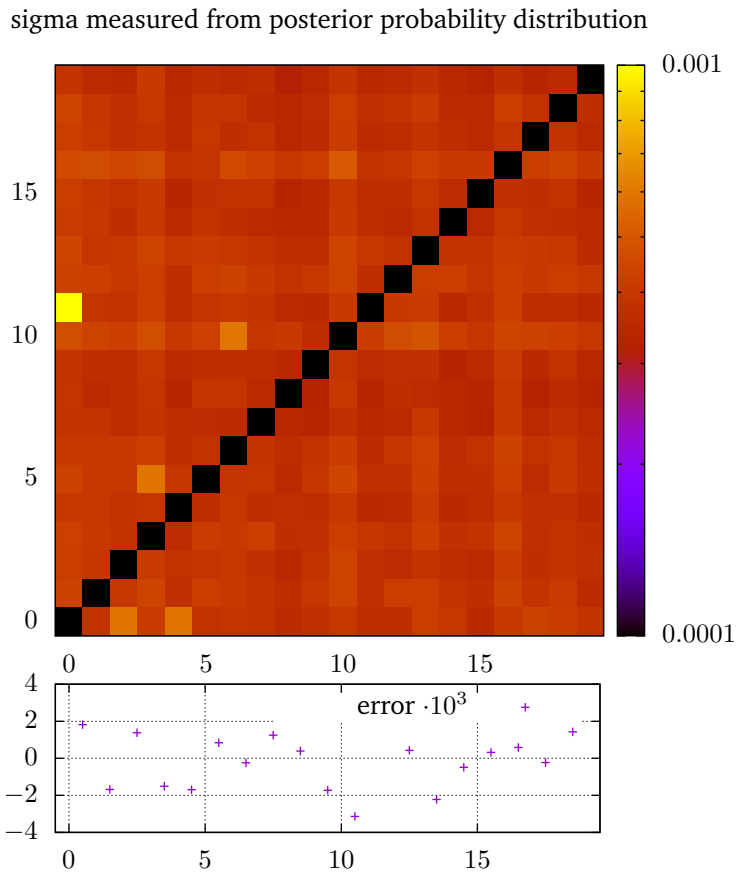


Figure 4.14: The upper panel depicts the remaining uncertainty (in an arbitrary scale) when the row epoch is measured by the column epoch. Notice that there are some combinations of epochs which lack statistical similarities and give greater uncertainties. The lower plot shows the remaining error on the ToAs of the epoch when combining the measurements from above.

we naively averaged the maximum a-posteriori values of the shifts and their variances using classical formulae for measurement value addition to find the ToAs and compare them with their true values. The dataset contained 2000 single pulses. The results of this simulated measurement is displayed in Fig. 4.14. We reach an accuracy of 1.9×10^{-3} radians, which is nearly the theoretical maximum for 100 pulses per epoch as can be seen from Fig.4.12. When timing with only 10 epochs, the accuracy dropped to about 3×10^{-2} radians. In that case, the theoretically possible accuracy for each epoch was missed. That practically means, that gathering knowledge of later epochs can improve measurements of the past by reprocessing them with the newer statistical templates.

ToAs from real data

The method carried out in such a way fails when tested with real data. Reducing the phase shift information of every statistical template to classical mean shifts and their variances before correlating the different templates' measurement flaws the precision of the method. This can be understood easily considering very noisy templates. These may cause different very likely shift values separated by a larger interval of unlikely phase values, e.g. different peaks of the profile may be mistaken for each other. When reducing, the mean and variance of only the most likely peak was determined giving a rather low sigma value on it, since in its neighbourhood, the probability density can be approximated as a gaussian. This sets a too large weight on the statistical template compared with the other templates if the algorithm accidentally locked on the wrong peak. The consequence is a large bias in the direction of the mistaken peak. However, if one does not reduce the data template by template, but examines the probability density on the relative phase shift of two epochs using all statistical templates at once, one circumvents the possibility of this local fallacy and gets an overall correct probability distribution.
An algorithm obeying these caveats has been outlined in Sec. 4.2.7. By measuring the same quantity (relative shift of two epochs) with different statistical templates we may rasterize the probability distribution for that quantity considering all references at once and in parallel. Since pulsar timing expects us to state a single ToA per epoch and the expected error on it, the MAP ansatz taken in the very last step is now justified.
We compare our algorithm using 48 epochs of ten second integrations of PSR J1713+0747 with a total observation duration of 25.8 hours covering about 1.6 years observed in Effelsberg. 31 epochs were observed at $1350\mathrm{MHz}$ and 17 epochs observed at $2650\mathrm{MHz}$, both with a receiver bandwidth of $200\mathrm{MHz}$. The reported errors on the ToAs are found to be in agreement with the RMS reported by TEMPO2 yielding low χ^2 values. Even though single pulse observations carry much more statistical information that the data at hand, the statistical templates generated show variations.
The ToAs generated reach an RMS of $967\mathrm{ns}$ and a weighted RMS of $533\mathrm{ns}$ at a fit χ^2 of 0.71. The classical method yields $895\mathrm{ns}$ unweighted or $337\mathrm{ns}$ weighted RMS with a reduced χ^2 of 17. We compare the ToAs generated by our algorithm with classical ones generated by the tool `pat`. To prevent overfitting we only fitted for phase shift in TEMPO2 and a shift between the two observation frequencies where appropriate using a recent ephemeris [EPTA data release, Desvignes et al., 2015] that is more accurate than the given observations. We collected the results in Table 4.1. Besides the slightly increased RMS and weighted RMS we note that the more appropriate χ^2 values of the Bayesian ToAs are strengthening our case.
With larger residuals being produced one could argue that the algorithm presented is less accurate and simply enlarges the error bars accordingly. We argue that this is not the case and the errors stated are more appropriately describing the accuracy of the data set at hand. First let us investigate the low χ^2 values which in some cases turn out to be smaller than unity. This

possibly points to overestimated errors. However, by using a more accurate ephemeris to time the data, we tend to line up the pulsar phases to a more accurate value automatically. Thus, a fit using only the data at hand is expected to be less precise, leading to an increase in χ^2 values. Furthermore, when reducing the probability density functions, we stated the maximum likely value for the ToAs. This educated guess must not agree with the expectation value and may turn out to be the better guess when having more data. In addition, the analysis was carried out using the statistical templates from two different frequencies. Inherent differences between the two frequencies' profiles, e.g. changed ISM effects and profile evolution, lead to different statistical templates. The two frequencies' templates consequently are not completely compatible with each other and tend to give enlarged errors expressing this discrepancy. Secondly let us address the worse weighted RMS values. As there are a few dominant observations in the classical case, the measured residuals reduce more when shifting the over all phase to the mean of the observations with the lowest error bars. Whereas the actual distribution of unweighted residuals shows about the same accuracy in both cases, shifting the brightest observations to small residuals highly reduces the weighted RMS in the classical case. At the end, the high reduced χ^2 values indicate that we may overfit quantities such as the phase. The more accurate error bars of the Bayesian algorithm prevent overfitting the weighted RMS value as there are no such bright observations to reduce the weighted RMS systematically by optimizing for these ToAs.
Comparing the change of RMS and actual residuals' position for different ephemerides shows that indeed the classical ToAs are scattered around further than their error bars whereas the Bayesian ToAs are moved within their believed accuracy. For ephemerides representing different states of knowledge, the later is the expected behaviour. New knowledge should in general not change drastically what we assumed beforehand but lead to a more accurate description within the statistical bounds of a data set containing less knowledge.
We argue that these desirable properties of the Bayesian algorithm have not been produced by scaling the error bars in a trivial way. Fig. 4.15 shows a histogram of the ToA errors reported by both algorithm. It is clear that the Bayesian timing has a different distribution of error estimates that cannot be explained by linear transformation. There are only two ToAs of bright observation that fall into the main range of the classical algorithm and the ToAs seem to be grouped differently. The overall shape of the distribution shows similarities but cannot be scaled in any obvious way. The error bars also do not scale trivially with brightness of the observation. When removing the brightest observation the classical timing's reduced χ^2 values drop to about seven to twelve. Removing the same observation for the Bayesian analysis does change the reduced χ^2 by less than ten percent.
We conclude that while the absolute residuals turn out to be less accurate for the Bayesian case, the timing result seems to be more stable to small changes of ephemeris or binning of data points. The method shows large improvements in χ^2 value outperforming the classical error estimation by a factor of about seven to twenty.

4.3.7 Moding

We examined the algorithm's timing behaviour on a moding pulsar. We used a dataset of pulsar B0329+54 observed in Effelsberg at 21 cm consisting of 5000 single pulse observations with 1024 bins each. Using the same formula (4.35) as for the nulling anlysis, the alorithm decides if a single pulse or a subset of pulses was emitted in a certain pulsar mode described by statistical template m_i (consisting of the same set of parameters like T_i). We auto-generated a moding analysis by generating a statistical template T over all data. Then we analysed the probability of each given single pulse to appear, given the template T describes the radiation correctly.

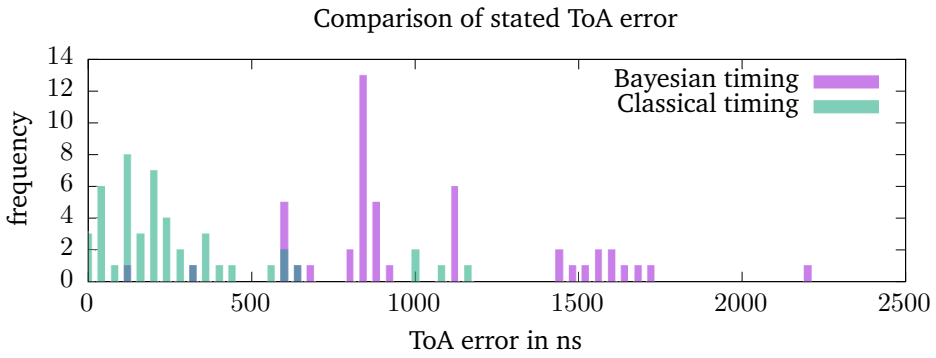


Figure 4.15: Comparison of the distribution of ToA errors as stated by both algorithms. Notice that the Bayesian algorithm's output is not a linear scaling of the classical output.

	Classical timing				Bayesian timing			
frequency	rms	Wrms	χ^2	red. χ^2	rms	Wrms	χ^2	red. χ^2
1360MHz	310ns	228ns	349	12	378ns	301ns	8.9	0.30
2650MHz	1445ns	1422ns	440	28	1542ns	1189ns	23.6	1.5
Both	895ns	337ns	789	17	967ns	533ns	32.5	0.71

Table 4.1: Comparison of timing residuals of observations taken at 1350MHz and model accuracy as reported by TEMPO2 for classical and Bayesian ToAs.

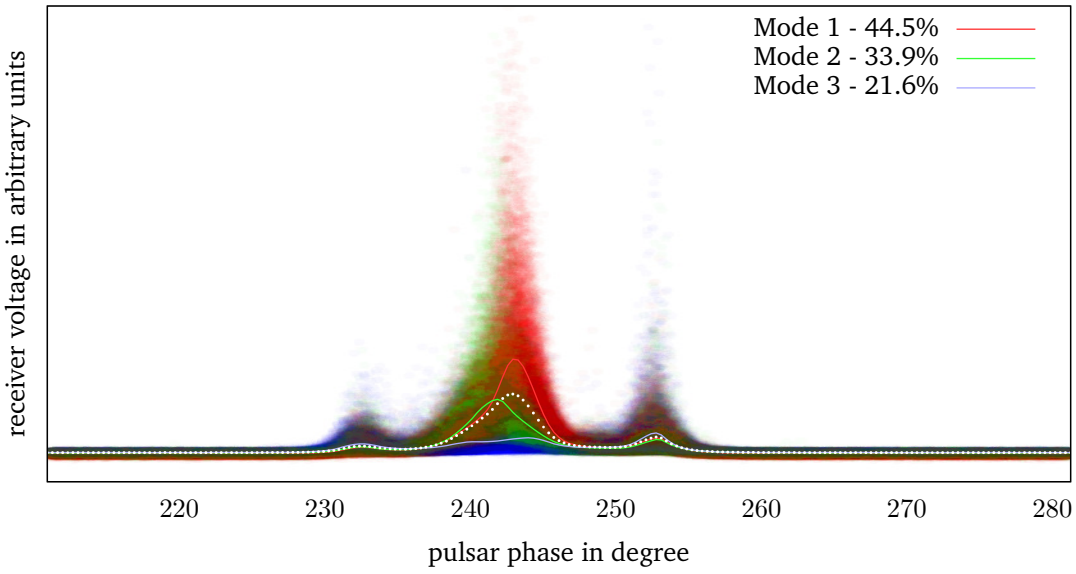


Figure 4.16: We overlayed the average profile curves of different modes over a scatter plot of the single pulses coloured according to the probability of belonging to mode 1-3. The dotted shape amounts to an average over all pulses.

Basically this means evaluating the joint probability for the phase (4.27) and amplitude (by integrating (4.11) over s). We ordered this list of probabilities descending and initially divided it into n_{modes} sections of equal length, where we assumed a certain n_{modes} to be the correct number.
In principle, the number of statistically distinct modes is also subject to uncertainty. One may also calculate the probability, that a certain n_{modes} holds. This could in principle further improve the results derived in the following but is of secondary interest for a first analysis of the behaviour of the algorithm.
Given this initial assignment of the single pulses to modes, we started an iteration procedure. In each step, a new set of statistical templates according to the weights of the previous step are calculated: $\{T_1^{(0)}, ..., T_{n_{modes}}^{(0)}\}$. These were then used as models in eq. (4.35) to assign the probability for each individual pulse to belong to a certain mode. These probabilities become the new weights for that pulse. We experimented with this iterating procedure and found about five steps to be sufficient that the difference from step to step is negligible.
The algorithm allows an unambiguous assignment of single pulses to a certain statistical template except for a very small number of pulses. The statistical nature of single pulses makes it difficult to grasp what classification is happening. Thus we decided to make a scatter plot (Fig. 4.16) of 5000 single pulses assigned to three modes. The single pulses were blurred to simulate a density kernel and plotted with an opacity of 1% in the colour of the probability to belong to a certain mode. For orientation, we overlayed the average pulse shapes each mode would generate if it was the sole mode observed along with an average over all pulses, drawn in dotted white.
The term moding usually refers to the appearance of few distinguishable shapes of the integrated pulse profile. As a single pulse profile is very different from its neighbouring pulses, it was very common to assign different modes to consecutive pulses. Thus the question arises, whether to still call this behaviour moding or not. On one hand, we could try to further develop the algorithm to make mode switching on such small timescales very unlikely. Then the algorithm would pick up “modes” in a more classical sense. On the other hand, the integrated profiles of these “sub”-modes are quite distinct and certainly their ratio has an impact on the average profile shape over a few minutes integrating and thus also on timing (as we will examine below). Understanding the relationship between these sub-modes and the astonishingly stable moding behaviour over a larger integration time could lead to a deeper understanding of the conditions in the pulsar magnetosphere.
These submodes also seem to be correlated over the whole profile. For example, mode three in the figure is the only one having an earlier rise of intensity at 230 degrees and additionally a very low slope in the middle of the profile.
Taking these different correlations into account can also archive a more accurate timing and make the timing stable against different integration times and noise. Fig. 4.17 shows two examples of how a moding template on single pulse level can improve detecting the phaseshift of a few tens to hundreds of pulses. Statistical templates were generated from 5000 pulses and later used to time simulated epochs consisting of consecutive subsets of 25 in the upper and 50 single pulses in the lower panel, taken from the same 5000 pulses. The most likely value for the shift of each set was binned to generate histograms of the observed frequency. Knowing that there was no subpulse drift over the 5000 pulses as a whole, we expect the phase to be measured as zero for every subset. This is indeed the case, if we assume the pulsar to radiate in different modes (depicted in green). When run over subsets of 25 pulses, the non-moding template (depicted in red) fails to detect the correct shift by repeatedly mistaking one peak for the other. Unfortunately the few outliers which would have identified the right peak are scattered around zero with a systematic bias. We suspect this bias to be caused by the single statistical

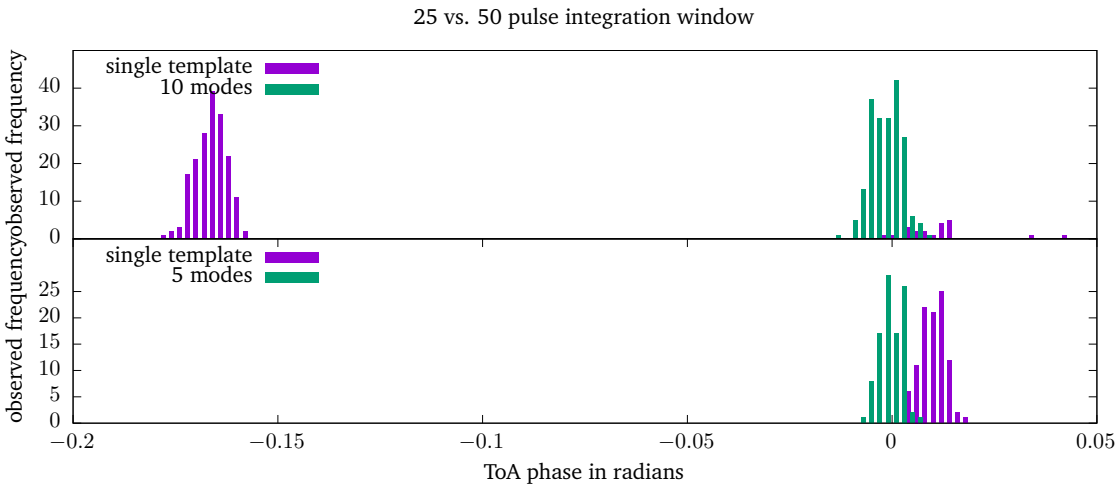


Figure 4.17: This figure compares the phase detection performance of a single or a moding statistical template on 25 respectively 50 pulse integrations.

template trying to cover two or more very distinct modes of radiation. Doubling the number of pulses in a subset fixes the problem of not matching the right peaks. However, there is still a bias and the results have a larger variance than the moding template. The moding template detects the correct pulse phase with satisfying accuracy while showing a lower variance. This can be understood having a look at the generated modes. The maximum value of the second peak for every single mode is further left for pulses which are less intense. In the average picture over all pulses this correlation cannot be accounted for. When looking at subsets of pulses however, the lower probability of reaching a high intensity mode shifts the peak to the left. Consequently, the reference template is detected to be shifted to the right. The ansatz using one statistical template reduces this inaccuracy to about half a degree, which is still a low value in the light that e.g. the peaks of the first and second mode, if we assume three modes (see Fig.4.16), are separated by 1.5 degrees and the average taken from 50 pulses is still very noisy. Using a statistical moding template, the variance in variability exceeds the inaccuracy introduced by integrating a smaller subset than the one the template was generated from.
We conclude that assuming even a low number of modes to be present can significantly improve the timing results both in variance and systematic error.

