

IMPROVED TECHNIQUES FOR PULSAR DATA ANALYSIS

Prajwal Voraganti
Padmanabh

MAX-PLANCK-INSTITUT FÜR RADIOASTRONOMIE

Improved techniques for pulsar data analysis

Dissertation

zur

Erlangung des Doktorgrades (*Dr. rer. nat.*)

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Prajwal VORAGANTI PADMANABH

aus

Jamshedpur, India

Bonn 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Referent: Prof. Dr. Michael Kramer
2. Referent: Prof. Dr. Frank Bertoldi
Tag der Promotion: 30.08.2021
Erscheinungsjahr: 2021

RHEINISCHE FRIEDRICH–WILHELMS–UNIVERSITÄT BONN

Abstract

by Prajwal Voraganti Padmanabh

for the degree of

Doctor rerum naturalium

Pulsars are rapidly rotating, strongly magnetised neutron stars with spin periods ranging from few milliseconds to tens of seconds. They emit beams of electromagnetic radiation that are predominantly detected as periodic pulsations at the radio end of the spectrum. Since the first pulsar detection in 1967, there have been close to 3000 such systems discovered and they have proven to be excellent tools for a variety of scientific applications. This includes areas in fundamental physics ranging from testing theories of gravity to implications for condensed matter physics via understanding the equation of state. Despite the large number of discoveries, there exist several open questions related to pulsar astronomy that can be addressed by pushing the boundary of pulsar discoveries beyond the current population census. Chapter 1 gives a detailed description of pulsar properties and motivates the need to search for pulsars. Chapter 2 describes the telescope signal chain and the methodology behind searching for new pulsars using radio telescopes. It also motivates the advantage of using interferometers for conducting large scale pulsar surveys, the scope for which has been, to date, limited.

Chapter 3 presents the pulsar searching infrastructure developed for MeerKAT and introduces the Transients and Pulsars with MeerKAT (TRAPUM) survey. This survey is targeted at sources where the probability of finding new radio pulsars is high. The TRAPUM infrastructure provides a large field of view via a multibeam beamformer that is capable of synthesising up to 864 coherent beams using the full array of MeerKAT dishes. The highlight is the use of open source tools like `Kubernetes` to build an automated cloud computing setup that can process multi-beam data in quasi real-time. In order to test the efficiency of the processing pipelines, a multi-beam search was conducted on globular cluster Terzan 5 and the results are discussed. Although no new pulsars were discovered, an upper limit of $\sim 18 \mu\text{Jy}$ was placed on the flux density of a possible new pulsar in the cluster. This limit holds for binary pulsars in circular orbits with orbital periods above 5 hours and whose companion masses are below $4 M_{\odot}$. Besides this, advantages of the multi-beam Terzan 5 data are put forth for conducting future searches on this globular cluster.

Chapter 4 introduces the MPIfR Galactic Plane survey, which is the most sensitive Galactic plane survey conducted in the Southern Hemisphere. This

multi-purpose survey serves a wide range of scientific fields including pulsars, transients, Galactic magnetic fields, continuum and spectral line studies. A major part of the survey is aimed at using the S-Band receivers (operating between 1.7 to 3.5 GHz) built at the MPIfR to probe deeper into the Galactic plane as well as the Galactic centre region. From the pulsar searching aspect, simulations show that 476 ± 23 and 220 ± 15 new pulsars are expected to be discovered at L-Band (1.284 GHz) and S-Band (2.4 GHz). A 24-hour pilot survey was conducted at L-Band (centered at 1284 MHz) on survey fields that have been previously imaged by MeerKAT. While candidate discoveries await confirmation, the known pulsar redetections show reasonable detection significance as expected from the radiometer equation. The survey coverage achieved assuming a minimum fractional telescope gain of 0.5 is 78 per cent.