4.4 Conclusions

We developed and evaluated a log-normal Bayesian model for single pulse analysis of pulsars. The algorithm described is able to reproduce the results of classical averaging procedures when considering non-fluctuating templates up to the approximations taken. The method is surprisingly versatile in situations of weak signal or a fluctuating template shape, since it is shown to be more robust against fluctuations. Thus it may enable one to examine fast changes in the radiation profile or distinguish different modes of radiation, such as the nulling analysis, an example for model discrimination, has shown.
Secondary parameters' reconstruction may be implemented easily by calculating their imprint on the data and then evaluating the joint probability using the model derived, as was demon-

strated with phase drift on simulated epochs. The parameter studies derived in such a way profit from additional statistical information incorporated in the template automatically since only the statistically significant part of the data will have an imprint on the probability distribution of the parameters under examination.
The benefit of rather understanding which parameter values are compatible with the data gathered than trying to reconstruct one possible signal out of the dataset might be worth the effort. This may give statistical access to hitherto poorly examined parameters and topics such as small time-scale variations of the radiation process and magnetosphere by providing strong and reasonable reconstructions. It may be an appropriate choice when a classical estimator leaves us with a definite but perhaps insignificant answer.
A first test to generate ToAs from real data showed that the algorithm developed can reduce χ^2 values while reaching the accuracy of classical methods. These values can further be improved as we did not use single pulses but few second integrations. It gave correct ToAs even though the statistical templates used were generated from the same data that was to time. The classical way of template generation is known to affect the results of timing if the same data is used for template generation [Hotan et al., 2005].
If one intends to use the method on routine basis as an alternative to the classical one, the data volume to keep for logging and generating reproducible output increases drastically. One would have to find a way of reducing the data without loosing the possibility to reprocess it or re-validate it.
Data compression and performance might be improved at the same time for Bayesian methods, as van Haasteren [2013] already points out. Computational feasibility sets a limit to the potential use of Bayesian analysis, but there exist ready to use methods for speeding up Bayesian calculations, such as Taylor et al. [2012] or the Metropolis-algorithm [Metropolis et al., 1953]. The formulae presented may also be sped up by analysing them in perturbative manner. Enßlin et al. [2009] and references therein develop an information field theory, taking the information Hamiltonian $H = -\log \mathcal{P}(\cdots)$ as a starting point. This allows for approximating the numerical integrations involved analytically and breaking their influence down to computationally simpler matrix arithmetic.
Furthermore consolidating different data channels, like radio data at different frequencies, is easily possible in a Bayesian framework.
Once the computational complexity is dealt with, the Bayesian method might provide a way to reduce observation time for low signal to noise pulsars and give a handle on the statistical uncertainties involved in template generation. It gives access to a broad range of secondary applications such as sophisticated methods of interference detection or testing pulsars long-term behaviour for template changes or short- and long-term moding behaviour affecting ToA-analyses.
Hence the basic analysis carried out has shown that Bayesian reconstruction of pulsar templates is not only feasible, but also forms a more complete picture on the pulsars template shape. The additional data turns out to be very valuable for subsequent analysis of secondary parameters mitigating the effects of insignificant fluctuations. Benefits known to arise by inspection of the stochastic behaviour in the analysis will be automatically propagated by implementing a measurement using the proposed Bayesian template.
The C++ implementation of the algorithm ran about a day on 25.8 hours of $10\mathrm{s}$ preintegrated data on a powerful desktop computer[6] to generate the TOAs of Table 4.1. Taking such efforts today to improve mainly the χ^2-values may seem like exaggeration. However, with the next generation telescopes like the SKA, we will face a high S/N ratio in single pulse or short observations. We will be entering a regime where we cannot afford losing statistical data any more.

[6] 32 GB RAM / Intel Hexacore i7-3960X CPU @ 3.30GHz

Improving timing performance in this regime will crucially depend on understanding the cause and controlling the impact of fluctuations in the signal. Waiting for the fluctuations and noise to average out may not be necessary or even inferior to a method like the one suggested.

4.5 Acknowledgements

The authors would like to thank Torsten Enßlin for valuable input on the measurement model. We thank Axel Jessner to provide a data set of B1133+16 conveniently preprocessed to access in the code and valuable discussion. We thank Kejia Lee for the extensive comments shaping and sharping the argumentative line of the paper. We thank Niels Oppermann for extended discussion on the mathematical difficulties associated with the evaluation of the posterior pdf. We thank Kuo Liu for valuable discussion about the results and ToAs. We thank Ben Hoyle for a quick review of the introduction. We would like to thank Aidan Hotan for valuable comments on the overall argumentative line and pointing out important aspects of the results.

Appendix to paper

4.6 Calculations

4.6.1 Integration over g

Starting from eq. (4.10):

$$
\begin{aligned}
H_B[A] =& \frac{1}{2}\left[(\sqrt{\mathbf{d}} - \widehat{\exp[\mathbf{f}]}\cdot\mathbf{g}\sigma_\mathrm{n})^\dagger \frac{1}{\sigma_\mathrm{n}^2}(\sqrt{\mathbf{d}} - \widehat{\exp[\mathbf{f}]}\cdot\mathbf{g}\sigma_\mathrm{n}) + \mathbf{f}^\dagger\mathbf{F}^{-1}\mathbf{f} + \mathbf{g}^\dagger\mathbf{G}^{-1}\mathbf{g}\right] \\
& \underbrace{-\log\mathcal{P}(\tau) + \frac{1}{2}\log(|2\pi\sigma_\mathrm{n}^2|^{N_\mathrm{tot}}|2\pi\mathbf{F}||2\pi\mathbf{G}||\mathbf{d}|)}_{=:\blacktriangle - .5\cdot\mathbf{f}^\dagger\mathbf{F}^{-1}\mathbf{f}} \\
=& \frac{1}{2}\left[\sqrt{\mathbf{d}}^\dagger\frac{1}{\sigma_\mathrm{n}^2}\sqrt{\mathbf{d}} - 2\frac{(\sqrt{\mathbf{d}}\widehat{\exp[\mathbf{f}]})^\dagger}{\sigma_\mathrm{n}}\mathbf{g} + \mathbf{g}^\dagger\widehat{\exp[\mathbf{f}]}^\dagger\widehat{\exp[\mathbf{f}]}\mathbf{g} + \mathbf{g}^\dagger\mathbf{G}^{-1}\mathbf{g}\right] + \blacktriangle = \\
=& \frac{1}{2}\left[\sqrt{\mathbf{d}}^\dagger\frac{1}{\sigma_\mathrm{n}^2}\sqrt{\mathbf{d}} + \mathbf{g}^\dagger\underbrace{\left(\mathbf{G}^{-1} + \widehat{\exp[\mathbf{f}]}^\dagger\widehat{\exp[\mathbf{f}]}\right)}_{=:\mathbf{D}_\mathbf{f}}\mathbf{g} - 2\frac{(\sqrt{\mathbf{d}}\widehat{\exp[\mathbf{f}]})^\dagger}{\sigma_\mathrm{n}}\mathbf{g}\right] + \blacktriangle = \\
=& \frac{1}{2}\left[\sqrt{\mathbf{d}}^\dagger\frac{1}{\sigma_\mathrm{n}^2}\sqrt{\mathbf{d}} + (\mathbf{g} - ...)^\dagger D_\mathbf{f}(\mathbf{g} - ...) - (\frac{\sqrt{\mathbf{d}}\widehat{\exp[\mathbf{f}]}}{\sigma_\mathrm{n}})^\dagger\mathbf{D}_\mathbf{f}^{-1}(\frac{\sqrt{\mathbf{d}}\widehat{\exp[\mathbf{f}]}}{\sigma_\mathrm{n}})\right] + \blacktriangle
\end{aligned}
$$

Integrating over g leads to:

$$\begin{aligned} H_B[A \setminus \mathbf{g}] =& \frac{1}{2}\left[\frac{\sqrt{\mathbf{d}}}{\sigma_{\mathrm{n}}^2}^{\dagger}\left(1-\widehat{\exp[\mathbf{f}]}^{\dagger}\mathbf{D}_{\mathbf{f}}^{-1}\widehat{\exp[\mathbf{f}]}\right)\frac{\sqrt{\mathbf{d}}}{\sigma_{\mathrm{n}}}+\log|2\pi\mathbf{D}_{\mathbf{f}}^{-1}|\right]+\blacktriangle = \\ =& \frac{1}{2}\left[\sqrt{\mathbf{d}}^{\dagger}\frac{1}{\sigma_{\mathrm{n}}^2}\left(1-\widetilde{\mathbf{D}}_{\mathbf{f}}^{-1}\right)\sqrt{\mathbf{d}}+\log|2\pi\mathbf{D}_{\mathbf{f}}^{-1}|\right]+\blacktriangle = \\ =& \frac{1}{2}\left[\sqrt{\mathbf{d}}^{\dagger}\frac{1}{\sigma_{\mathrm{n}}^2}\left(1-\widetilde{\mathbf{D}}_{\mathbf{f}}^{-1}\right)\sqrt{\mathbf{d}}+\mathbf{f}^{\dagger}\mathbf{F}^{-1}\mathbf{f}+\right. \\ & \left.+\log(|2\pi\mathbf{D}_{\mathbf{f}}^{-1}||2\pi\sigma_{\mathrm{n}}^2|^{N_{\mathrm{tot}}}|2\pi\mathbf{F}||2\pi\mathbf{G}||\mathbf{d}|)\right]-\log\mathcal{P}(\tau) \end{aligned} \tag{4.36}$$

where $\widetilde{\mathbf{D}}_{\mathbf{f}} = \widehat{\exp[-\mathbf{f}]}\mathbf{G}^{-1}\widehat{\exp[-\mathbf{f}]}+1$ is always greater than 1 leading to a broadening of the likelihood in case of bad signal noise ratio. We may reformulate the matrix in the data-dependent term as

$$\begin{aligned} 1-\widetilde{\mathbf{D}}_{\mathbf{f}}^{-1} &= [\widetilde{\mathbf{D}}_{\mathbf{f}}^{-1}][\widetilde{\mathbf{D}}_{f}-1] = [\widehat{\exp[-\mathbf{f}]}\mathbf{G}^{-1}\widehat{\exp[-\mathbf{f}]}+1]^{-1}\widehat{\exp[-\mathbf{f}]}\mathbf{G}^{-1}\widehat{\exp[-\mathbf{f}]} \\ &= [1+\widehat{\exp[\mathbf{f}]}\mathbf{G}\widehat{\exp[\mathbf{f}]}]^{-1} \end{aligned} \tag{4.37}$$

which leads us to the final form

$$\begin{aligned} H_B[A] = \frac{1}{2}\left[\sqrt{\mathbf{d}}^{\dagger}\frac{1}{\sigma_{\mathrm{n}}^2}[1+\widehat{\exp[\mathbf{f}]}\mathbf{G}\widehat{\exp[\mathbf{f}]}]^{-1}\sqrt{\mathbf{d}}+\right. \\ \left.+\mathbf{f}^{\dagger}\mathbf{F}^{-1}\mathbf{f}+\log(|2\pi\mathbf{D}_{\mathbf{f}}^{-1}||2\pi\sigma_{\mathrm{n}}^2|^{N_{\mathrm{tot}}}|2\pi\mathbf{F}||2\pi\mathbf{G}||\mathbf{d}|)\right]-\log\mathcal{P}(\tau) \end{aligned} \tag{4.11}$$

4.6.2 Discrete Fourier coefficients of a detuned signal

For the signal consisting of a single Fourier coefficient with frequency $\omega' = \frac{2\pi n}{T} = \frac{2\pi na}{\tau}$ the measured coefficient over a period $[K\tau - \frac{\tau}{2} : K\tau + \frac{\tau}{2}]$ of pulse K is readily calculated as

$$\begin{aligned} \widetilde{d}_{k,K} &= \int_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} \mathrm{d}t \quad s_n \exp\left[i(\omega'_m-\omega_k)t\right] = \\ &= s_n\left[\frac{\exp\left[2\pi i\frac{(na-k)}{\tau}t\right]}{2\pi i(na-k)}\right]_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} = \\ &= s_n\left[\frac{\exp\left[2\pi ik\frac{(\frac{n}{k}a-1)}{\tau}t\right]}{2\pi ki(\frac{n}{k}a-1)}\right]_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} = \\ &= s_n \exp\left[2\pi ik(\frac{n}{k}a-1)K\right]\frac{\sin\pi k(\frac{n}{k}a-1)}{\pi k(\frac{n}{k}a-1)} \\ &= s_n \exp\left[2\pi ik(\frac{n}{k}a-1)K\right]\cdot\mathrm{sinc}[\pi k(\frac{n}{k}a-1)] \end{aligned} \tag{4.18}$$

For the full spectrum of a real signal, where $s_n = s^*_{-n}$ we get

$$
\begin{aligned}
d_{k,K} &= \int_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} \mathrm{d}t \quad \sum_{n=-\infty}^{\infty} s_n \exp\left[i(\omega'_m - \omega_k)t\right] = \\
&= \left[\sum_{n=-\infty}^{\infty} s_n \frac{\exp\left[2\pi i\frac{(na-k)}{\tau}t\right]}{2\pi i(na-k)}\right]_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} = \\
&= \sum_{n=1}^{\infty}\left[s_n \frac{\exp\left[2\pi i\frac{(na-k)}{\tau}t\right]}{2\pi i(na-k)} + s_n^* \frac{\exp\left[2\pi i\frac{(-na-k)}{\tau}t\right]}{2\pi i(-na-k)}\right]_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} = \\
&= \sum_{n=1}^{\infty} a_n \left[\frac{1}{2\pi i(k^2-(na)^2)}\left((k+na)(\cos\Phi_{s_n} + i\sin\Phi_{s_n})\exp\left[2\pi i\frac{(na-k)}{\tau}t\right]\right.\right. \\
&\qquad \left.\left. +(k-na)(\cos\Phi_{s_n} - i\sin\Phi_{s_n})\exp\left[2\pi i\frac{(-na-k)}{\tau}t\right]\right)\right]_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} = \\
&= \sum_{n=1}^{\infty} a_n \left[\exp\left[2\pi i\frac{-k}{\tau}t\right]\left(\cos\Phi_{s_n}\frac{-ik\cos(\frac{2\pi na}{\tau}t) + na\sin(\frac{2\pi na}{\tau}t)}{\pi(k^2-(na)^2)}\right.\right. \\
&\qquad \left.\left. +\sin\Phi_{s_n}\frac{ik\sin(\frac{2\pi na}{\tau}t) + na\cos(\frac{2\pi na}{\tau}t)}{\pi(k^2-(na)^2)}\right)\right]_{K\tau-\frac{\tau}{2}}^{K\tau+\frac{\tau}{2}} = \\
&= \pm\sum_{n=1}^{\infty} a_n \left[\frac{2\sin(\pi na)}{\pi(k^2-(na)^2)}\left(\cos\Phi_{s_n}(ik\sin(2\pi naK) + na\cos(2\pi naK))\right.\right. \\
&\qquad \left.\left. +\sin\Phi_{s_n}(-na\sin(2\pi naK) + ik\cos(2\pi naK))\right)\right] = \\
&= \pm\sum_{n=1}^{\infty} a_n \left[\frac{2\sin(\pi na)}{\pi(k^2-(na)^2)}\left(i\sin(2\pi naK)(\cos\Phi_{s_n}k + i\sin\Phi_{s_n}na)\right.\right. \\
&\qquad \left.\left. +\cos(2\pi naK)(\cos\Phi_{s_n}na + i\sin\Phi_{s_n}k)\right)\right] = \\
&= \sum_{n=1}^{\infty} a_n \left[\frac{2\mathrm{sinc}[\pi(na-k)]}{k+na}\left(i\sin(2\pi naK)(\cos\Phi_{s_n}k + i\sin\Phi_{s_n}na)\right.\right. \\
&\qquad \left.\left. +\cos(2\pi naK)(\cos\Phi_{s_n}na + i\sin\Phi_{s_n}k)\right)\right] =
\end{aligned}
$$

where we now introduce $\overline{k} = \frac{an+k}{2}, d = \overline{k} - k$ and simplify

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} a_n \left[\frac{\mathrm{sinc}[\pi(na-k)]}{\overline{k}} \cdot \big(i \sin(2\pi naK)(\cos\Phi_{s_n}(\overline{k}-d) + i\sin\Phi_{s_n}(\overline{k}+d)) \right. \\
&\left. \cdot \; + \cos(2\pi naK)(\cos\Phi_{s_n}(\overline{k}+d) + i\sin\Phi_{s_n}(\overline{k}-d)) \big) \right] = \\
&= \sum_{n=1}^{\infty} a_n \left[\frac{\mathrm{sinc}[\pi(na-k)]}{\overline{k}} \cdot \big(\overline{k}(i\sin(2\pi naK) + \cos(2\pi naK))(\cos\Phi_{s_n} + i\sin\Phi_{s_n}) \right. \\
&\left. \quad \cdot \; + d(-i\sin(2\pi naK) + \cos(2\pi naK))(\cos\Phi_{s_n} - i\sin\Phi_{s_n}) \big) \right] = \\
&= \sum_{n=1}^{\infty} a_n \mathrm{sinc}[\pi(na-k)] \left[\underbrace{\exp\left[i(2\pi naK + \Phi_{s_n})\right]}_{I} + \underbrace{\frac{d}{\overline{k}} \exp\left[-i(2\pi naK + \Phi_{s_n})\right]}_{II} \right] =: \sum_{n=1}^{\infty} A_{kn} a_n
\end{aligned}
\tag{4.20}
$$

where we assumed b_0 to be zero and used $s_n =: a_n(i\sin\Phi_{s_n} + \cos\Phi_{s_n}), a_n \mathrm{real}$.

4.7 Some considerations on the evaluation of the receiver Posterior

Having rigorously marginalized out the stationary process g, we seek to maximize the posterior probability density function with respect to the template parameters. When projecting this equation into Fourier space, G is a diagonal matrix and the multiplication with $\hat{f}$ transforms to a convolution that breaks this diagonality, leading to a coupling of the Fourier coefficients. While in the diagonal case the pdf completely factorizes in separate Fourier coefficients, which are independently and quickly integrated, there is only one N_{bin}-dimensional integral in the exact case. For typical N_{bin}s of 1024 the computational demand is simply too high. Thus we decided for imposing diagonality in Fourier space, which discards the non-diagonal terms of f leading to a solvable integral. However, this discards phase information as now f has also the properties of a stationary process. Consequently we have derived an amplitude-only model. An Amplitude model alone does not contain valuable information about TOAs since the amplitude is invariant to time shifts. Therefore, we impose a model on the phase of the signal based on wrapped gaussians and being evaluated phenomenologically. This model uses the amplitude information to estimate the phase error of an observation at hand. The phenomenological way was chosen as the exact calculations for quadrature of the stochastic signal mix the phases of different Fourier channels in a highly non-trivial fashion again leading to an unwanted coupling of phases present only at the single pulse level.
This ansatz tries to find a compromise between numerical feasibility and accuracy. For the case of real measurements containing additional noise and fluctuations, we expected to grasp the uncertainties left with acceptable precision and lead to acceptable χ^2-values at the end. However losing the mathematical precision, we expect not to reach the accuracy of the classical timing codes when it comes to evaluating exact, nonfluctuating test data sets. Indeed our round mean square errors turned out to be 25% worse in these cases while the reached χ^2 was comparable accurate.
The question surely arises why we did not evaluate the pdf in time domain. In time domain, the

main operator acting on the data is the inverse of a diagonal matrix ($\hat{f}$) plus a Toeplitz-matrix G [7] that can be further reduced to a circular Toeplitz-Matrix by the symmetry in $\hat{f}$. There exist fast ways (up to $\mathcal{O}(N \log N)$) of calculating the operators scalar product with the data [Ng and Pan, 2010] however this does not reduce the dimensionality of the integration when it comes to varying the template parameters of interest. Unfortunately, expanding the operator linearly (to develop a mean field theory for the signal field, which should lead to a very good estimate and first order corrections for the correlations present in the problem) does not help in this case, as the derivatives refer to specific entries of the inverse matrix. Calculating the inverse matrix itself scales $\mathcal{O}(N^3)$ or best $\mathcal{O}(N^2)$ in our case, and as we need N entries for the calculation of the mean field gradient, the computational demand is again unacceptable. A class of transformations on f that leave the determinant of the operator invariant, combined with a pdf on the invariants of the transformation, could make exact evaluation of the derived equations possible. In lack of such a tool we decided to approximate the problem in Fourier space.

4.8 Iterative solver for fields

Finding the maximum a-posteriori (MAP) probability set equals finding a global maximum for $\overline{\mathbf{f}}$ and $\sigma_{\mathbf{f}}$ both at the same time. Equations (4.16) and (4.17) explicate optimization w.r.t. one parameter only in the context of correct prior knowledge. Furthermore, we want to find the optimal solution w.r.t. both parameters. We may resolve these issues by assuming that there is only one maximum (which might not be the case as there might be several modes of radiation in the system etc.) and rely on a suitable iteration scheme. We implemented a method of steepest ascend. The Hamiltonian may be seen as a potential to be broken up in parts dependent on $\overline{f}_k$ and σ_k for every Fourier coefficient. For the sake of shortness let us call this part $\Phi(\overline{\mathbf{f}}(t), \sigma(t))$. We now try to find the path of the steepest ascend, parametrized by t: $\gamma_t = (\overline{\mathbf{f}}, \sigma)$. Infinitesimally, isolines may be found using $\mathrm{d}\Phi(\gamma_t) = 0$. Working this out for our Hamiltonian leads to

$$0 = \langle \frac{(\mathbf{f} - \overline{\mathbf{f}})^2}{\sigma^3} - \frac{1}{\sigma} \rangle \frac{\partial \sigma}{\partial t} + \langle \frac{\mathbf{f} - \overline{\mathbf{f}}}{\sigma^2} \rangle \frac{\partial \overline{\mathbf{f}}}{\partial t} \tag{4.38}$$

Consequently, the direction of the steepest ascend/descent follows the differential equation

$$0 = -\langle \frac{\mathbf{f} - \overline{\mathbf{f}}}{\sigma^2} \rangle \frac{\partial \sigma}{\partial t} + \left(\frac{\langle (\mathbf{f} - \overline{\mathbf{f}})^2 \rangle}{\sigma^3} - \frac{1}{\sigma} \right) \frac{\partial \overline{\mathbf{f}}}{\partial t} \tag{4.39}$$

where orthogonality is easily shown. Since for most cases occurring setting $\overline{\mathbf{f}}$ according to (4.16) is an acceptable method while a reasonable update procedure of σ prior to optimizing it was missing, we update σ when updating $\overline{\mathbf{f}}$ by the method, but not vice versa. Integrating (4.39) is approximated by

$$\Delta\sigma = \frac{\partial \sigma}{\partial \overline{\mathbf{f}}} \Delta \overline{\mathbf{f}} = \left(\frac{\langle (\mathbf{f} - \overline{\mathbf{f}})^2 \rangle}{\sigma} - \sigma \right) \frac{\Delta \overline{\mathbf{f}}}{\langle \mathbf{f} - \overline{\mathbf{f}} \rangle} \tag{4.40}$$

$$\Rightarrow \sigma_{new} = \frac{\langle (\mathbf{f} - \overline{\mathbf{f}}_{new})^2 \rangle + \langle (\mathbf{f} - \overline{\mathbf{f}}_{old})^2 \rangle}{2\sigma_{old}} \tag{4.41}$$

[7] a matrix whose entries depend only on the distance to the diagonal

and determining $\bar{\mathbf{f}}_{new}$ as in (4.16). As is evident by derivation, this method is generic for a prior of the form of a Gaussian, however could be refined by also updating the $\bar{\mathbf{f}}$ value by similar considerations.

CHAPTER 5

Paper : Rambrain
a library for virtually extending physical memory

Imgrund, M. and Arth, A.
This chapter has been submitted in slightly modified form to Astronomy and Computing

Abstract: We introduce Rambrain, a user space C++ library that manages memory consumption of data-intense applications. Using Rambrain one can overcommit memory beyond the size of physical memory present in the system. While there exist other more advanced techniques to solve this problem, Rambrain can save development time by providing a fast, general and easy-to-use solution. Rambrain takes care of temporarily swapping out data to disk and can handle multiples of the physical memory size present. Rambrain is thread-safe, OpenMP and MPI compatible and supports asynchronous IO. The library is designed to require minimal changes to existing programs.

5.1 Introduction

Facing large amounts of data, be it simulations or observation results, many astrophysicists have become part-time software engineers. As the primary target of their work focuses on producing astrophysical results, developing e.g. code to analyse data, is an inevitable obstacle on their way to the actual goal. In the case of the authors this goal is either to analyse extensive data sets of pulsar timing information [based on Imgrund et al., 2015] or to post-process large snapshots of cosmological simulations (see Arth et al. in prep.). While typical software-engineering amounts to serialising given tasks to be executed as quickly as possible, many everyday codes evaluating data or simulation results are written to be run only a few times. In this light, the primary focus of an astrophysicist often lies on saving development time and not execution time.
One of the time consuming tasks is writing code that needs to deal with large data sets. When developing applications which use large amounts of main memory, a single larger dataset may suffice for the system to run out of memory. The typical ad hoc solution to this is finding a machine with more main memory. It is obvious that this solution is only temporary when facing growing amounts of data. The other extreme is to code memory management functions specifically in an optimised way for the problem at hand, so called “out-of-core computing”. This, however, is very time-consuming.
Therefore we introduce Rambrain, a library that facilitates quick development of applications in need of large main memory. It is built to easily integrate with existing C++ code on Linux

and helps applications to swap out temporarily unneeded data to transparently access multiples of the actual physical memory available on the system.
While there may exist other problem specific solutions that have a slightly better performance, we argue that in most situations the flexibility of a fast, reliable and out-of-the-box solution is preferred to a few percent performance gain. Before introducing Rambrain, let us quickly review other solutions to the problem at hand and discuss when these are to be preferred and in what cases they are to be discarded. While there may exist other problem specific solutions that have a slightly better performance, we argue that in most situations the flexibility of a fast, reliable and out of the box solution is preferred to a few percent performance gain. Before introducing Rambrain, let us quickly review other solutions to the problem at hand and discuss when these are to be preferred and in what cases they are to be discarded.

5.2 Common strategies to avoid out-of-memory errors

The most basic strategy to still run an application in a situation of scarce free memory is using native system swapping. Modern operating systems like Linux manage association of physical memory to various processes running at a given moment. As an application developer, you are presented a more or less consecutive virtual memory address space. It is in general not clear whether a chunk of virtual memory, a so called “page”, is residing in a physical main memory location, called a “frame”, at a given time or not. This layer of abstraction facilitates assignment of memory to a process, so that the system can overcommit physical memory and reassign virtual pages to physical frames, when desired. When free frames become scarce, the system writes out currently unused pages to secondary storage (such as hard disks) in order to free frames. When a process tries to access a non-resident page, a page fault is triggered and the page is read in from secondary storage by the memory manager of the system [Ligh et al., 2014, p.20] and if necessary, according frames are freed by writing the occupying pages out beforehand. While this process is efficient under normal operation, the system typically slows down to being unusable when actively consuming nearly all physical memory. Especially when multiple processes compete for the remaining space (a typical situation for a developer working and debugging), the computer is virtually unusable until the memory-intense calculation has finished.
This swapping mechanism is also limited by the available swap space on the secondary storage. While adding more swap space with the system’s on-board mechanisms[1] is possible, it needs super user privileges and reserves the whole swap size on the disk even if it is not used completely. Furthermore, it aggravates the situation when multiple processes are competing for memory, as more and more parts of other programs can be swapped out and need to be swapped in again in order to continue execution.
Using system swapping as a mechanism for overcommitting main memory can also provoke the action of the so called “Out-Of-Memory Killer (OOM-Killer)". As available memory becomes sparse, the system tries to keep most processes running. In order to free memory for other processes, the OOM-Killer will kill one or more processes by assigning a score correlated with importance, memory consumption, execution and idle times of the candidate process. The OOM-Killer thus can abort simulation or analysis at the very last step and protections against are hard to find [see e.g. Rodrigues, 2009]. The OOM-killer can by now be controlled a bit finer via the */proc* file system, but shutting it off for a certain process needs administrator privileges. However, one has to keep in mind that even if one can force the own application to stay alive, the OOM-killer can simply shut down system processes which may trigger secondary effects on

[1] Using the system tools *mkswap/swapon* as root.

the target process.
A more controllable user-space solution is desirable, such as the memory mapping system calls combined with moderate sized swap files on the secondary storage. Memory mapping techniques are fast because they use the same paging and copy mechanisms such as system swapping, but are subject to stronger limitations than letting the system handle the paging itself.[2] The consecutive logical address space that is handed over to the process has to be managed by the user. This means that the user has to take care of allocating multiple data structures on top of the space, a mechanism that the new/delete operators deal with in C++, normally. While handling for example a vector of fixed size structures in a memory map is simple, allocating objects of different sizes will be highly non-trivial. As the system is responsible for writing out the memory mapped regions to the file on secondary storage, efficient interaction with the kernel when changing the memory-mapped region is challenging when trying to optimise this process for performance. This renders such a technique possible, but complicates robust implementation and favourable run time behaviour in highly dynamic situations.
Of course, there exist already such solutions, such as the STXXL [Dementiev et al., 2008] that facilitate out-of-core computation providing large standard containers in analogy to the Standard Template Library (STL). While this is a very useful idea, it has still some drawbacks imposed by it's specialised approach. Rambrain has built in class support for the full C++ standard in contrast to the limitation to POD-support of the STXXL. Rambrain provides direct access to pointers in memory and thus will pose no overhead over heap allocation once the pointers have been provided. Additionally, objects created with Rambrain can be used in association with normal STL-containers and will be swapped, too.
An alternative approach, using parallel virtual file systems is also imaginable [see for example Tang et al., 2004]. However, this kind of approach still leaves the programmer with the burden to write IO operations himself, even if they may be encapsulated e.g. as a function.
Furthermore, optimizing the data flow on this level comes near to developing an out-of-core algorithm for the problem at hand that takes control over all input and output operations manually. Introductory reviews of such algorithms can be found in Toledo [1999a], Vitter [2001]. Of course one can design a very clever way of handling input and output data to boost performance. This, however, opposes the goal to find a more generic solution that gives the developer moderate control over input and output flow while taking from him the burden of handling the input and output manually. Specialised solutions cover for example n-body codes [Salmon and Warren, 1997] or linear algebra calculations [Reiley and van de Geijn, 1999, Toledo, 1999b].
From the view of the application developer, the situation is very simple: When writing a program the developer knows what data he uses, what he will use next, and what is not needed for longer time. This information is always present directly in the source code. In the next section we will introduce the interface which communicates this information to the library.

5.3 Interfacing Rambrain

In order to manage the storage needs of a C++ application, we are faced with the problem of designing an interface to tell Rambrain, which data is to be managed and when it has to be present. In this chapter we introduce this interface built to require minimal changes of existing code while at the same time providing rich convenience features when possible.

[2] Both the number and size of memory maps are limited by the system.

Listing 5.1: Typical two dimensional field initialisation

```
double k_x=1.,k_y=1.;
unsigned int x_max=1024, y_max=1024;

double *arr[x_max];
for (int x=0;x<x_max;++x) //allocate rows
  arr[x] = new double[y_max];
for (int x=0;x<x_max;++x){ //initialize field
  double *line = arr[x];
  double xx = x / (double) x_max;
  for (int y=0;y<y_max;++y){
    double yy = y/(double) y_max;
    line[y] = sin((xx*k_x+yy*k_y));
  }
}
//do something and delete afterwards:
for (int x=0;x<x_max;++x)
  delete arr[x]; //deallocate lines
```

5.3.1 Basic usage

As a memory manager keeping track of data has some overhead on its own, it is only useful when the data managed is large. Rambrain can manage simple primitives, arrays, whole classes and also supports nesting of managed objects into managed classes. For a start, consider the code in listing 5.1 that is initialising a two dimensional plane wave field of data type double on heap memory. We allocate an array of pointers to the respective field rows in line 4, allocate the actual rows in line 6, and set up a plane wave over all field values in lines 7 to 14. Some calculations are executed prior to the deallocation of the rows in line 17.
If we assume now that `y_max` and `x_max` take large values, the allocated doubles will consume a non-negligible amount of RAM, passing a gigabyte at roughly 11600^2 elements. Thus, the developer would have to swap out elements if he seeks to avoid system-swapping to occur, to ensure that the program does not run out of physical memory. Manual implementation inserts many lines of code when allocating memory and around line 8. Alternatively, the user would write his own memory manager version calling functions to load and unload data. When several objects are needed at once, loading and unloading become the dominant part of the code.
Furthermore the additional lines start to obfuscate algorithmic code structure. The nested `for`-loops as well as the essential initialisation done will be difficult to spot. Minimal changes to this passage of code will allocate the arrays so that Rambrain is aware of them and dynamically loads and unloads the lines if needed, as can be seen in listing 5.2.
The overall structure is minimally changed. Up to adding line 8 we only wrap data objects. We introduce two template classes here, `managedPtr<>` and `adhereTo<>` to emplace Rambrain. When using Rambrain in a minimal way, these two classes will be the only ones referenced by the developer.
The first class, `managedPtr<>`, replaces allocation and deallocation by Rambrain wrappers. This replacement is necessary to hide away the pointer to the actual data in logical memory, as the element may or may not be present when the user dereferences that pointer.
Consequently, we need a way to give back access to the data. This is done by `adhereTo<>` which states its meaning in camel-case: This objects adheres to the data. While the respective

Listing 5.2: typical two dimensional field initialisation with Rambrain

```
double k_x=1.,k_y=1.;
unsigned int x_max=1024, y_max=1024;

managedPtr<double> *arr[x_max];
for (int x=0;x<x_max;++x) //allocate rows
  arr[x] = new managedPtr<double>(y_max);
for (int x=0;x<x_max;++x){ //initialize field
  adhereTo<double> glue(arr[x]);
  double *line = glue;
  double xx = x / (double) x_max;
  for (int y=0;y<y_max;++y){
    double yy = y/(double) y_max;
    line[y] = sin((xx*k_x+yy*k_y));
  }
}
//do something and delete afterwards:
for (int x=0;x<x_max;++x)
  delete arr[x]; //deallocate lines
```

`adhereTo<>` object exists according to scoping rules, it is guaranteed that the user can fetch a valid pointer to the data by assigning the `adhereTo<>` object to the pointer, as is done in line 9. In the following, we will also refer to this as “pulling the pointer".
The scoping relieves the user from the need to explicitly state that the data is no longer used for the moment. While the corresponding `adhereTo<>` object exists, the pointer to the data remains valid. When this “glue” to a `managedPtr<>` is deleted, for example by going out of scope, the object may be swapped out to disk in order to free space in physical memory for other objects, if needed. This already concludes what a developer needs to know about Rambrain to write his own code using the library in the most basic fashion.

5.3.2 Advanced usage

Currently, Rambrain is equipped with the following advanced features that give more detailed control or convenience. The line numbers given refer to the code examples in listing 5.3. The advanced features show that the interface is both minimalistic and powerful enough to facilitate development with Rambrain.

- **Allocation of simple datatypes.** The user may allocate a single object or multiple objects at once, passing an initial value. Also multidimensional arrays are supported, that will be collapsed to an array of `managedPtr<>`s of the size of the last dimension. (lines 1-4)
- **Class allocation.** Class objects may have nested `managedPtr<>`s which can be swapped out independently of the class object. Rambrain supports parametrized as well as default constructors. Destructors will be called in the correct sequence. Furthermore, the member hierarchy can be tracked. Finally, Rambrain will ensure correct deallocation of the object. As some or all parts of it may have been swapped out, this is a non-trivial task. The code supports array initialisation on classes, too. (lines 6-15)
- **Different kinds of loading stages.** The user may explicitly state whether to load objects immediately or delay actual loading until the first pointer is being pulled from the

Listing 5.3: Advanced features

```
managedPtr<double> a1; //single element
managedPtr<double> a2(5); //array of five elements
managedPtr<double> a3(5,1.); //five elements, all set to 1.
managedPtr<double,2> a1(5,5,0); //two dim., vals set to 0.

class B { public:
        B(); B(double &a, double &b);
        ~B();
        void someFunction();
        managedPtr<double> data; } //Class with ctors/dtor

managedPtr<B> b1; //single element, default constructor
managedPtr<B> b2(1) //single element, default constructor
managedPtr<B> b2(1,a,b); //single element, param. ctor
managedPtr<B> b2(5,a,b); //5 elements, parametrised ctor

adhereTo<double> glue1(a1);//Load right away
adhereTo<double> glue2(a2,false); // Load when used
const adhereTo<double> glue3(a3); // Access const

double *c1=glue1;
double *c2=glue2; //If not present, will be fetched here
const double *c3 = glue3;

//= adhereTo<double> a1_glue(a1); double* a1data = a1_glue;
ADHERETOLOC(double, a1, a1data);

void B::someFunction(){
        ADHERETO(double,data); //shadows member B::data
        data[0] = 42.; }

//MT: Do not fail if too much memory is requested:
managedMemory::defaultManager->setOutOfSwapIsFatal(false);
//MT: Avoid deadlock when needing multiple data at once:
double *c5,*c6;
adhereTo<double> c5_glue(a1),c6_glue(a2);
{LISTOFINGREDIENTS
        c5 = c5_glue;
        c6 = c6_glue; }
```

`adhereTo<>` object.
Rambrain can profit from `const`-accessing the data. In case of the object having been swapped out already, the swap file copy is not changed and reused and thus another write-out is not necessary. If the developer requests write access, the object has to be rewritten to the file system for a swap-out. Therefore, when only reading data, using *const*-pointers is highly encouraged as will be seen in section 5.5.4. (lines 17-19)

- **Convenience macros.** When adhering to an object and pulling a pointer should happen in the same slot, we provide convenience macros that create the `adhereTo<>`-object together with pulling a pointer in a single line. For class members this may happen shadowing a parameter. In this case, the resulting code reads as if the class would contain an unmanaged array of the same name. Of course, *const*-versions of these macros exist, too. (lines 21-30)

- **Multithreading options.** When using Rambrain in a single threaded context, Rambrain throws an exception when the user tries to pull pointers referencing more data than the physical memory limit at once. This can be disabled by a function call to enable overcommitment in multithreaded situations. In this case, pulling a pointer that would violate the limits blocks until enough RAM has become available by other threads destroying their `adhereTo<>`s. (line 33). However, this can potentially introduce a deadlock. Take for example a couple of threads that need two pointers each to start their calculation. Assume only half or less of these `managedPtr<>`s fit into RAM. In this case, all or some threads may have requested the first of the needed two pointer in parallel. Since Rambrain cannot free pulled pointers while the respective `adhereTo<>`s in scope exist, it blocks all threads and waits for memory to become available to swap-out. This, however, will never happen, as all threads are waiting and no thread is eventually finishing to unlock data for swapping. To circumvent this situation, the user may use a globally locking scope conveniently provided by Rambrain (lines 37-39). It is however highly encouraged not to over-commit memory also in multi-threaded situations as performance may drop by this forced serialisation.

5.3.3 Design considerations for user code

Having introduced the basic usage style of the library, let us evaluate the impact of using Rambrain on code design. While the syntax suggests that there would be nothing to keep in mind, a few limits and caveats apply.

Maximum problem size

Rambrain's physical memory usage is limited to a certain amount the `managedPtr<>`s may consume.[3] As Rambrain cannot use the native OS paging mechanisms, it is bound to the memory limits set by the user. Consequently, the set of currently existing `adhereTo<>`s[4] marks data as in-use and determines what cannot be swapped out. Additional managed pointers may only consume the remaining free memory. Thus, Rambrain will be unable to manage problems that demand the simultaneous use of more data than this limit. The code has to be written in a way that the maximum simultaneously accessed data amounts to less bytes than the limit. This usually is the case anyway as algorithms are being formulated in a local way on the data.

[3] Currently we do not track the overhead imposed by the usage of Rambrain, as well as other heap allocations. This is planned for a future release.

[4] Explicit delayed loading can be emplaced to limit this to the set of adhereTo<>s that a pointer was pulled from.

Data structures

The size of the simultaneously used data structures relates to the way of solving a problem. A matrix operation, for example, can typically be formulated on various matrix representations such as rows, columns, sparse single elements or smaller submatrices. To gain something from managing such a subobject, the user has to take care that the payload per managed pointer is large enough, so that the overhead of managing the data becomes small. We propose allocating smaller structures via traditional mechanisms and leaving the data-intense elements to Rambrain. If however a `managedPtr<>` is chosen, it is vital to keep in mind that this block of data can only be swapped out and in as a whole.
Ideally, all elements of a single requested `managedPtr<>` will be needed in one step of a calculation. If not, Rambrain might end up having to swap in many excess bytes to use just one or two elements. Fortunately enough, the same argument applies for normal CPU cache locality and developers are used to developing for this consecutive, local access scheme. For a review of the term locality and further hints please see for example Chellappa et al. [2008], Denning [2005]. Therefore, existing and highly optimised libraries are perfectly suited to be used together with Rambrain.