Chapter 5 describes a collaborative effort between MPIfR and multinational software company SAP for pulsar candidate classification on a large scale. Toolkits were developed for deployment of pulsar candidate classifiers on a large scale using the SAP Data Intelligence platform. As a use-case, 3.2 million candidates generated from the High Time Resolution Universe South low-lat survey were put through a pipeline developed using Data Intelligence containing multiple redetections of 60 known pulsars. The pipeline uses publicly available machine learning models (trained in different ways on different data sets) to generate scores for each candidate. The score indicates the likelihood of a candidate being a pulsar. Based on standard machine learning evaluation metrics, the models were compared against each other to test their robustness. The most robust model is shown to be a classifier whose training set contained pulsar detections with a wide range of spin periods and duty cycles. This model is reported to have an area under the receiver operating characteristic (ROC) curve of 0.988, which for an ideal classifier would be 1.0.

Chapter 6 presents a detailed analysis of the pulse profile instability of PSR J1022+1001, a pulsar that is regularly timed by pulsar timing arrays with the aim of detecting nanohertz gravitational waves. The pulsar has a rich and conflicting history of results pertaining to its purported profile instability and its likely cause. In this analysis, data spanning 20 years including two different recording instruments from the Effelsberg telescope and a recording instrument from the Parkes telescopes were used. Previously used methods of characterising the profile stability were modified and adapted to improve their robustness. The results show a significant amount of long term instability in the pulse shape. By accounting for instrumental effects like polarisation miscalibration and propagation effects like interstellar scintillation, up to 25 per cent of the instability could be explained. This indicates that the observed instability is most likely intrinsic to the pulsar.

To my family and friends

In science it often happens that scientists say, 'You know that's a really good argument; my position is mistaken,' and then they would actually change their minds and you never hear that old view from them again. They really do it. It does not happen as often as it should, because scientists are human and change is sometimes painful. But it happens every day. I cannot recall the last time something like that happened in politics or religion.

Carl Sagan

To work alone you have the right, and not to the fruits. Do not be impelled by the fruits of work. Nor have attachment to inaction.

Chapter 2, Verse 47
Translated from the Bhagvad Gita

It is our choices, Harry, that show what we truly are, far more than our abilities.

Albus Dumbledore
Harry Potter and the Chamber of Secrets

Acknowledgements

This four year long journey has been a roller-coaster ride with an unimaginable number of unexpected turns and twists. However, I would not have managed to accept and enjoy the challenge if not for some wonderful people in my life.

I first offer my heartfelt gratitude to Michael for giving me an opportunity to pursue a PhD at MPIfR. Your knowledge, attitude and work ethic is awe-inspiring. I remember meeting you on the first day which was nothing short of a celebrity moment (having already been a fan of the handbook). Throughout these years, you have always provided encouragement and solutions at important junctions of my PhD. I am also grateful that you gave me an opportunity to conduct quizzes in the Christmas parties despite having your own agenda. Finally, you have been a constant source of motivation and optimism in the midst of the pandemic via the weekly emails filled with trivia and humour.

About my day-to-day supervisor, Ewan, I have no idea where to begin. If I need to sum up this journey in one sentence I would say this - "From not understanding any of your stupendous ideas (layered with a Scottish accent), to eventually building something that works, I have learnt so much". I consider myself extremely fortunate to having had you as my supervisor. Its not only the insane depth of your knowledge and skills (on literally any topic under the sun and above!) but most importantly your patience and knack of teaching that has significantly simplified my life. Your approach to problem solving fascinates me till date. Thanks for being *pure dead brilliant*.

To Ramesh, I want to thank you for encouraging me to apply to MPIfR for a PhD back when I contacted you in my undergraduate days. You have always backed me and given valuable inputs at important times.

To David, my sincerest thanks for providing me a fresh new project at a crucial point and also for some reassuring words in times of adversity. I owe a lot to you for improving my communication and giving me a reminder every now and then about meeting goals without getting lost in the details.

I thank Frank Bertoldi for giving valuable inputs during every TAC meeting.

I have shared some memorable times with colleagues at MPIfR and am grateful for all the help they provided throughout my PhD. First of all, my office-mate Vishnu. From pair programming to enjoying Disneyland, and of course the never ending thirst of Rafa fandom (culminating in eventually watching him live!), its been a crazy ride. I hope that Rafa does become the GOAT (is he not already?) and you manage to mine enough Bitcoins to buy us an outing with him. To Hans, you have been like a big brother to me and I will forever cherish the Badminton tournament in Utrecht. I still believe that we should have split the Gold. To the Badminton group: Hans, Jens, Henning, Gautam and Rohit, thanks for keeping my love for the sport alive all these years.