5.4 Architecture and Design

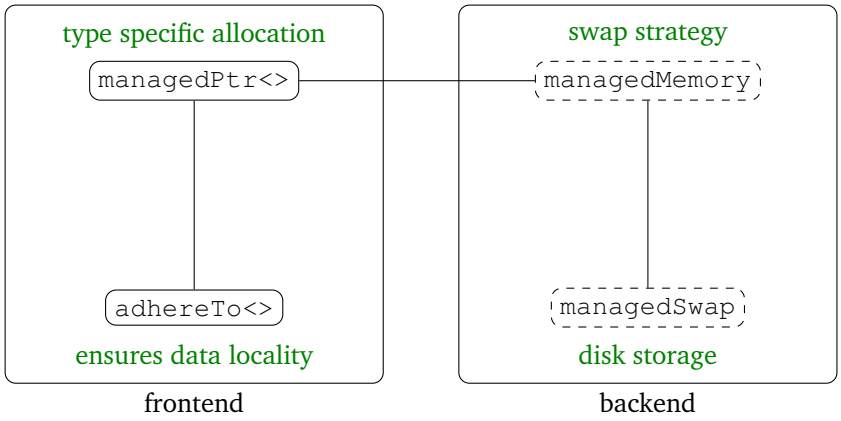


Figure 5.1: **Architecture of Rambrain:** Rambrain is divided into four major classes, each serving a distinct purpose. The classes in dashed boxes are abstract classes.

Having described the interface of Rambrain, let us now describe how Rambrain is internally implemented and what design decisions have been taken to serve the user's data requests. As depicted in Fig. 5.1, Rambrain is divided into four independent classes. While the user front end is implemented in a standardized way by the two classes `managedPtr<>` and `adhereTo<>`, whose functioning has been described above, the abstract backend classes can be inherited to implement a custom strategy which elements to select for swapping. We currently serve two implementations of these classes each. One amounts to a dummy class that is used for testing purposes. The other implementations, `cyclicManagedMemory` as well as `managedFileSwap`, will be described in the following sections. We provide profound source code documentation for all classes. The documentation can be compiled from source code using doxygen [van Heesch, 2015] or viewed online [Imgrund and Arth, 2015a,b] in a daily

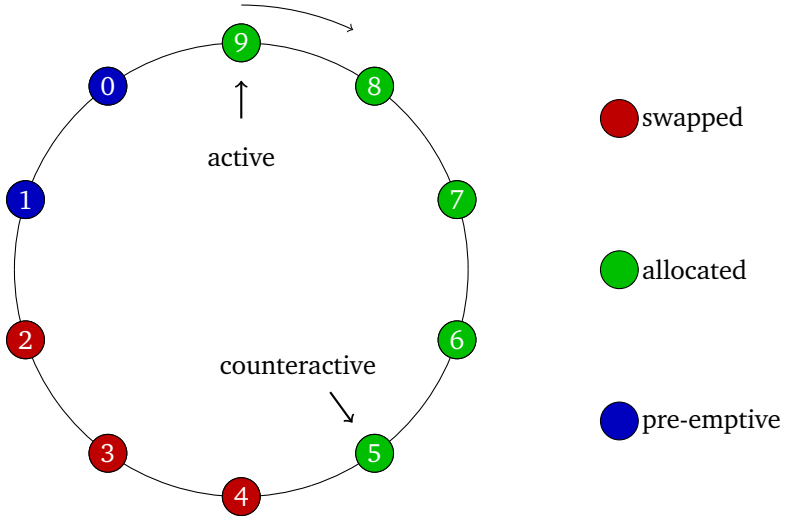


Figure 5.2: **Cyclic managed memory:** Having accessed one element, it is very likely that the former next element will be the next one this time, too. Obeying this ordering, the algorithm will asynchronously pre-fetch “pre-emptive" elements and swap out allocated but unused elements when necessary.

generated version.

5.4.1 Swapping Strategy

It is a major design decision which elements to choose for swap-out to secondary storage when facing many currently not used objects. In this section we argue that a generic strategy should be at least capable of handling random access and access in the same order in an efficient way and describe the actual implementation.
When swapping out the same amount of data to media not capable of fast random access, swap-out size and fragmentation factors limit the speed achieved in a practical situation: The throughput per byte to be written/read is reduced when writing small chunks only, as the overhead of managing the transfer both physically and logically will take a greater fraction of execution time of the request. This is especially true when using hard disks as secondary storage: When fragments of the data needed are distributed over larger parts of the disk, the read/write head of the disk has to be positioned differently at every fragment. This process consumes more time than accessing consecutively stored data. While this argument does not apply for modern solid state disks anymore, splitting data over multiple locations still poses an overhead as there must exist structures to describe and manage the splitting. Consequently a strategy writing out and reading in larger and consecutive parts at once will in general be faster than a strategy swapping out small chunks.
With no prior knowledge on what access pattern the user will impose on the data we can only make general assumptions and search for a strategy which can learn access patterns. The actual pattern encountered will lie somewhere in between the two extremes of a completely ordered and repeated sequence and random access patterns. Thinking of looping over an array of data, which is very common in scientific codes, the most simple strategy is based on the assumption that if one element has been accessed right after the other, it repeatedly may be requested in that sequence in the future. Having accessed all elements, it is most likely that the first element will be accessed again. When there are multiple array objects, this also holds when a subset of

objects is under consideration. Even when needing only a subset of all arrays, it is likely that the elements of the array will be accessed in the same order. This assumption suggests a cyclic strategy which we implement in the `cyclicManagedMemory` class and illustrate in Fig. 5.2. This order is represented as a doubly linked list of element pointers with connected end points. To organize this as an effective queueing system, the most recently accessed element is marked with a so called “active" pointer and the last still allocated and not swapped out element as “counteractive". The counteractive element is followed by swapped out elements or elements that are in the process of being written to secondary storage. When accessed in an ordered way, we may keep elements in physical memory for as long as possible. The cycle defines a reasonable sequence of swap-out: the elements that have not been accessed for the longest time are the next candidates for swap-out. They are conveniently found by dereferencing the counteractive pointer and moving this pointer backwards as elements are swapped. This will write large chunks of data consecutive into the swap files. When a swapped out element is requested by the user, also the elements that are presumed to be needed next will be loaded pre-emptively and the elements will be placed in front of the former active element.
In this way, accessing the next element in a local sequence will be very fast as it can have already been loaded and no re-ordering has to be done to the cycle at all. Only the active pointer has to be moved backwards one element to apparently move all active elements one position forward in the cycle. As long as the arrays themselves will be accessed consecutively, local ordering is also preserved by this scheme when interchanging access to various arrays.

5.4.2 Pre-emptive element swap-in and decay

It is a non-trivial question to decide the amount of bytes which are to be swapped in pre-emptively. A pre-emptively swapped in element will use up free physical space. Thus one has to make sure to not load unneeded elements that would be swapped out again immediately. This could cause major increase of IO-operations, thereby slowing down the system. It is prevented by tracking the amount of pre-emptively swapped in bytes. Pre-emptive swap-in will take place only as long as only a certain number of pre-emptively loaded bytes or less are present. If a pre-emptively loaded memory element is accessed by the user, it's size will be subtracted from the pre-emptive budget. If an element has to be swapped in from the swap file, the next elements will be fetched too, until the pre-emptive budget is filled up again. In this way, random access does not cause additional overhead by swapping in unnecessary bytes as the pre-emptive budget will always be near its limit and thus no further pre-emptive elements are swapped in.
This procedure however can lead to a constantly filled up pre-emptive budget. Imagine that an array A fills the RAM completely before an array B is accessed consecutively. Given that some elements of A have been loaded pre-emptively, they will never be used while B is accessed. Thus, they effectively block the pre-emptive budget that would be useful in loading B consecutively. To avoid this situation, Rambrain implements a decay of pre-emptive elements. The amount of decaying pre-emptive elements is determined by probabilistic arguments to prevent random access from producing too many useless pre-emptive bytes in the following way:
The maximum size of the pre-emptive budget can be used to estimate the probability of hitting a pre-emptive element at random:[5]

$$P_{\mathrm{preemptive}} \approx L_{\mathrm{preemptive}}/(L_{\mathrm{ram}} + L_{\mathrm{swap}}) \leq L_{\mathrm{preemptive}}/L_{\mathrm{ram}}$$

Where L_{ram} is the maximum physical memory allowed, L_{swap} the amount of occupied swapped

[5] Assuming equally distributed element sizes which are only a fraction of the pre-emptive budget.

out bytes and $L_{\text{preemptive}}$ the size of the pre-emptive budget. Now, every time an element is not available in RAM, we determine the amount of pre-emptive elements that have been accessed since the last element had to be swapped in. The probability that these N elements have been accessed randomly consequently can be estimated by $P^N_{\text{preemptive}}$. If this value drops below 1 percent, we let decay twice the amount of the free pre-emptive budget, but at least one byte. Decaying implies swapping out pre-emptive elements to make space for new pre-emptive elements. This typically implies loading at least two elements pre-emptively, as the pre-emptive swap-in fraction is by default set to ten percent and this fraction squared equals the significance level assumed above.

5.4.3 Swap file usage

When loaded into RAM, the data area of a `managedPtr<>` has to be allocated consecutively as pulling a pointer guarantees consecutive layout. On secondary storage devices we may split up the data over various swap file locations. While this is not desirable, it is of use when free swap file location is running out and we want to use smaller left-over chunks from previous deallocations.
Another major difference to managing heap memory, like the memory allocator in the standard libraries that is interfaced by the `new/delete` operator implementations, is that one cannot easily use the free space for the management overhead. This is because the managing structures have to be accessible very fast and would cause considerable latency when resident in secondary storage.
Of course managing the chunks of the swap file in physical memory poses unavoidable overhead. It will limit the amount of managed memory as this overhead grows over the physical size of memory available. At the moment the user has to manage large enough data amounts in one `managedPtr<>` to keep this overhead small. While this sounds like reintroducing the problem we sought out to solve, we find a typical memory overhead to be 5 to 10 percent of the amount of allocated structures when the data content is about 1kB. This amounts to being able to manage half a terrabyte of data as if it were in RAM on a 32GB machine. The data would be saved in roughly $5 \cdot 10^8$ `managedPtr<>`s of this size. It is advisable to switch to higher memory loads per `managedPtr<>` which reduces the overhead by the according factor, making more space addressable on disk. We plan to pack up objects into larger sets in future versions of the library to further reduce the overhead. It is also planned to monitor the overhead and strictly constrain it to the overall limit in future releases.[6]
Thus, given the task to swap out a `managedPtr<>`, the standard implementation of this component, `managedFileSwap`, checks its list of free chunks of memory in the swap files and tries to find the first free chunk the `managedPtr<>` fits into. If it fails to find such a chunk, it starts to split the data consecutively over the remaining gaps. If this also fails, it cleans up cached `managedPtr<>`s produced by `const` accesses and tries again. If no free space is left, it will simply fail. As this unfortunate case may happen after days of calculation, we also provide a swap policy mechanism that states how the library should react in that case. Policies amount to “fail in case of a full swap", “ask the user if he wants to assign more swap space" or “automatically extend swap space if free disk space is left to do so".

[6]This, however, is a non-trivial task as typically the standard memory allocation implementation has the control over the system call extending heap size.

5.4.4 Asynchronous IO and Direct Memory Access

The main techniques to write out large data sets to secondary storage are Memory Mapping (MM), Direct Memory Access (DMA) and using Asynchronous IO (AIO) or a mixture of these. We briefly review the different approaches with respect to the task of transferring objects from primary to secondary storage:

- **Memory Mapping**: The memory management unit in control of the virtual address space can be used to seemingly load contents of a whole file into physical memory. The same process used for paging will be utilised to write out or read in missing pieces and let an application use all space at once. When dealing with large files, this technique is very popular, as it is fast (may use DMA internally). However, when files become too big, the memory management unit quickly runs into similar problems to the one encountered with native swapping. A possible fix may be to map only parts of the swap files. In this case, however, one has to control tightly which mappings to close first, as closing will block when the mapped region is not written to disk completely. While there exists kernel hinting, a technique to tell the kernel which pages to write out first, the one-to-one mapping of allocations to the page file poses a bigger obstacle. Optimal decisions where to store certain elements are hard to find in a generic way and one is again limited to consecutive memory allocations. Splitting data would render pulling a pointer to consecutive memory impossible. Furthermore, the advantage of directly mapping allocations to swap file locations quickly can become a problem when the data has to be moved to still use a minimal memory mapped region. We thus quickly deferred using this method. There may be some interesting features to it, as automatic pre-fetching might already mimic an early stage of pre-emptive loading. Cleverly opening and closing such page-file “windows", however, is hard to handle having no guarantees for future access patterns.

- **Direct Memory Access**: DMA can in principle copy parts of memory directly to secondary storage without routing the data through the CPU. It is fast in both throughput and latency. However, it imposes memory alignment restrictions on both sides and supports only writing chunks of a certain size (typically 512kB for hard disks). Since writing is direct, the action bypasses any buffering by the kernel and thus directly leads to disk access. While this can be advantageous in situations where one writes out many consecutive datasets and implements a write cache on ones own, it typically leads to overhead in our use case. Together with the imposed alignment restrictions, it is not clear how to write an efficient implementation without writing complex scheduling code or having lots of overhead when user objects do not fit into the DMA alignment. DMA, while fast, is very complex to handle in situations where a priori it is not clear what the user requests from Rambrain. Thus the benefits of fast IO and low CPU impact vanish in light of kernel file system buffering efficiency. There is a long going discussion involving Linus Torvalds who highly discourages the use of DMA by the user [please see Torvalds, 2002].

- **Asynchronous IO:** The Linux kernel provides the user with the possibility to load and save data asynchronously to the files. Primary actions are taken only on the file system cache which has gone through a long evolution and is by now a very fast and efficient way to use free physical space without negative effects under high load. Furthermore, DMA or Memory Mapping techniques may be present in the background to bring the cache in sync with the secondary storage. Implementing Asynchronous IO upon normal buffering

implies fast execution and efficient write-out while at the same time being robust to architecture changes. Finally the most efficient way of actually carrying out a certain storage operation may only be found out at system level.
The interested reader may be warned, however, that there currently exist three AIO implementations: kio (Kernel Asynchronous IO), libaio (which is just a C wrapper for the former) and POSIX AIO. The latter is currently implemented as blocking AIO, the former is not guaranteed to be truly asynchronous, as its implementation is file system driver specific. We use a pool of submitting threads using AIO to provide true AIO where possible and simulated AIO otherwise, using the libaio wrapper for the system calls. In this way, IO operations will be non-blocking and have a low impact on CPU load.
By using asynchronous read and write requests, Rambrain is capable of loading data in the background with small impact on the CPU load. A technique for doing this is to first create the `adhereTo<>`-object, which triggers swapping in of the object. While the asynchronous IO is swapping in the element, other calculations can be done. When finally pulling the requested pointer, it may already have been copied in in the background. A graphical scheme comparing synchronous and explicit asynchronous requests to Rambrain is available in Figure 5.3 and a schematic listing of the code producing this access scheme can be found in listing 5.4. Putting the highlighted line four after line six would constitute a synchronous version of the code. As the application can already process other data while fetching in next needed objects, this can effectively hide latency similar to GPU programming techniques or pre-fetching for caches [see e.g. Callahan et al., 1991].

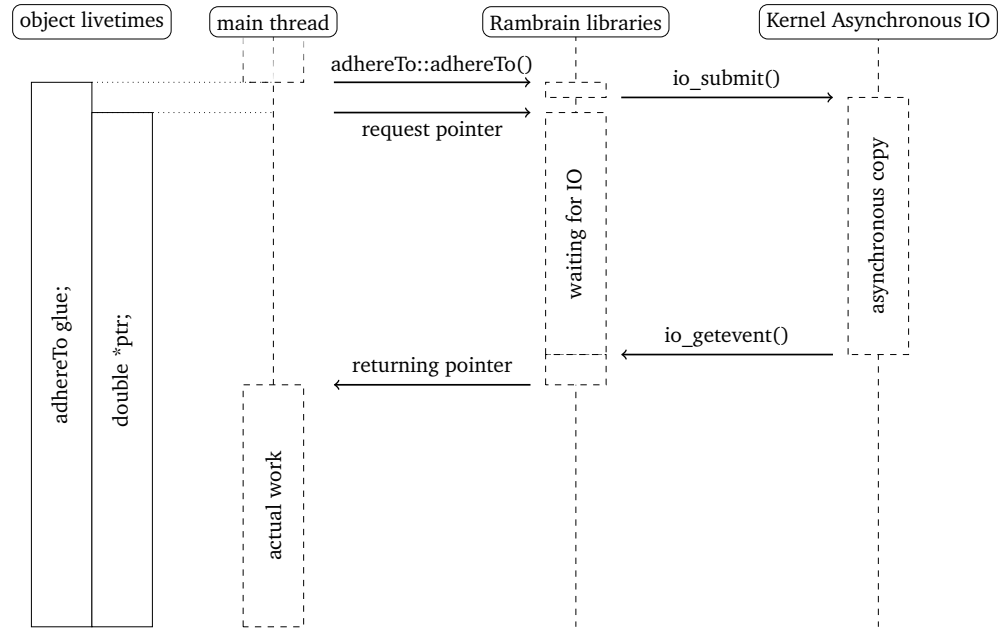
Listing 5.4: Explicit asynchronous access

```
managedPtr<double> data(1024), data2(1024);
...
adhereTo<double> glue(data);
adhereTo<double> glue2(data2);
double* ptr = glue;
do_something_on_data(ptr);
double* ptr2 = glue2;
do_something_on_data2(ptr2);
```

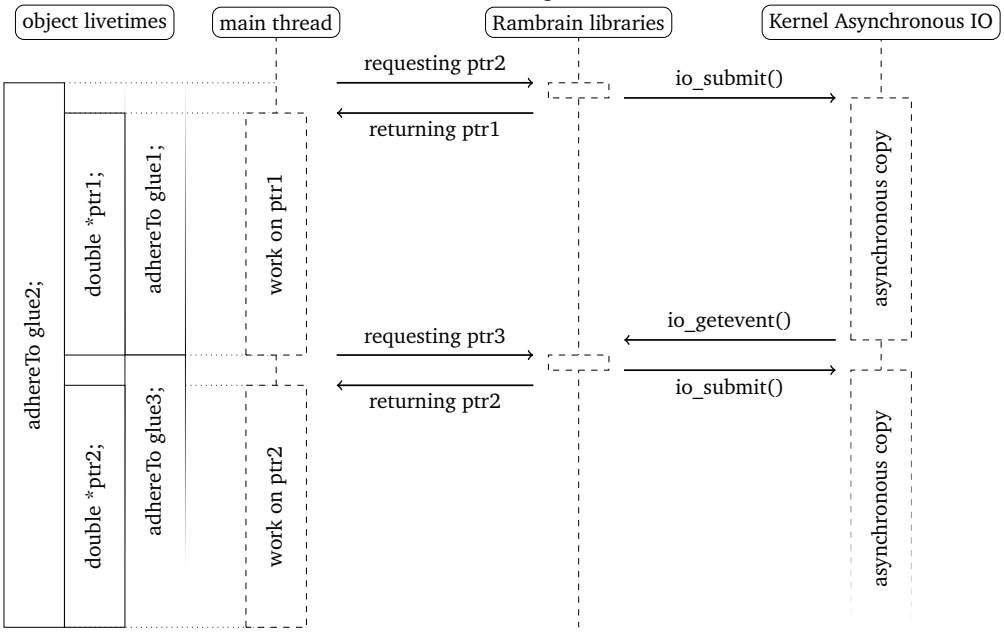
Having chosen AIO for transferring the data to secondary storage, the actual implementation is simple on the interface side but quite demanding on the scheduler side, as the scheduler has to deal with non-complete swap-outs and swap-ins when scheduling further action. As a rule of thumb, it has been found very useful to “double-book" memory in the sense that chunks moving from or to physical memory will demand their size in both budgets. At the same time we also track the amount of memory which will be freed by such actions (and thus can be waited for when needed). When completed, the budget of free memory on the source side will be restored to the correct value and the bytes which were pending before will be subtracted from the pending bytes count. In this way, the scheduler can find the right strategy, given currently pending IO, and demand a small amount of IO to satisfy its constraints imposed by user requests.

5.4.5 Compatibility to multithreading

Multithreading complicates writing the scheduler code a lot since one has to be very careful that the needs of one thread do not interfere with the needs of another thread. Scheduler and swap both are written as one instance shared by all threads. This design decision was taken



(a) Blocking IO



(b) Explicit asynchronous IO

Figure 5.3: **Exemplary interaction of user code with Rambrain library.** Rambrain may be faster when giving clues about upcoming data requirements. While in (a) the time waiting for data to arrive is wasted, the user may use this idle time for calculations on already arrived data, as depicted in (b) and written in listing 5.4. As preventing idle time is highly desirable, Rambrain tries to behave like case (b) without the user explicitly hardcoding this. In order to do so Rambrain tries to guess the upcoming data demands of the program and automatically pre-fetches elements that will be needed.

as data may be shared among threads and thus needs a common swapping procedure. Copying data between threads however will result in various `managedPtr<>`s for each instance. Consequently, passing `managedPtr<>`s and `adhereTo<>`s from one to another thread has to happen thread safe, as well as access to one `managedPtr<>` from multiple threads. Thread safety in this sense does not mean that one thread has exclusive access to a managed pointer, but that the mechanisms ensuring the availability of the data are written in a way that the object is present if at least one `adhereTo<>` in any thread is present and that the object may be swapped out at destruction of the very last `adhereTo<>` instance.

5.5 Results and Discussion

In this section we measure how code which utilises Rambrain compares to a code without Rambrain. Measuring performance is a non-trivial task for technical as well as theoretical reasons. First of all, tests should be reproducible and measure the overhead imposed by Rambrain. However, reaching this goal is non-trivial, as file system operations, kernel Asynchronous IO or scheduler performance in a multithreaded situation may affect the overall performance as well. Especially the typical use case - a developer seeking to work and debug on the same system - is hard to simulate in a reproducible and meaningful way. Separating library-imposed overhead and IO performance would be of no use either, as the user is interested in overall performance. Most of the carried out tests however will be highly speeded up by disk caching, which is also found in a productive system. We emphasize that while only RAM-to-RAM copying is done by the OS in these cases, these tests measure best the overhead implied by the workings and logic of the Rambrain library, since once the user is I/O limited, test results will be dominated by hardware performance.
In order to provoke swapping actions we set up a test system finding a PC with the smallest physical RAM module sizes removing all RAM modules up to one. The tests were then carried out using OpenSuse 13.2 (based on kernel 3.16) on an Intel(R) Core(TM)2 Quad CPU Q6700 operating at 2.66GHz on an ASUSTeK P5NT WS mainboard with 32Kb L1 Cache, 4MB L2 Cache and a standard unbranded 2GB memory module. The hard disk used is a Samsung SpinPoint S250.

5.5.1 Library overhead without swapping

We present the overhead the library imposes on the execution time of a user code in a regime, where actually nothing has to be swapped. This allows to judge whether Rambrain reaches near-to-native performance and thus can be employed if it is unclear whether it will be needed on the target system. We propose a test in which we perform a rather simple n-body simulation of a fixed set of particles using a Forward-Euler integrator [Euler, 1768]. While each timestep only depends on the last position and velocities of all particles, we save the trajectories and velocities along the way in two dimensional arrays. A typical use case for this is in place visualisation of such a simulation. Therefore, the memory used by the program grows over time, adding two vectors per particle in each iteration. The results of both runs are shown in Fig. 5.4.

In the beginning of the simulation, when hardly any data is present, we notice quite a big relative overhead of the Rambrain library. However, this only amounts to an absolute difference of only one to two seconds. From a few MB of data on, both curves show the same scaling with time, which is given by the algorithm itself. The relative overhead presented by the blue line declines very rapidly and finally converges down to a value between one and two percent close

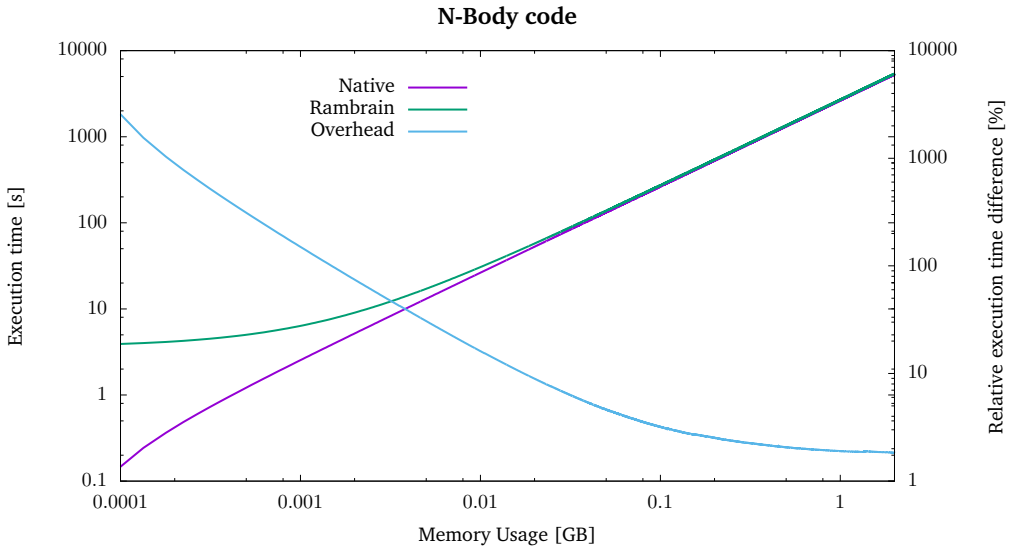


Figure 5.4: **Execution time of a n-body code:** We present timing information from a simple n-body code which accumulates data by saving particle trajectories and velocities. By comparing a version with and without rambrain we see that the overhead of the library amounts for only a few percent of execution time in the regime of reasonable data sizes.

to the two GB mark.
In conclusion, a code utilising Rambrain is always a bit slower in the regime where no data has to be swapped out compared to native code. However, the impact on execution time is not a very big factor and we see no strict need for user to completely switch off Rambrain in this case.

5.5.2 Matrix operations

In this subsection we demonstrate the internal movement of data for a common problem: Transposing a big matrix which itself does not completely fit in memory. We save matrices block wise, as it is done in many linear algebra libraries [see e.g. Blackford et al., 2002]. This allows for a straight forward migration to a Rambrain version of the algorithm, simply replacing one layer of pointers by a `managedPtr<>` class.

The result is shown in Figure 5.5. The left part of the plot shows the data allocation phase. At first the main memory is filled up very quickly with data, then data is consecutively swapped out to make room for more allocations. In the transposition phase afterwards, data is exchanged from swap to memory and back, loading all necessary blocks for the current transposition step. Please note that the asynchronous nature of Rambrain makes it very difficult to measure these values at a few discrete time points, since it is not clear when exactly the AIO events are handled in the background. Finally, the deletion of data is also plotted in the graph, but happens so fast that it is below the resolution limit of this plot.

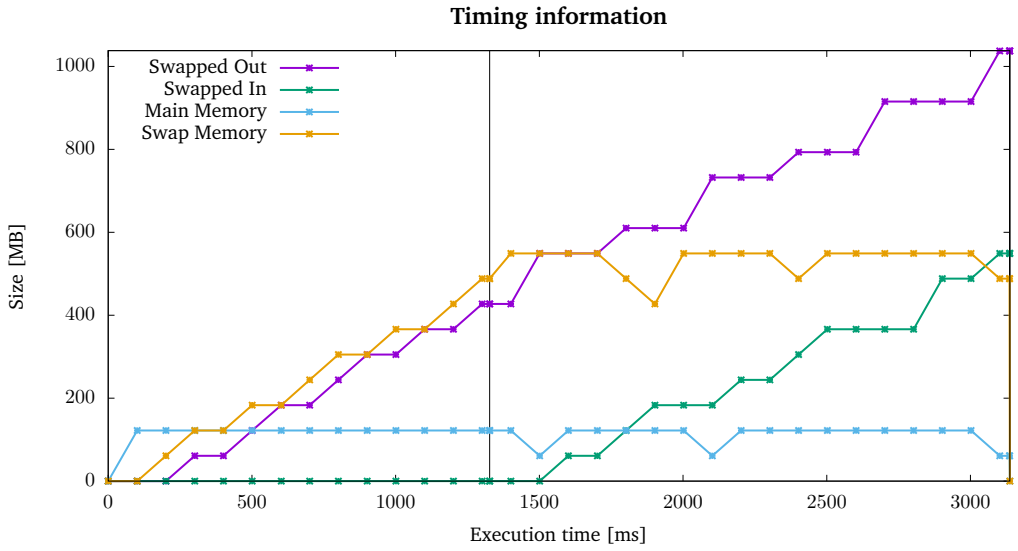


Figure 5.5: **Data movement for one 'Block' algorithm matrix transpose:** We show how data is moved between main memory and swap in one matrix transpose run. The vertical line marks the time point when the execution progresses from data allocation to the actual transposition.

5.5.3 Asynchronous IO and pre-emptive reading/writing

In this subsection we address the possible speed-up in execution time one can gain by efficiently using the asynchronous nature of Rambrain and the possibility to pre-emptively load and unload elements automatically.
To measure the performance of this mechanism, we propose the test shown in listing 5.5. We set up a two dimensional array which is realised by a list of managed pointers. While keeping the first dimension (i.e. the amount of one dimensional arrays) fixed at 1024, we vary the size of the underlying arrays (second dimension, $bytesize$). In order to measure the speed-up by asynchronism and pre-emptive actions we need to give the library some time to work in the background. Therefore, as in a typical use case, we iterate over the arrays in consecutive order and write the result of a simple integer multiplication into the respective array. We vary the percentage of the array that data is written to ($load$) as a second parameter, simulating an arbitrary computational load that scales with the data. The results of this test are presented in Figure 5.6.

It is clearly observable that the execution time decreases due to pre-emptive strategy. Increasing the work which is done on the data in the left plot, the library's overhead is already masked at a few tens of percent of touched array elements. Working on the file buffer cache only, this test shows the minimal overhead of the Rambrain libraries. In a real use case scenario, the required computational load to completely mask swapping is increased. This result clearly encourages the user to leave the standard behaviour of pre-emptive support enabled whenever possible. Even if the data access is completely random, it does not imply a big performance drawback to try to be pre-emptive. Of course a problem-specific approach pre-fetching exactly the next needed elements without trying to guess can improve performance here. However, this strongly violates our assumption, that we value development time over execution time. We therefore argue that this optimisation leads towards developing a customized out-of-core

Listing 5.5: Standard implicitly asynchronous loading

```
unsigned int numel = 1024, bytesize;
managedPtr<managedPtr<char>> arr ( numel, bytesize );
ADHERETOLOC ( managedPtr<char>, arr, ptr );
float load;
float rewritetimes = ( float ) load / 100.;
int iterations = 10230;

for ( int i = 0; i < iterations; ++i ) {
        unsigned int use = ( i % numel );
        //AdhereTo
        adhereTo<char> glue ( ptr[use] );
        //Pull the pointer to the object
        char *loc = glue;

        //Produce some computational load
        for ( int r = 0; r < rewritetimes * bytesize; r++ ) {
                loc[r % bytesize] = r * i;
        }
}
```

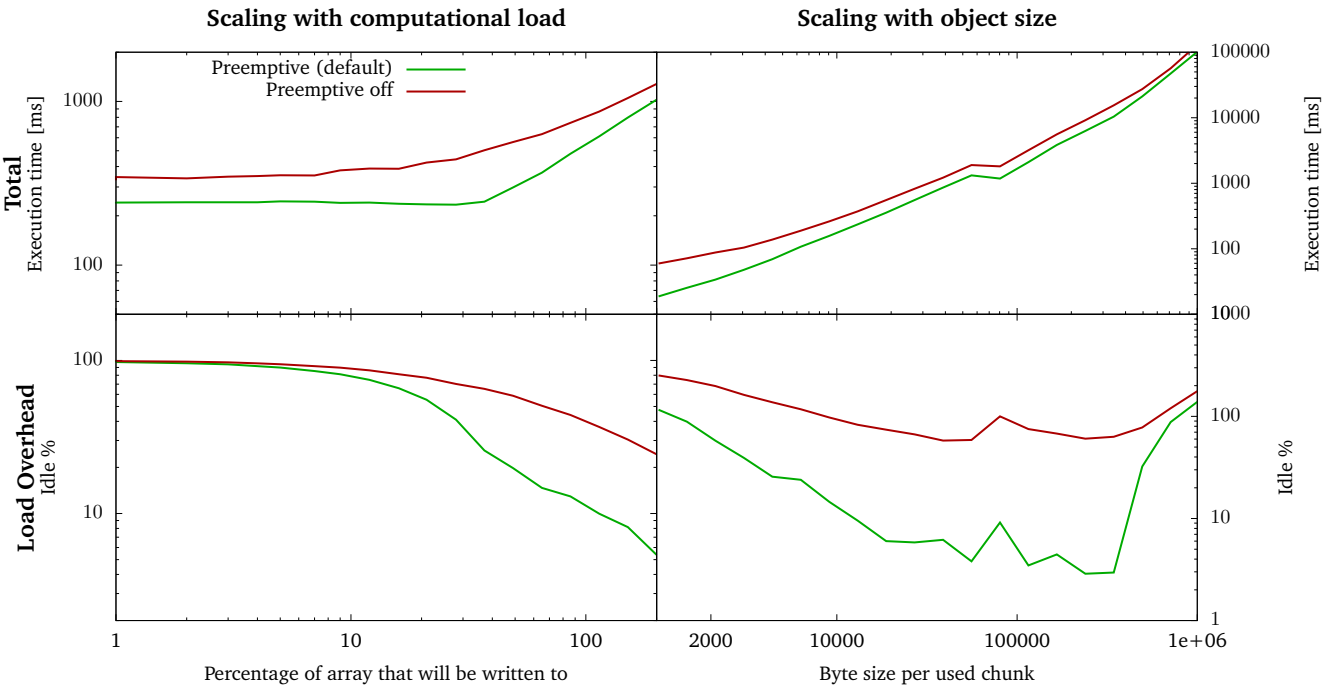


Figure 5.6: **Pre-emptive loading:** We compare enabled and disabled pre-emptive mechanism of Rambrain and find that the pre-emptive behaviour of Rambrain results in a significant performance boost.

algorithm, something no generic memory manager can substitute for.

5.5.4 Constant vs Non-Constant

Our next test is designed to examine how much time is saved by properly pulling `const` pointers when possible. As outlined in section 5.3.2 it is possible to request a pointer to constant data from an `adhereTo<>` object instead of a pointer to mutable data. This should be done in general, see e.g. Meyers [2012], but is of special importance to the case of Rambrain. Not following this best practise will leave Rambrain with no clue on whether the data has been modified and forcing Rambrain to write the data out to the swap again. Hence, if the data has already a representation in the swap and is addressed as constant, this copy is kept as long as the swap has enough free space. When the in-memory copy of the data pointer is later deleted and a swap-out occurs, the data needs not to be written out again, saving expensive writing operations.
In order to test this mechanism we allocate two blocks of data consisting of an array of smaller data chunks. The first one we call the real data while the second one is the dummy data which we will adhere to and pull a pointer from to ensure the real data being swapped out due to memory restrictions. Afterwards we access the real data and the dummy data in alternating sequence, once swapping in the data `const` and once non-`const`. We measure the time it takes to swap in the dummy data in both cases, ergo capturing the time it takes to also swap out the real data. We present the resulting behaviour for different sizes of data blocks in Figure 5.7.
We notice that the change in execution time by `const`-access obviously scales with the amount of data, since it is highly dependent on the time it takes to complete the swap-out. In the regime of a data block amounting to between one and ten megabytes, we decrease the execution time of the relevant code sections by about 20 to 30 percent. Since these are relatively small data sizes in comparison to the main memory, we can assume that these data swap-outs are completely handled by the disk cache. Therefore we save only the time for cache management and basically a memory copy. When we enter the regime of secondary storage IO we can expect the difference in execution time to be even larger since the secondary storage itself is much slower than the main memory. For most storage types, storing data takes longer than reading data, thus we expect this mechanism to save even more time in this case. It is strongly advised to use `const`-access whenever possible, also in light of other caches' properties and optimizations being used by the compiler.

5.5.5 Comparison with native OS swapping

Finally, let us compare the performance of Rambrain in the use case against system swapping. In principle, a local administrator can equip a Linux system with more swap space than usual by creating additional swap files or partitions with the system command `mkswap` and enable them for use with the command `swapon`. However, please note that it is not possible to do so as a normal user. Additionally, this approach requires the allocation of the whole swap file space on secondary storage already in the beginning - regardless of how much of it will be actually used. Using this technique we create and enable a 10GB swap file on the described test machine.
We compare a code which uses Rambrain to a non-managed code utilising system swapping. We carry out two different runs: In the first one, data is written consecutively to an 8GB sized matrix. In the second one, the application randomly writes to elements of this matrix. In the latter test we explicitly disabled the pre-emptive swapping algorithm.
On some attempts to run unmanaged, the native application is killed by the OOM-killer. This

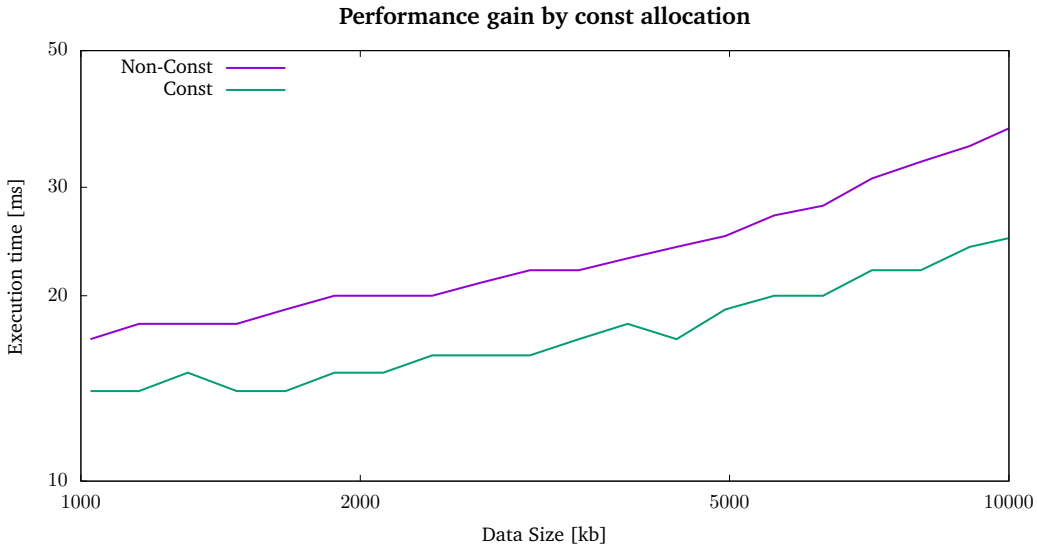


Figure 5.7: **Speed-up by pulling const pointers:** We run a simple test where data is drawn once as constant and once as writeable pointers and compare the time it takes to swap out the data afterwards in a regime where all the data still fits in the disk cache.

probably happens due to the fast growth of heap memory. Also having a swap file which is not at least about 25 percent bigger than the actual swapped size often provokes the OOM-killer to terminate the process. Even if the OOM-killer does not kill the test process, it may be that it may shut down other processes in the background to free memory for the test process. When the attempts succeed, the system is virtually unusable as even opening another shell prompt takes minutes. Furthermore, the interference of the native code with the system does not stop when the application exits, but leaves the system in a slowly reacting state for minutes to hours of usage, as large parts of other applications and system processes have been swapped out to disk. We expect that running other applications such as an integrated development environment in parallel will aggravate the situation when trying to solve the problem using OS swapping.
But also the actual execution time of Rambrain-managed code is favouring the use of our library. In the case of consecutive access, the version using Rambrain is about 10 percent faster than the native version. In case of random access, Rambrain is only 2% faster than the native code, if we obey the design limitation that at least all elements of a single `managedPtr<>` will be accessed.
This test result is further confirmed by daily experience of the authors being able to develop code on the same machine their analysis software runs in parallel without being disturbed by the process which uses Rambrain.