To Alessandro, for being such a genuine person and a great teacher - *Grazie!* To Jason and Ann, for being so welcoming at your wedding and of course discussing Tennis! To Paulo, for being the most enthusiastic pulsar person I have ever met. I truly believe

that each pulsar is one of your personal friends (the closest ones being binaries with a possible Shapiro delay measurement). To Eleni and Natasha, for always clarifying my doubts. To Mary, from explaining everything about IMPRS on my first day to guiding me with thesis submission in these last few months, *Gracias!* To Tilemachos, for showing me around in Thessaloniki, *Sas efcharisto!*

Weiwei, thank you for picking me up at the train station when I first arrived in Bonn and also helping me move into my place. I will never forget the day we conquered Table Mountain after being totally ignorant about choosing the toughest route! Madhuri, thanks for helping me get adjusted to life in Germany and being a pillar of strength during some difficult times. Maitraiye, thanks for finishing a point short of me in the World cup betting league. Arshia, thanks for interesting conversations and assisting me in dealing with thesis related matters. Shalini, thanks for making me feel more Bengali at times and for interesting conversations on music.

Vivek, I cannot thank you enough for your guidance on all fronts. Be it research, job applications or cracking mokkas. I have a lot of respect for you as a scientific researcher and as a person. *Rumba thanks Thala!* Adi, from the amazing hikes to the Samosa driven meetups (a shout out to Pikky too!), we have had some very memorable times. Your larger than life attitude was much needed in this pandemic phase and I cannot thank you enough for also letting me work from your place at times (In fact I am writing this from your place!).

I thank Ewan, David, Ramesh, Paulo, Laura, Robert Wharton, Vivek and Michael for proof reading various parts of my thesis and giving valuable comments.

Thanks to Kira, Mariya and Le for being so kind and helpful when it came to administrative issues. Thanks to Jan and Marcel from the IT section for solving many computer related issues. Thanks to Rainer and Simone for enabling all IMPRS related activities.

To the family I have made after coming to Germany: Hem, Bhavya, Maya, Abhi, Onkar, Sonia and Siri. From rehearsing for that once in a lifetime Mime performance to numerous meetups since, all of you have been so welcoming and at times taken care of me like a child. I am forever indebted to you all. A special mention to someone without whom I would not have known all of you: Mandeep. A shout out to Khushi and aunty for hosting us several times, the icing on the cake being the absolutely sumptuous food.

To Tushita, for being a wonderful friend, co-Bangalorean and talking some sense into me when needed the most. Thanks kano!

To my friends since high school: Arjun, Sashank, Hari, Aashish and Adi Nagaraj. Thanks for being a constant source of entertainment via mindless banter and out-of-the-box thinking (as we have named our group!). I thank Arjun in particular for being great company in high school as well as undergraduate days. As for Sashank and Hari, no matter what lies ahead, Rafa is already the GOAT in our hearts. VAMOS!

To Thejas, for always being there for me as a mentor and a friend in need since school days. Thanks kano!

To Srinivas, for various tips and suggestions when looking out for job opportunities. Thanks kano!

To my mentors back in India: I thank Avinash Deshpande for introducing me to pulsars and Joydeep Bagchi for thought provoking scientific conversations. Special mention to Sanjay and Kaushal for being more of a friend than a mentor. To my music guru, Sagar Gururaj. Thank you for bringing back music into my life and being such a wonderful mentor.

The lives of two sportsmen have always inspired me to improve constantly at work and as a person: Rafael Nadal and Rahul Dravid.

And finally, whoever I am today is solely because of my family: Amma, Nana, Sushma, Shivani, Ravi mama, Anu aunty, Ajji and not to forget Hachi. To Ravi mama, for always believing in me and instilling a disciplined way of living. To Sushma and Shivani, for being the best sisters I could have ever asked for. Thank you for being with me through thick and thin especially in the thesis writing phase. I end my acknowledgements by thanking the two most important people in my life: my parents. Amma and Nana, thank you for giving me the freedom to pursue a career of my choice.