5.6 Conclusions And Outlook

We introduced the reader to writing code that utilises the Rambrain library. We described in detail why the proposed interface is sufficient to consistently handle data swap-out automatically and leads to satisfactory performance. We have demonstrated that the outlined mechanisms not only work properly, but also outperform naive approaches to mimic their strategy. Of course

the library cannot compete with a fully specialized out-of-core algorithm, but can save a lot of development time in providing automatic facilities for large data sets. The library handles asynchronous transfer of data which provides latency hiding of disk IO operations and reduces idle times to a few percent if computational load allows. Furthermore, we have shown that the memory and CPU overhead of the library are both in the acceptable regime of only several percent. As all of this is provided by minimal user-side interaction, we feel the goal of writing a memory manager that enables the user to transparently access multiples of the physical memory to be fulfilled. As memory management is a short-cut to just stating what data is currently needed, the user can focus on the main goals of his application at the price of only a small overhead.
The interested reader may find the code released as open source project [Imgrund and Arth, 2015a] accompanied by extensive further documentation, a list of the small set of prerequisites, notes about the (also system-wide) configuration options, a complete list of features and code examples. Interesting features are planned for future releases, such as direct mapping of file content to `managedPtr<>`'s so that loading the data beforehand is not necessary any more.
Carrying out over 100 automatic tests partly consisting of random interaction with the library on every development step and keeping track of performance has proven very useful to find bugs which only occur under rare circumstances e.g. in multithreaded situations and improved robustness of the code a lot.
We feel this library to be ready for use by a more general scientific audience.

5.7 Acknowledgements

We thank Karsten Wiesner and Christian Storm for helpful comments on presentation of advanced topics in this paper. We thank Arno Riffeser and all others who helped us to find bugs in the actual implementation. We thank Susanna Maurer for helping us organising a workshop on Rambrain at LMU.

CHAPTER 6

Paper : Does Pulsar Radio Emission come with a twist?

M. Imgrund, H. Lesch
This chapter has been submitted in slightly modified form to Astronomy&Astrophysics

Abstract: Pulsars are expected to wind up their magnetic field to some extend as the plasma far from the rotation axis cannot co-rotate faster than the speed of light. We estimate the approximate additional magnetic field energy caused by this twist, assuming a dynamical equilibrium and typical densities of the Goldreich-Julian model. Assuming an electron positron plasma, we calculate the magnetic energy being dissipated by an annihilation event and show that it is of the order of $> 10^{22} \frac{\mathrm{erg}}{s}$. We calculate the plasma's radiation reaction to the displacement field built up by the magnetic field decay. We find the expected radiation to be of the same magnitude than the decayed energy if radiated coherently.

6.1 Introduction

Since the discovery of pulsars by Hewish et al. [1968], their highly non-thermal coherent pulsed radio emission (brightness temperatures of $10^{29}K$ or even $2 \cdot 10^{41}K$ for nanoshots in Hankins and Eilek [2007], see also Lorimer and Kramer [2004], Lyne and Graham-Smith [2012] and references therein) remains one of the most intriguing puzzles of astrophysics. Especially the mechanism producing this coherent radio emission is not adequately understood. Over the last fifty years there have been many proposals on the actual origin on both emission mechanisms and its relation to the structure of the pulsar magnetosphere. The principal problem is to explain why relativistic leptons (energies of several MeV) during their one-dimensional motion along the strong magnetic fields emit photons with energy of only 10^{-5} eV (in the radio range), but in a coherent way. Given the observed luminosities a huge number of particles has to perform the same energy loss process at exactly the same time. Proposed were for example mechanism such as maser emission, curvature radiation, radiation by plasma instabilities such as the two stream instability or modulational instability, radiation by strong acceleration in the pulsar field, solitons and their interaction as well as up-scattering of photons by Langmuir waves [see e.g. Melrose and Gedalin, 1999, and references therein].

Most of these processes generate the necessary power, but lack a natural explanation for one of the following features:

- Ubiquity among pulsar population
- Specific origin in magnetosphere

- Coherent radiation process
- Microsecond fluctuations
- Radiation possibly consists of nanosecond shots

Recent developments [Hankins et al., 2003, Krzeszowski et al., 2014, Löhmer et al., 2008] suggest that pulsar radio emission might consist of sub-nanosecond shots of radiation. This poses a remarkable challenge to theoretical models as sub-nanosecond time-scales correspond to fractions of the plasma wavelength.
We want to stress that electron-positron-annihilation within well established and observational confirmed assumptions about the pulsar parameters on the base of Maxwell-Equations may suffice to explain the power and physical origin of the coherent pulsar's radio emission. We show that the reaction of the plasma to the displacement field generated by the overall magnetic field at the site of an electron-positron annihilation event may convert this energy to electromagnetic emission in the radio regime.

6.2 Topological aspects and magnetic field spin-up

It is a well-established fact that the co-rotating part of the pulsar magnetosphere cannot extend farther than the so called light cylinder radius $r_L = P \cdot c$ where c denotes the speed of light and P the pulsar's rotation period. Rigid rotation of plasma would exceed the speed of light and thus is relativistically forbidden. This divides the field lines up into two classes, one of which is co-rotating with the pulsar inside the light cylinder radius and one that extends and opens to the outside region. The second class of "open" field lines passing the light cylinder will be bent by the inertia of the slower rotating plasma leading to a toroidal magnetic field component. This winding up of magnetic field lines can drive strong winds and transport angular momentum outward, as is the case in AGN and accretion disk jet physics[Blandford and Payne, 1982, Meier et al., 2001]. In pulsar physics the short rotation period of the neutron star enhances this effect further as the plasma's inertial forces quickly bend the field lines against the rotation. Ampere's law with Maxwell's extensions states that in this case we either build up an electric field or a current will be driven through the area:

$$\vec{\nabla} \times \vec{B} = \frac{1}{c}(\partial_t \vec{E} + 4\pi \vec{j}) \tag{6.1}$$

where $\vec{B}$ and $\vec{E}$ denote the magnetic and electric field and $\vec{j}$ the current density. This can be easily seen by integrating over the surface that is perpendicular to the spin axis and crossed by the open field lines at some point. This simple reasoning strongly supports the assumption that a large current will be driven through the open field lines. However, this current does not solve the topological problem of induced magnetic field twisting caused by spinning. The pulsar is still rotating and spinning up the open field lines from its polar cap onwards. As the field lines get effectively sheared, this rotation constantly pumps energy into the magnetic field, as the shear leads to an increase of magnetic field line length and thus magnetic field strength.
Let us estimate this effect of twisting using a toy model of a cylinder of length r_L and polar cap radius r_p as depicted in Figure 6.1. In this (homogeneous) case, the spin-up leads to an toroidal field component that grows as

$$B_\perp(t) = B_{||} \cdot \frac{2\pi r_p t/P}{r_L}$$

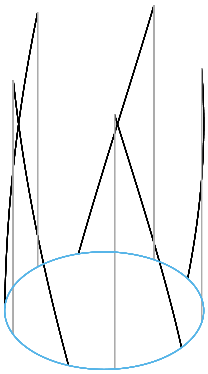


Figure 6.1: **Toy model of magnetic twisting:** when the polar cap depicted in blue rotates, it drags along the field lines. As the field lines are solidly connected at their ends, they twist up and will be stretched.

where $B_{||}$ is the strong parallel magnetic dipole field typical for a pulsar and $B_{\perp}$ the toroidal field. We will neglect the back-reaction of the now twisted current on the field configuration as we seek an estimate for the twisting angle necessarily present when a current flows. For the angle of twisting we find

$$\alpha = \tan\epsilon = \frac{B_{\perp}}{B_{||}}$$

and the field's energy density becomes

$$\mathcal{E} = \frac{B^2}{8\pi} = \frac{B_{||}^2}{8\pi}(1+\alpha(t)^2) \tag{6.2}$$

Thus the effectively differential rotation caused by relativistic limits will constantly pump energy into the magnetic field unless the twisting angle is somehow decreased. We stress here that irrespective of the actual geometry that will be established, once the field lines pass the light cylinder, differential rotation must take place in an ideal MHD scenario. Allowing no slippage, the field lines will also twist for the case of an oblique rotator, as the polar cap is still rotating around itself when projected along the rotation vector. The magnetic field energy stored in this twist will propagate by transporting the twist and current into the magnetosphere, which is a well-established idea in solar flare physics [see e.g. Spicer, 1982]. As Spicer [1982] stresses, the stored energy may be regained by current decay. While it is clear that the exact amount of twisting and thus field energy is determined by geometrical factors such as the actual length of the field line, we argue that this is not important seeking for a dynamic equilibrium solution. If we assume that there exists a dynamical equilibrium between twisting up and current decay and possible other processes, the twisting will be an energy source that is partly consumed by local current decay, and the twisting angle will remain constant as magnetic field energy will be restored by currents and field twist moving into the area. Nevertheless, locally the power that the magnetic field loses due to current decay may be calculated by deriving equation (6.2) w.r.t. time:

$$\mathcal{P} = \partial_t \mathcal{E} = \frac{B_{||}^2 \alpha^2}{4\pi}\frac{\dot{\alpha}}{\alpha} \tag{6.3}$$

We emphasize that this loss term is not the only loss term of the pulsars rotational energy but just a small fraction of total losses experienced. Both dipole and pulsar wind losses are dominant to this - as we will see below - comparably small contribution.

If a long term dynamical equilibrium can be reached, the derivative of the electric field w.r.t. time in Ampere's law has to vanish - an assumption that is also established by assuming ideal MHD. Even if an electric field is built up, this field would not grow indefinitely and the time derivative of the electric field is zero. Accordingly we are left with both other terms that let us calculate the value of α at a certain radius r from the polar cap's centre axis by integrating over the area and using Stokes' theorem. This yields

$$\frac{B_\perp}{B_{||}} = \alpha = \frac{2\pi r j}{c B_{||}} = 2\pi e \frac{r n \beta}{B_{||}} \tag{6.4}$$

where n denotes the plasma density, e the elementary charge and β the plasma velocity as fraction of light speed c. For simplicity we assume the current to be uncharged and consisting of both electrons and positrons of same density and velocity in an counter-current flow. The second factor in equation (6.3), the relative time derivative of α can consequently be identified with

$$\frac{\dot{\alpha}}{\alpha} = \frac{\dot{n}}{n} \tag{6.5}$$

as the background magnetic field and the mean current velocity is expected to stay constant in an equilibrium case. The change in density however occurs naturally by annihilation events in an electron positron plasma. To compensate for the change of α and reach an equilibrium, we expect this process to be a consumer of the twist that is produced by the rotation of the pulsar. This change leads to a steady inflow of magnetic twist and particles to compensate for the loss.

6.3 Spin-down of field lines in a pulsar magnetosphere

Goldreich and Julian [1969] calculated the excess charge density due to plasma motion through an magnetic field to be proportional to the ratio of the magnetic field and the light cylinder radius. This leads to a particle density of

$$n_{GJ} = \frac{1}{ec} \frac{B_{||}}{P} \delta \tag{6.6}$$

where we introduced a multiplicity factor δ that accounts for possibly larger background densities. In contrast to the interpretation of Goldreich and Julian [1969] we assume that we have an electron-positron plasma which density is on this order and assume for simplicity that no net charges arise. Observations indeed support a current of $n_{GJ}c$ to be on the right order [Kramer et al., 2006a]. Yuen et al. [2012] furthermore carry out RM measurements that give density values that are consistent with the Goldreich-Julian density. We assume in the following that

- The pulsar magnetosphere's plasma density follows the density proposed by Goldreich and Julian [1969]
- The plasma consists mainly of electrons and positrons [see e.g. Chen and Beloborodov, 2014, for a recent numerical simulation].

The current density associated with the Goldreich-Julian density is

$$j = ec\beta n = \frac{B_{||}}{P} \beta \delta \tag{6.7}$$

Inserting this into (6.4) leaves us with

$$\alpha = \frac{2\pi}{c} \frac{r}{P} \beta \delta \tag{6.8}$$

leading to a magnetic energy density that is increased over its a priori value by

$$\mathcal{E}_{ex} = \frac{1}{8\pi}\left[B_{||}^2(1+\alpha^2) - B_{||}^2\right] = \frac{\pi}{2c^2}\frac{B_{||}^2}{P^2}r^2\beta\delta \tag{6.9}$$

In an electron positron plasma, we expect annihilation to occur naturally. We estimate the cross section of this reaction by the Klein-Nishina cross section[Klein and Nishina, 1929] for low values values of $\gamma = 1/\sqrt{1-\beta^2}$:

$$\sigma_{KN} = \frac{\pi}{2}r_e^2/\gamma$$

where r_e is the classical electron radius. Thus, the rate at which density decreases naturally by annihilation events amounts to

$$\Gamma = \dot{n} = \sigma_{KN}cn^2\beta = \frac{\pi r_e^2 c}{2}\frac{n^2\beta}{\gamma} \tag{6.10}$$

Other current destructive processes like positronium formation[Baring and Harding, 2001] could possibly dominate the process of annihilation alone. As discussed in Baring and Harding [2001], bound positronium is neutral, the charges can no longer form an electric current. This would increase the effect calculated here. The Thompson cross section alone might be enhanced by the large magnetic field [Gonthier et al., 2014] also in case of positron electron annihilation as there is free energy in the magnetic field to decay. For simplicity we present the arguments here with the unmodified Thomson cross section, following the line of a worst case scenario.
The decrease in charge density by annihilation leads to a decrease in current flow. A spontaneous current decrease will trigger a high induction voltage that is trying to increase the current again if the associated energy is not dissipated otherwise. In the latter case, magnetic twist will be decreased. We assume for the moment that for a dynamic equilibrium, both electric fields and $gamma$ values of the plasma should merely stay at the same level and the energy stored in the magnetic field is dissipated. In this case annihilation acts as a sink of the magnetic twist while the pulsar's rotation acts as its source. We may calculate the total dissipation by inspecting (6.4) and noticing that $\dot{\alpha}/\alpha = \dot{n}/n$ and inserting this in (6.3). We are left with

$$\mathcal{P} = \frac{r_e^2 c}{8}\frac{B_{||}^2\alpha^2 n\beta}{\gamma}$$

where we now may insert our values for α and n to find

$$\mathcal{P} = \frac{\pi^2 r_e^2 e^2 c}{2}\left(\frac{1}{ec}\frac{B_{||}\delta}{P}\right)^3\frac{r^2\beta^3}{\gamma} = \frac{\pi^2 r_e^2}{2ec^2}\left(\frac{B_{||}\delta}{P}\right)^3\frac{r^2\beta^3}{\gamma} \tag{6.11}$$

Let us now calculate the total energy dissipation by this process, noticing that $B_{||}$ falls with $(z/R_{NS})^3$ where z is the distance from the polar cap. As this decay is rapid, we may calculate in cylindrical geometry. Let us use $R_P = \sqrt{2\pi R_{NS}^3/(cP)}$ as expression for the polar cap radius

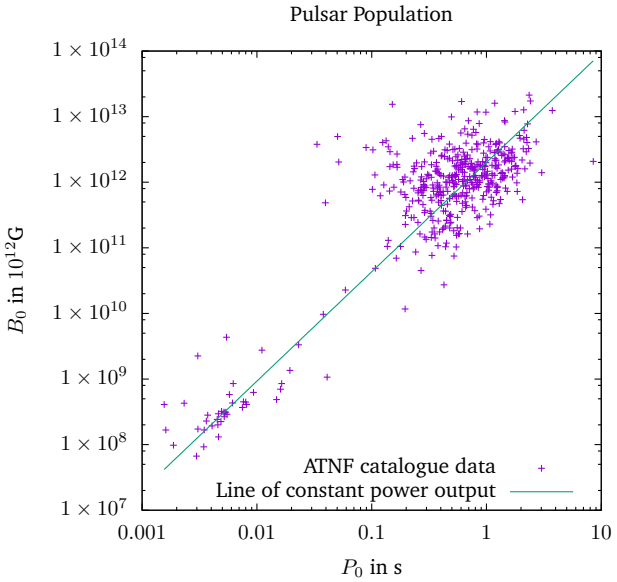


Figure 6.2: **Scaling of power loss due to current decay:** Plotting the scaling of the power loss due to current decay over the range of the pulsar population provided by the ATNF Pulsar Catalogue[Manchester et al., 2005] we find good agreement between the scaling of the proposed process and observational data.

as in Sturner [1995]. We calculate the total energy output as:

$$
\begin{aligned}
P &= \frac{r_e^2\pi^2}{2ec^2}\left(\frac{B_{||}\delta}{P}\right)^3 \int_0^{R_P} \mathrm{d}r 2\pi r \int_{R_{NS}}^{\infty} \mathrm{d}z \left(\frac{R_{NS}^3}{z^3}\right)^3 r^2 \\
&= \frac{\pi^3 r_e^2}{32ec^2}\left(\frac{B_{||}\delta}{P}\right)^3 \frac{\beta^3}{\gamma} R_P^4 R_{NS} && (6.12) \\
&= \frac{\pi^5 r_e^2 e^2}{8c}\left(\frac{1}{ec}\frac{B_{||}\delta}{P}\right)^3 \frac{1}{P^2}\frac{\beta^3}{\gamma} R_{NS}^7 && (6.13) \\
&= 7.7 \cdot 10^{21} \cdot \frac{B_{12}^3\delta^3}{P_0^5}\frac{\beta^3}{\gamma} R_{NS,6}^7 \frac{\mathrm{erg}}{\mathrm{s}} && (6.14)
\end{aligned}
$$

where a number N in the indices of the last equation denote that quantities have been normalized to the N^{th} power of ten of the value in cgs-units and R_{NS} is the neutron star's radius. Note that the strong dependence on both the pulsar radius and charge density factor may serve as an explanation for the observed radio luminosities given there is a strong dissipative process that leads to conversion of the magnetic energy into radio radiation. The calculated power misses the observed roughly $10^{25}\frac{\mathrm{erg}}{\mathrm{s}}$ [Allen and Horvath, 2000, Figure 3, assuming a bandwidth of 100MHz to GHz] by about three dimensions. Depending on the seventh power of actual values of neutron star radii and a possible density factor δ to the power of three, the exact total value is subject to high uncertainty. However, the scaling with pulsar parameters is very accurate. As is evident from figure 6.2, plotting a line of constant power depending on the input parameters magnetic field and period yields a line connecting both the millisecond and primary pulsar population supporting our case.

6.4 Induced field and radio radiation

When we calculate the energy associated with each elementary annihilation, we find

$$E_{\text{event}} = \frac{\mathcal{P}}{\Gamma} = \frac{\pi e}{c}\left(\frac{B_{||}\delta}{P}\right) r^2\beta^2 \tag{6.15}$$

which is on the order of 10 erg at the border of the polar cap. To estimate on what time-scale the energy is being released, let us consider the spacetime volume a single event occupies:

$$l_a = \sqrt[4]{c/\Gamma} = \sqrt[4]{\frac{2\gamma}{\pi r_e^2}\frac{1}{n^2\beta}} \tag{6.16}$$

This amounts to a length scale of about 6.4cm or 0.2ns. calculating the ratio of this length scale and the plasma wavelength we find:

$$\frac{l_a}{\lambda_{pl}} = \frac{1}{2\pi c}\sqrt[4]{\frac{32\pi}{r_e^2}\frac{1}{\beta}\cdot\frac{e^4}{m^2\gamma^5}} = 0.5\frac{1}{\beta\gamma^{5/4}} \tag{6.17}$$

This means that already at mildly relativistic particle flows, the expected length scale of the area the displacement current is dominant is small compared to the plasma wavelength and will take a minor role in the reaction of the plasma. Thus we may approximate the reaction of the plasma in the spacetime volume as a collective process of the plasma particles in a volume l_a^3. As the process of annihilation itself is happening very fast, the plasma needs some time to react to the field built up instantaneously by magnetic induction. If we assume that the electric field built up is of the strength of the energy drop of the magnetic field, the above yields a field of:

$$\begin{aligned} E_{el} &= \sqrt{8\pi\frac{E_{\text{event}}}{l_a^3}} \\ &= (2^{9/8}\pi^{11/8}er_e^{3/4})n^{5/4}r\frac{\beta^{7/4}}{\gamma^{3/8}} &(6.18)\\ &= 0.92\cdot\frac{B_{||}^{5/4}}{P_0^{5/4}}\frac{r}{R_P}r_{NS,6}^{3/2}\frac{\beta^{7/4}}{\gamma^{3/8}}\text{statV} &(6.19)\end{aligned}$$

6.5 Collective plasma reaction and estimated radiation output

Modelling this field to appear instantaneously we calculate the reaction of the plasma via Ampere's law, where we assume the rotation of the magnetic field to be cancelled out by the net current β_0:

$$\frac{\partial E}{\partial t} = \dot{E} = -4\pi enc(\beta-\beta_0) \Rightarrow \ddot{E} = -4\pi ecn\dot{\beta} \propto \partial_t j$$

Modelling the flow via the equation of motion for a single charge carrier, we find

$$(mc\dot{\gamma}\beta) = Ee = mc\gamma^3\dot{\beta}$$

where we used that

$$\dot{\gamma} = \gamma^3 \beta \dot{\beta} \Rightarrow (\dot{\gamma\beta}) = \dot{\gamma}\beta + \gamma\dot{\beta} = \gamma^3 \dot{\beta}(\underbrace{\beta^2 + \frac{1}{\gamma^2}}_{\beta^2+(1-\beta^2)=1}) = \gamma^3 \dot{\beta} \tag{6.20}$$

Combining the two above equations we find an harmonic oscillator for the electric field:

$$\ddot{E} + E \underbrace{\frac{4\pi e^2 n}{m\gamma^3}}_{\omega^2_{\mathrm{p},rel}} = 0$$

It is clear that while the approximated field takes a moderate value, it has not been generated by a plasma inherent but much faster process. Consequently the plasma reacts collectively being kicked by this field. We may estimate the instantaneously radiated energy by Lienard's formula [Jackson, 1998] for the relativistic radiation which in our case simplifies as follows:

$$P_{\mathrm{single}} = \frac{2e^2\gamma^6}{3c} \left[(\dot{\beta})^2 - (\underbrace{\beta \times \dot{\beta}}_{0})^2 \right]$$

which induces a damping factor to our harmonic oscillator that will diminish the available free energy. As we expect the plasma to behave collectively, we have to introduce a factor of N^α where N denotes the number of particles and α a coherence factor taking a value in the interval of $[1:2]$ where 2 amounts to full coherence while 1 represents incoherent emission. We calculate the total radiated energy per particle, and thus the damping per particle as:

$$P_{\mathrm{coherent,single}} = \frac{2e^2\gamma^6}{3c} N^{\alpha-1} \dot{\beta}^2$$

While the exact solution implies a non-linear interaction, let us estimate the time-scales and impact of the radiation process by assuming a damping term proportional to $\dot{E}$. Let us assume that the electric field induced by an annihilation event decays as fast as possible. This amounts to assuming critical damping which equals $\lambda = \omega_{\mathrm{p,rel}}$. Since the damping does not grow linear with acceleration but quadratic we expect the actual solution to decay even faster. The well-known solution to this case is

$$E(t) = a_1 \mathrm{exp}\left[-\omega_{\mathrm{p,rel}} t\right] + a_2 t \mathrm{exp}\left[-\omega_{\mathrm{p,rel}} t\right] \tag{6.21}$$

where we now solve for an initial field value of $E(0) = E_0$ and postulate that the plasma was initially in equilibrium, leading to $\dot{E}(0) \propto (\beta - \beta_0) = 0$. This yields $a_1 = E_0$ and $a_2 = E_0\omega_0$ and thus conclude for the field and its second derivative:

$$E(t) = E_0 \mathrm{exp}\left[-\omega_{\mathrm{p,rel}} t\right] + \omega_{\mathrm{p,rel}} E_0 t \mathrm{exp}\left[-\omega_{\mathrm{p,rel}} t\right]$$

$$-4\pi e n c \dot{\beta}(t) = \ddot{E}(t) = \omega_{\mathrm{p,rel}} \mathrm{exp}\left[-\omega_{\mathrm{p,rel}} t\right] \left(-E_0 \omega_{\mathrm{p,rel}} + E_0 \omega^2_{\mathrm{p,rel}} t\right) \tag{6.22}$$

As the received power is proportional to E^2, the received electric field spectrum will scale with $\dot{\beta}$. The expected time dependence of this field is given by equation (6.21) and can be written as a linear combination of the spectra demanded by [Krzeszowski et al., 2014, equation (5)].

The mean total power emitted can be estimated by calculating the total energy of a single event multiplied with the annihilation rate.

$$\begin{aligned}\mathcal{P}_{\mathrm{total}} &= \Gamma_A(n)\int_0^\infty \mathrm{d}t N P_{\mathrm{single,coherent}}(t)\\ &= \Gamma_A(n)\frac{2e^2\gamma^6}{3c}N^\alpha\int_0^\infty \mathrm{d}t\left(\frac{\omega_{\mathrm{p,rel}}}{4\pi enc}\exp\left[-\omega_{\mathrm{p,rel}}t\right]\right.\\ &\qquad\left.\cdot\left(-E_0\omega_{\mathrm{p,rel}}+E_0\omega_{\mathrm{p,rel}}^2 t\right)\right)^2\\ &= \Gamma_A(n)\frac{2e^2\gamma^6}{3c}N^\alpha\frac{E_0^2}{4\pi mc^2}\frac{\omega_{\mathrm{p,rel}}}{4\gamma^3 n}\end{aligned} \tag{6.23}$$

As the field is given by equation (6.19) and the annihilation rate by (6.10), we may readily calculate the total power density dissipated:

$$\begin{aligned}\mathcal{P}_{total} = &\frac{2^{(3\alpha-3)/4}\pi^{(13-3\alpha)/4}e^5 r_e^{(7-3)\alpha/2}}{3m^{3/2}c^2}n^{4-\alpha/2}r^2\cdot\\ &\cdot\beta^{9/2-3/4\alpha}\gamma^{(3\alpha-1)/4}\end{aligned} \tag{6.24}$$

$$\mathcal{P}_{total}^{\alpha=2} = 1.5\cdot 10^8\cdot\left(\frac{B_{12}\delta}{P_0}\right)^3\left(\frac{r}{r_P}\right)^2\gamma^{5/4}\beta^3\frac{\mathrm{erg}}{\mathrm{cm^3 s}} \tag{6.25}$$

$$\mathcal{P}_{total}^{\alpha=1.5} = 3.5\cdot 10^1\cdot\left(\frac{B_{12}\delta}{P_0}\right)^{13/4}\left(\frac{r}{r_P}\right)^2\gamma^{7/8}\beta^{27/8}\frac{\mathrm{erg}}{\mathrm{cm^3 s}} \tag{6.26}$$

Let us compare these power densities with the power expected to dissipate in equation (6.11). The most striking feature is that regardless of the coherence factor assumed, the scaling parallel to the polar cap is the same for both the input and output processes. In the case of complete coherence, $\alpha = 2$, we even find the same scaling w.r.t. the magnetic field and thus in the whole volume. Evaluating equation (6.11) we find

$$\mathcal{P} = 1.8\cdot 10^8\cdot\left(\frac{B_{12}\delta}{P_0}\right)^3\left(\frac{r}{r_P}\right)^2\frac{\beta^3}{\gamma}\frac{\mathrm{erg}}{\mathrm{cm^3 s}} \tag{6.27}$$

Thus the total power of the independent microscopic dissipation matches the expected value from Ampere's law astonishingly well. If we allow for an over-density described by δ larger than one, we may reproduce the same or more energy output at however higher but still moderate particle energies.

6.6 Conclusion

We have shown that annihilation in a mildly relativistic current with charge densities of the order of the Golreich Julian density can convert magnetic energy to electric energy. Assuming this energy to be converted roughly in the volume of spacetime associated with electron positron annihilation, we deduced that the electric energy is radiated nearly completely. The process that can be modelled as a classical damped harmonic oscillator will dissipate this energy on the time-scales of $T = 1/\omega_{p,rel}$ a quantity that evaluates to nanoseconds and below for typical pulsar densities. The time evolution of the electric field follows the form demanded by measured spectra [Krzeszowski et al., 2014]. Only the total energy output does not easily reach observed

values. Since it is strongly dependent on the radius of the neutron star and magnetosphere density, the actual number is uncertain and may be higher. A satisfying scaling behaviour with pulsar energies was discovered. As the described process only accounts for part of the energy output observed, other processes are not excluded [e.g. plasma instabilities such as Weatherall, 1998, also provide nanosecond time variability]
Even if the derived mechanism does not completely account for the emission received it suggests that the inevitable twisting of the magnetic field lines and its transport can be expected to play a major role. Perhaps treating the pulsar as a giant electromagnetic dynamo with an electric circuit as proposed by Shibata [1991] and others will at the end be more successful than searching for the source of radio emissions on the ground of micro- and plasma physics alone. For if annihilation, an effect that decays only a tiny fraction of the plasma density, unleashes $10^{21}\frac{\mathrm{erg}}{\mathrm{s}}$ of magnetic energy - what would the right mechanism do?

Acknowledgements

We thank Axel Jessner for helpful discussions on the nanoshots and spectra. We thank Alexander Arth and Ramesh Karuppusamy for general comments and discussion.

CHAPTER 7

Final remarks

Having analysed the radio emission problem theoretically in Chapter 1 and observationally in Chapter 2, we identified some of the biggest still unresolved issues. At the present state of theory, the problem of energy transport and conversion is still not satisfactorily solved. Especially in light of the newest simulation results, the kinetic energy of the plasma might not suffice to reach the necessary energy densities. While there is virtually only the modulational instability left as candidate process to explain the radio emission, the exact way of supplying the energy density necessary and the boundary condition in which the modulational instability takes place are not clear.
Our theoretical findings however also point to current decay being a versatile converter of magnetic energy to electric fields that could then drive emission. We expect more detailed numerical simulations to improve the magnetospheric models further in the next years. While the radio emission is of clearly microscopic origin and may perhaps not be generated in large-scale simulations, the new numerical results will serve as a guide to realistic boundary conditions for the radio emission process. Since the energy loss due to radio emission is negligible to other loss terms, we expect magnetospheric models to reach very high precissions also if they were incapable of revealing the radio emission process itself.
It has become clear that a high relativistic flow is not the only possibility of powering radio emission and there are many arguments against this scenario. First of all, numerical simulations do not reproduce the inner gap proposed by Ruderman and Sutherland [1975]. Secondly, there exist heuristical arguments against high Lorentz factors: For example, the high inertia of the charge carriers leads to a very low plasma frequency. Besides, it is unclear how a high relativistic flow can be powered in a setting that naturally produces currents.
Deciding that the best contribution can be made to statistical problems of data analysis, we focused on a thorough reinvestigation of the receiver equation. We understood that it is a direct consequence of this equation, that the brightest pulses are not the best timers. We furthermore showed how to improve the estimates of the remaining uncertainties. This led to more precise and accurate timing results yielding a better reduced χ^2 value than classical methods of ToA generation.
Future generation telescopes and receivers will put phase jitter and thus correlated noise in the focus of the observation data. As we argued, being able to process data losslessly and thus perhaps on the single pulse niveau, to fit global parameters, needs a fast and reliable interface to transparently access single pulse data. We developed a library that helps analyzing such huge datasets. We suspect that in the next years, an even more extended framework will be necessary to handle the data challenges ahead. Having shown that the interpretation of the

data microscopically will depend on the macroscopic data fitted, we emphasize that the best possible parameter reconstruction will only be achieved, when reprocessing the data with the gathered posterior distributions is possible.
Finally we strongly suggest further investigation of the radio emission process. Understanding the actual physical events is not only an interesting question of plasma physics, but turns out to be crucial for the understanding of secondary parameters like ToAs and the noise they are subjected to.
If we take the chance to analyse data the best way possible, and try to learn of the existing theoretical problems, we believe that the problem of radio emission can be solved by mutual progess in theory and observation.

Bibliography

M. P. Allen and J. E. Horvath. Pulsar radio luminosity laws revisited. *Monthly Notices of the Royal Astronomical Society*, 317:23–27, September 2000.

S. V. Antipov, M. V. Nezlin, A. S. Trubnikov, and I. V. Kurchatov. Experimental studies of Langmuir solitons. *Physica D Nonlinear Phenomena*, 3:311–328, July 1981.

J. Arons. Pair creation above pulsar polar caps - Geometrical structure and energetics of slot gaps. ApJ, 266:215–241, March 1983.

J. Arons and J. J. Barnard. Wave propagation in pulsar magnetospheres - Dispersion relations and normal modes of plasmas in superstrong magnetic fields. ApJ, 302:120–137, March 1986.

E. Asseo. Lattice of Langmuir solitons in the pulsar magnetosphere. *Physica Scripta Volume T*, 52:87, 1994.

E. Asseo, G. Pelletier, and H. Sol. A non-linear radio pulsar emission mechanism. MNRAS, 247: 529–548, December 1990.

D. C. Backer. Pulsar Nulling Phenomena. Nature, 228:42–43, October 1970.

D. C. Backer. Pulsar Fluctuation Spectra and the Generalized Drifting-Subpulse Phenomenon. ApJ, 182:245–276, May 1973.

M. Bailes. The art of precision pulsar timing. In S. A. Klioner, P. K. Seidelmann, and M. H. Soffel, editors, *IAU Symposium*, volume 261 of *IAU Symposium*, page 212–217, January 2010.

M. G. Baring and A. K. Harding. Photon Splitting and Pair Creation in Highly Magnetized Pulsars. ApJ, 547:929–948, February 2001.

J. J. Barnard and J. Arons. Wave propagation in pulsar magnetospheres - Refraction of rays in the open flux zone. ApJ, 302:138–162, March 1986.

A. M. Beloborodov. Polar-Cap Accelerator and Radio Emission from Pulsars. ApJ, 683:L41–L44, August 2008.

I. B. Bernstein, J. M. Greene, and M. D. Kruskal. Exact Nonlinear Plasma Oscillations. *Physical Review*, 108:546–550, November 1957.

D. Bhattacharya and E. P. J. van den Heuvel. Formation and evolution of binary and millisecond radio pulsars. Phys. Rep., 203:1–124, 1991.

L. S. Blackford, J. Demmel, J. Dongarra, I. Duff, S. Hammarling, G. Henry, M. Heroux, L. Kaufman, A. Lumsdaine, A. Petitet, R. Pozo, K. Remington, and R. C. Whaley. An Updated Set of Basic Linear Algebra Subprograms (BLAS). *ACM Trans. Math. Softw.*, 28(2):135–151, June 2002.

R. D. Blandford and D. G. Payne. Hydromagnetic flows from accretion discs and the production of radio jets. MNRAS, 199:883–903, June 1982.

M. Blaskiewicz, J. M. Cordes, and I. Wasserman. A relativistic model of pulsar polarization. ApJ, 370:643–669, April 1991.

David Callahan, Ken Kennedy, and Allan Porterfield. Software Prefetching. In *Proceedings of the Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS IV, page 40–52, New York, NY, USA, 1991. ACM.

Srinivas Chellappa, Franz Franchetti, and Markus Püschel. *Generative and Transformational Techniques in Software Engineering II: International Summer School, GTTSE 2007, Braga, Portugal, July 2-7, 2007. Revised Papers*, chapter How to Write Fast Numerical Code: A Small Introduction, page 196–259. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

A. Y. Chen and A. M. Beloborodov. Electrodynamics of Axisymmetric Pulsar Magnetosphere with Electron-Positron Discharge: A Numerical Experiment. ApJ, 795:L22, November 2014.

K. S. Cheng, C. Ho, and M. Ruderman. Energetic radiation from rapidly spinning pulsars. I - Outer magnetosphere gaps. II - VELA and Crab. ApJ, 300:500–539, January 1986.

W. Coles, G. Hobbs, D. J. Champion, R. N. Manchester, and J. P. W. Verbiest. Pulsar timing analysis in the presence of correlated noise. MNRAS, 418:561–570, November 2011.

I. Contopoulos and A. Spitkovsky. Revised Pulsar Spin-down. ApJ, 643:1139–1145, June 2006.

I. Contopoulos, D. Kazanas, and C. Fendt. The Axisymmetric Pulsar Magnetosphere. ApJ, 511: 351–358, January 1999.

J. M. Cordes. The detectability of planetary companions to radio pulsars. In J. A. Phillips, S. E. Thorsett, and S. R. Kulkarni, editors, *Planets Around Pulsars*, volume 36 of *Astronomical Society of the Pacific Conference Series*, page 43–60, January 1993.

J. M. Cordes. Pulsar Observations I. – Propagation Effects, Searching Distance Estimates, Scintillations and VLBI. In S. Stanimirovic, D. Altschuler, P. Goldsmith, and C. Salter, editors, *Single-Dish Radio Astronomy: Techniques and Applications*, volume 278 of *Astronomical Society of the Pacific Conference Series*, page 227–250, December 2002.

J. M. Cordes. Pulsar State Switching from Markov Transitions and Stochastic Resonance. ApJ, 775:47, September 2013.

J. M. Cordes and R. M. Shannon. A Measurement Model for Precision Pulsar Timing. *ArXiv e-prints*, October 2010.

J. M. Cordes, M. Kramer, T. J. W. Lazio, B. W. Stappers, D. C. Backer, and S. Johnston. Pulsars as tools for fundamental physics and astrophysics. New A Rev., 48:1413–1438, December 2004.

T. M. Cover and J. A. Thomas. *Elements Of Information Theory*. 2nd edition. Wiley, 2006.

R. Dementiev, L. Kettner, and P. Sanders. STXXL: standard template library for XXL data sets. *Software: Practice and Experience*, 38(6):589–637, 2008.

P. B. Demorest. Cyclic spectral analysis of radio pulsars. MNRAS, 416:2821–2826, October 2011.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1): pp. 1–38, 1977.

Peter J. Denning. The Locality Principle. *Commun. ACM*, 48(7):19–24, July 2005.

A. A. Deshpande and J. M. Rankin. Pulsar Magnetospheric Emission Mapping: Images and Implications of Polar CAP Weather. ApJ, 524:1008–1013, October 1999.

Gregory Desvignes et al. EPTA data release V1. 2015.

A. J. Deutsch. The electromagnetic field of an idealized star in rigid rotation in vacuo. *Annales d'Astrophysique*, 18:1, January 1955.

F. D. Drake and H. D. Craft. Second Periodic Pulsation in Pulsars. Nature, 220:231–235, October 1968.

R. T. Edwards and B. W. Stappers. Pulse-to-pulse intensity modulation and drifting subpulses in recycled pulsars. A&A, 407:273–287, August 2003.

R. T. Edwards, G. B. Hobbs, and R. N. Manchester. TEMPO2, a new pulsar timing package - II. The timing model and precision estimates. MNRAS, 372:1549–1574, November 2006.

T. Enßlin. Information field theory. In U. von Toussaint, editor, *American Institute of Physics Conference Series*, volume 1553 of *American Institute of Physics Conference Series*, page 184–191, August 2013.

T. A. Enßlin, M. Frommert, and F. S. Kitaura. Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. Phys. Rev. D, 80(10):105005, November 2009.

L. Euler. *Institutionum calculi integralis*. Number Bd. 1 in Institutionum calculi integralis. imp. Acad. imp. Saènt., 1768.

L. A. Fowler, D. Morris, and G. A. E. Wright. Unusual properties of the pulsar PSR 1822-09. A&A, 93:54–61, January 1981.

G. Ghisellini. *Radiative Processes in High Energy Astrophysics*, volume 873 of *Lecture Notes in Physics*. Springer, 2013.

M. V. Goldman. Strong turbulence of plasma waves. *Reviews of Modern Physics*, 56:709–735, October 1984.

P. Goldreich and W.H. Julian. Pulsar Electrodynamics. *ApJ*, 157:869, August 1969.

P. L. Gonthier, M. G. Baring, M. T. Eiles, Z. Wadiasingh, C. A. Taylor, and C. J. Fitch. Compton scattering in strong magnetic fields: Spin-dependent influences at the cyclotron resonance. Phys. Rev. D, 90(4):043014, August 2014.

M. Greiner, D. H. F. M. Schnitzeler, and T. A. Ensslin. Tomography of the Galactic free electron density with the Square Kilometer Array. *ArXiv e-prints*, December 2015.

I. A. Grenier and A. K. Harding. Gamma-ray pulsars: A gold mine. *Comptes Rendus Physique*, 16:641–660, August 2015.

T. H. Hankins and J. A. Eilek. Radio Emission Signatures in the Crab Pulsar. ApJ, 670:693–701, November 2007.

T. H. Hankins, J. S. Kern, J. C. Weatherall, and J. A. Eilek. Nanosecond radio bursts from strong plasma turbulence in the Crab pulsar. *Nature*, 422:141–143, March 2003.

T. H. Hankins, J. M. Rankin, and J. A. Eilek. What is the Physics of Pulsar Radio Emission? In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, 2009.

D. J. Helfand, R. N. Manchester, and J. H. Taylor. Observations of pulsar radio emission. III - Stability of integrated profiles. ApJ, 198:661–670, June 1975.

A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins. Observation of a Rapidly Pulsating Radio Source. *Nature*, 217:709–713, February 1968.

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

E. W. Hones, Jr. and J. E. Bergeson. Electric Field Generated by a Rotating Magnetized Sphere. J. Geophys. Res., 70:4951–4958, October 1965.

A. W. Hotan, W. van Straten, and R. N. Manchester. PSRCHIVE and PSRFITS: An Open Approach to Radio Pulsar Data Storage and Analysis. PASA, 21:302–309, 2004.

A. W. Hotan, M. Bailes, and S. M. Ord. PSR J0737-3039A: baseband timing and polarimetry. MNRAS, 362:1267–1272, October 2005.

M. Imgrund and A. Arth. *Github repository for Rambrain*. `https://github.com/mimgrund/rambrain/`, August 2015a.

M. Imgrund and A. Arth. *Daily auto-generated documentation for Rambrain*. `http://mimgrund.github.io/rambrain/`, August 2015b.

M. Imgrund, D. J. Champion, M. Kramer, and H. Lesch. A Bayesian method for pulsar template generation. MNRAS, 449:4162–4183, June 2015.

Maximilian Imgrund. On The Two-Stream Instability In Pulsar Magnetospheres. Master's thesis, 2012.

J. D. Jackson. *Classical Electrodynamics, 3rd Edition*. July 1998.

E. T. Jaynes and G. L. Bretthorst. *Probability Theory*. April 2003.

A. Jessner, H. Lesch, and T. Kunzl. Charge Densities above Pulsar Polar Caps. *ApJ*, 547: 959–966, February 2001.

A. Jessner, M. V. Popov, V. I. Kondratiev, Y. Y. Kovalev, D. Graham, A. Zensus, V. A. Soglasnov, A. V. Bilous, and O. A. Moshkina. Giant pulses with nanosecond time resolution detected from the Crab pulsar at 8.5 and 15.1 GHz. A&A, 524:A60, December 2010.

S. Johnston and D. Galloway. Pulsar braking indices revisited. MNRAS, 306:L50–L54, July 1999.

E. F. Keane. Radio pulsar variability. In J. van Leeuwen, editor, *Neutron Stars and Pulsars: Challenges and Opportunities after 80 years*, volume 291 of *IAU Symposium*, page 295–300, March 2013.

O. Klein and T. Nishina. Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. *Zeitschrift fur Physik*, 52:853–868, November 1929.

M. Kramer, A. G. Lyne, J. T. O'Brien, C. A. Jordan, and D. R. Lorimer. A Periodically Active Pulsar Giving Insight into Magnetospheric Physics. *Science*, 312:549–551, April 2006a.

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

K. Krzeszowski, O. Maron, A. Słowikowska, and A. Jessner. Flicker noise pulsar radio spectra. *ArXiv e-prints*, June 2014.

S. R. Kulkarni. Self-noise in interferometers - Radio and infrared. AJ, 98:1112–1130, September 1989.

T. Kunzl, H. Lesch, A. Jessner, and A. von Hoensbroech. On Pair Production in the Crab Pulsar. ApJ, 505:L139–L141, October 1998.

K. J. Lee, C. G. Bassa, G. H. Janssen, R. Karuppusamy, M. Kramer, K. Liu, D. Perrodin, R. Smits, B. W. Stappers, R. van Haasteren, and L. Lentati. Model-based asymptotically optimal dispersion measure correction for pulsar timing. MNRAS, 441:2831–2844, July 2014.

J. C. Lemm. Bayesian Field Theory: Nonparametric Approaches to Density Estimation, Regression, Classification, and Inverse Quantum Problems. *ArXiv Physics e-prints*, December 1999.

L. Lentati, P. Alexander, M. P. Hobson, F. Feroz, R. van Haasteren, K. J. Lee, and R. M. Shannon. TEMPONEST: a Bayesian approach to pulsar timing analysis. MNRAS, 437:3004–3023, January 2014.

H. Lesch, A. Jessner, M. Kramer, and T. Kunzl. On the possibility of curvature radiation from radio pulsars. A&A, 332:L21–L24, April 1998.

M.H. Ligh, A. Case, J. Levy, and A. Walters. *The Art Of Memory Forensics*. Wiley, 2014.

K. Liu, J. P. W. Verbiest, M. Kramer, B. W. Stappers, W. van Straten, and J. M. Cordes. Prospects for high-precision pulsar timing. MNRAS, 417:2916–2926, November 2011.

K. Liu, E. F. Keane, K. J. Lee, M. Kramer, J. M. Cordes, and M. B. Purver. Profile-shape stability and phase-jitter analyses of millisecond pulsars. MNRAS, 420:361–368, February 2012.

D. R. Lorimer and M. Kramer. *Handbook of Pulsar Astronomy*. December 2004.

A. Lyne and F. Graham-Smith. *Pulsar Astronomy*. Cambridge University Press, March 2012.

A. Lyne, G. Hobbs, M. Kramer, I. Stairs, and B. Stappers. Switched Magnetospheric Regulation of Pulsar Spin-Down. *Science*, 329:408, July 2010.

A. G. Lyne and B. J. Rickett. Measurements of the Pulse Shape and Spectra of the Pulsating Radio Sources. Nature, 218:326–330, April 1968.

O. Löhmer, A. Jessner, M. Kramer, R. Wielebinski, and O. Maron. Observations of pulsars at 9 millimetres. A&A, 480:623–628, March 2008.

R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs. The Australia Telescope National Facility Pulsar Catalogue. AJ, 129:1993–2006, April 2005.

A. McCann. Detection of the Crab Pulsar with VERITAS above 100 GeV. *International Cosmic Ray Conference*, 7:208, 2011.

D. L. Meier, S. Koide, and Y. Uchida. Magnetohydrodynamic Production of Relativistic Jets. *Science*, 291:84–92, January 2001.

D. B. Melrose. Coherent Radio-Emission Mechanisms for Pulsars. In T. H. Hankins, J. M. Rankin, and J. A. Gil, editors, *IAU Colloq. 128: Magnetospheric Structure and Emission Mechanics of Radio Pulsars*, page 306, 1992.

D. B. Melrose and M. E. Gedalin. Relativistic Plasma Emission and Pulsar Radio Emission: A Critique. *ApJ*, 521:351–361, August 1999.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys., 21:1087–1092, June 1953.

S. Meyers. *Effective C++ Digital Collection: 140 Ways to Improve Your Programming*. Pearson Education, 2012.

F. C. Michel. The state of pulsar theory. *Advances in Space Research*, 33:542–551, 2004.

A. G. Muslimov and A. K. Harding. High-Altitude Particle Acceleration and Radiation in Pulsar Slot Gaps. ApJ, 606:1143–1153, May 2004.

M. Ng and J. Pan. Approximate Inverse Circulant-plus-Diagonal Preconditioners for Toeplitz-plus-Diagonal Matrices. *SIAM Journal on Scientific Computing*, 32(3):1442–1464, 2010.

A. Noutsos, C. Sobey, V. I. Kondratiev, P. Weltevrede, J. P. W. Verbiest, A. Karastergiou, M. Kramer, M. Kuniyoshi, A. Alexov, R. P. Breton, A. V. Bilous, S. Cooper, H. Falcke, J.-M. Grießmeier, T. E. Hassall, J. W. T. Hessels, E. F. Keane, S. Osłowski, M. Pilia, M. Serylak, B. W. Stappers, S. ter Veen, J. van Leeuwen, K. Zagkouris, K. Anderson, L. Bähren, M. Bell, J. Broderick, D. Carbone, Y. Cendes, T. Coenen, S. Corbel, J. Eislöffel, R. Fender, H. Garsden, P. Jonker, C. Law, S. Markoff, J. Masters, J. Miller-Jones, G. Molenaar, R. Osten, M. Pietka, E. Rol, A. Rowlinson, B. Scheers, H. Spreeuw, T. Staley, A. Stewart, J. Swinbank, R. Wijers, R. Wijnands, M. Wise, P. Zarka, and A. van der Horst. Pulsar polarisation below 200 MHz: Average profiles and propagation effects. A&A, 576:A62, April 2015.

S. Osłowski, W. van Straten, G. B. Hobbs, M. Bailes, and P. Demorest. High signal-to-noise ratio observations and the ultimate limits of precision pulsar timing. MNRAS, 418:1258–1271, December 2011.

F. Pacini. Energy Emission from a Neutron Star. Nature, 216:567–568, November 1967.

G. Pelletier, H. Sol, and E. Asseo. Magnetized Langmuir wave packets excited by a strong beam-plasma interaction. Phys. Rev. A, 38:2552–2563, September 1988.

S. A. Petrova. Axisymmetric force-free magnetosphere of a pulsar - II. Transition from the self-consistent two-fluid model. MNRAS, 446:2243–2250, January 2015.

J. R. Pierce. Limiting Stable Current in Electron Beams in the Presence of Ions. *Journal of Applied Physics*, 15:721–726, October 1944.

Wesley C. Reiley and Robert A. van de Geijn. POOCLAPACK: Parallel Out-of-Core Linear Algebra Package. Technical report, Austin, TX, USA, 1999.

B. J. Rickett. Amplitude-modulated noise - an empirical model for the radio radiation received from pulsars. ApJ, 197:185–191, April 1975.

B. J. Rickett. Interstellar scattering and scintillation of radio waves. ARA&A, 15:479–504, 1977.

G Rodrigues. Taming the OOM killer. *LWN.net*, February 2009.

M. A. Ruderman and P. G. Sutherland. Theory of pulsars - Polar caps, sparks, and coherent microwave radiation. ApJ, 196:51–72, February 1975.

John K. Salmon and Michael S. Warren. Parallel, Out-of-Core Methods for N-body Simulation. In *PPSC*. SIAM, 1997.

H. Schamel. Stationary solutions of the electrostatic Vlasov equation. *Plasma Physics*, 13: 491–505, June 1971.

H. Schamel and A. Luque. Kinetic theory of periodic hole and double layer equilibria in pair plasmas. *New Journal of Physics*, 7:69, February 2005.

P. A. G. Scheuer. Amplitude Variations in Pulsed Radio Sources. Nature, 218:920–922, June 1968.

R. M. Shannon, S. Osłowski, S. Dai, M. Bailes, G. Hobbs, R. N. Manchester, W. van Straten, C. A. Raithel, V. Ravi, L. Toomey, N. D. R. Bhat, S. Burke-Spolaor, W. A. Coles, M. J. Keith, M. Kerr, Y. Levin, J. M. Sarkissian, J.-B. Wang, L. Wen, and X.-J. Zhu. Limitations in timing precision due to single-pulse shape variability in millisecond pulsars. MNRAS, 443:1463–1481, September 2014.

S. L. Shemar and A. G. Lyne. Observations of pulsar glitches. MNRAS, 282:677–690, September 1996.

S. Shibata. Magnetosphere of the rotation-powered pulsar - A DC circuit model. ApJ, 378: 239–254, September 1991.

S. R. Spangler. Observations of diffractive interstellar scintillation phenomena. In J. M. Cordes, B. J. Rickett, and D. C. Backer, editors, *Radio Wave Scattering in the Interstellar Medium*, volume 174 of *American Institute of Physics Conference Series*, page 32–46, 1988.

D. S. Spicer. Magnetic energy storage and conversion in the solar atmosphere. Space Sci. Rev., 31:351–435, 1982.

I. H. Stairs, A. G. Lyne, and S. L. Shemar. Evidence for free precession in a pulsar. Nature, 406: 484–486, August 2000.

D. Stinebring. Effects of the interstellar medium on detection of low-frequency gravitational waves. *Classical and Quantum Gravity*, 30(22):224006, November 2013.

S. J. Sturner. Electron Energy Losses near Pulsar Polar Caps. ApJ, 446:292, June 1995.

P. A. Sturrock. A Model of Pulsars. *ApJ*, 164:529, March 1971.

Jianqi Tang, Binxing Fang, Mingzeng Hu, and Hongli Zhang. A Parallel Out-of-core Computing System Using PVFS for Linux Clusters. In *Proceedings of the International Workshop on Storage Network Architecture and Parallel I/Os*, SNAPI '04, page 33–39, New York, NY, USA, 2004. ACM.

T. M. Tauris and E. P. J. van den Heuvel. *Formation and evolution of compact stellar X-ray sources*, page 623–665. April 2006.

J. H. Taylor. Pulsar Timing and Relativistic Gravity. *Royal Society of London Philosophical Transactions Series A*, 341:117–134, October 1992.

J. H. Taylor and J. M. Cordes. Pulsar distances and the galactic distribution of free electrons. ApJ, 411:674–684, July 1993.

J. H. Taylor, R. N. Manchester, and G. R. Huguenin. Observations of pulsar radio emission. I - Total-intensity measurements of individual pulses. ApJ, 195:513–528, January 1975.

S. R. Taylor, J. R. Gair, and L. Lentati. Using Swarm Intelligence To Accelerate Pulsar Timing Analysis. *ArXiv e-prints*, October 2012.

A. Tchekhovskoy, A. Spitkovsky, and J. G. Li. Time-dependent 3D magnetohydrodynamic pulsar magnetospheres: oblique rotators. MNRAS, 435:L1–L5, August 2013.

A. N. Timokhin. On the force-free magnetosphere of an aligned rotator. MNRAS, 368: 1055–1072, May 2006.

A. N. Timokhin and A. K. Harding. On the Polar Cap Cascade Pair Multiplicity of Young Pulsars. ApJ, 810:144, September 2015.

Sivan Toledo. A survey of out-of-core algorithms in numerical linear algebra. In James M. Abello and Jeffrey Scott Vitter, editors, *External Memory Algorithms*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, page 161–179. American Mathematical Society, 1999a.

Sivan Toledo. External Memory Algorithms. chapter A Survey of Out-of-core Algorithms in Numerical Linear Algebra, page 161–179. American Mathematical Society, Boston, MA, USA, 1999b.

L. Torvalds. O_DIRECT performance impact on 2.4.18. *Newsgroup fa.linux.kernel*, May 2002.

R. van Haasteren. Accelerating pulsar timing data analysis. MNRAS, 429:55–62, February 2013.

Dimitri van Heesch. *Doxygen project webpage*. `http://www.stack.nl/~dimitri/doxygen/index.html`, December 2015.

S. J. Vigeland and M. Vallisneri. Bayesian inference for pulsar-timing models. MNRAS, 440: 1446–1457, May 2014.

Jeffrey Scott Vitter. External Memory Algorithms and Data Structures: Dealing with Massive Data. *ACM Comput. Surv.*, 33(2):209–271, June 2001.

C. Wang, D. Lai, and J. Han. Polarization changes of pulsars due to wave propagation through magnetospheres. MNRAS, 403:569–588, April 2010.

J. C. Weatherall. Modulational Instability, Mode Conversion, and Radio Emission in the Magnetized Pair Plasma of Pulsars. ApJ, 483:402–413, July 1997.

J. C. Weatherall. Pulsar Radio Emission by Conversion of Plasma Wave Turbulence: Nanosecond Time Structure. ApJ, 506:341–346, October 1998.

J. M. Weisberg and J. H. Taylor. General Relativistic Geodetic Spin Precession in Binary Pulsar B1913+16: Mapping the Emission Beam in Two Dimensions. ApJ, 576:942–949, September 2002.

P. Weltevrede and S. Johnston. Profile and polarization characteristics of energetic pulsars. MNRAS, 391:1210–1226, December 2008.

P. Weltevrede, R. T. Edwards, and B. W. Stappers. The subpulse modulation properties of pulsars at 21 cm. A&A, 445:243–272, January 2006a.

P. Weltevrede, R. T. Edwards, and B. W. Stappers. Statistics of the Drifting Subpulse Phenomenon. *Chinese Journal of Astronomy and Astrophysics Supplement*, 6(2):13–17, December 2006b.

S. M. White. Thermal effects in the ultrarelativistic two-stream instability. Ap&SS, 116: 173–187, November 1985.

Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Wiley, 1949.

R. Yuen, R. N. Manchester, M. Burgay, F. Camilo, M. Kramer, D. B. Melrose, and I. H. Stairs. Changes in Polarization Position Angle across the Eclipse in the Double Pulsar System. ApJ, 752:L32, June 2012.

V. E. Zakharov. Collapse of Langmuir Waves. *JETP*, 35:908, 1972.

Danksagung

Ich möchte Dir, Harald, vielen Dank sagen. Es war nicht das erste spannende bis waghalsige Projekt mit Dir, und so bin ich jetzt zugleich froh aber auch traurig, dass es ein Ende findet. Denn ich denke, „wir sind hier noch lange nicht fertig."

Ich danke Dir, Michael, dafür dass Du Dir Zeit für meine Arbeit und Ideen genommen hast. Ebenso danke ich Dir und dem Max-Planck-Institut für Radioastronomie für die Finanzierung meiner Arbeit.

Ganz besonders möchte ich mich bei meiner Frau, Katharina, und meinem Sohn Konrad bedanken. Kathi, Dir gebührt die Hälfte der Arbeit, denn ohne Dich hätte ich das nicht geschafft - mit Dir dafür umso besser. Konrad, Dir gebührt die andere Hälfte. Du hast, während ich diese Arbeit geschrieben habe, atmen, laufen und sprechen gelernt und dich zu einem wunderbaren Menschen entwickelt. Das war Deine Arbeit und es hat mir große Freude bereitet, Dir dabei zu helfen. Sei Dir gewiss, Du hast mir ebenso geholfen.

Ich danke meinen Eltern für die fortwährende Unterstützung während des Studiums.

I thank my colleagues in Bonn for the great time and discussion we had, und den Kollegen und Mitarbeitern an der USM für die vielen produktiven Pausen. Es ist schade Euch als Kollegen zu verlieren, aber schön, einige Freunde gewonnen zu haben.

Ich danke Claudio Winter, Alexander Arth, Patrick Boehl, Susanne Goerke und Lukas Milles für ihre seelische Unterstützung.