Contents

1	Introduction to Pulsars	19
1.1	A brief history	19
1.2	Pulsar properties	20
1.2.1	Formation	20
1.2.2	Structure and composition	21
1.2.3	Electrodynamics and emission	21
1.2.4	Polarisation	23
1.2.5	Spin evolution	24
1.2.6	Magnetic Field	24
1.2.7	Characteristic age	25
1.2.8	Sub-populations	26
1.3	Effects of the Interstellar Medium	32
1.3.1	Dispersion	32
1.3.2	Scattering	33
1.3.3	Scintillation	34
1.3.4	Faraday Rotation	35
1.4	Motivation for pulsar searching	36
1.4.1	Improving the understanding of the neutron star structure	36
1.4.2	Testing general relativity and alternative theories of gravity	37
1.4.3	Detecting gravitational waves using pulsar timing arrays	38
1.4.4	Probing the Galactic centre region	40
1.4.5	Improving Galactic pulsar population models	40
1.4.6	Understanding different millisecond pulsar evolution mechanisms	41
1.4.7	Discovering sources beyond the current pulsar census	43
1.5	Thesis outline	43
2	Finding pulsars	45
2.1	Radio telescope instrumentation for pulsar data acquisition	46
2.2	Pulsar search procedure	49
2.2.1	Data storage	49
2.2.2	Interference mitigation techniques	49
2.2.3	Dedispersion	52
2.2.4	Barycentering	58
2.2.5	Discrete Fourier Transform	58
2.2.6	Red noise removal	60
2.2.7	Harmonic Summing	60
2.2.8	Folding	63
2.2.9	Candidate Sifting	65
2.2.10	False alarm probability and minimum S/N	66

2.2.11	Acceleration searching	68
2.2.12	Other search techniques	75
2.3	Pulsar Software	78
2.4	Overview of Pulsar Surveys	80
2.4.1	Large area surveys	80
2.4.2	Targeted surveys	83
2.5	Pulsar surveys with interferometers	84
3	Pulsar Searching with MeerKAT	87
3.1	Introduction to MeerKAT	88
3.2	System description	89
3.2.1	Correlator/Beamformer (CBF)	89
3.2.2	FBFUSE	91
3.2.3	APSUSE	92
3.3	Introduction to TRAPUM	94
3.4	An autonomous high-speed pulsar search pipeline	96
3.4.1	Background	96
3.4.2	Computing infrastructure	98
3.4.3	Processing pipelines	103
3.5	A case study with Terzan 5	108
3.5.1	Introduction to Terzan 5	108
3.5.2	Overview of radio pulsar searches with Ter5	110
3.5.3	Benefits of TRAPUM for Ter5 searches	111
3.5.4	Observations and setup	112
3.5.5	Sensitivity	113
3.5.6	Search strategy and processing	115
3.5.7	Results	120
3.5.8	Discussion	122
3.5.9	Conclusions	123
4	The MPIfR Galactic Plane Survey: System setup and early results	125
4.1	Introduction	125
4.2	Survey strategy	126
4.3	Instrumentation	130
4.3.1	Front-end	130
4.3.2	Back-end	131
4.4	Sensitivity	131
4.5	Simulations	135
4.6	Processing	136
4.6.1	Candidate viewing	138
4.7	Pilot survey	139
4.7.1	Search strategy and processing	139
4.7.2	Results	142
4.8	Conclusions and Future Prospects	148

5	Pulsar candidate classification using SAP Data Intelligence	149
5.1	Abstract	149
5.2	Introduction	150
5.3	Machine Learning	151
5.3.1	Artificial Neural Networks	152
5.3.2	Common evaluation metrics	153
5.3.3	Feature Selection	155
5.3.4	Overview of ML classifiers for pulsar candidate classification	157
5.4	Candidate processing schemes for current and next generation telescopes	158
5.5	SAP Data Intelligence	159
5.6	Data preparation	160
5.7	Processing	162
5.8	Preliminary results	162
5.9	Discussion and Future work	165
5.10	Conclusions	168
6	Revisiting profile instability of PSR J1022+1001	169
6.1	Abstract	171
6.2	Introduction	171
6.3	Previous profile stability studies of J1022+1001	172
6.4	Observations and Data Reduction	174
6.5	Methods	176
6.5.1	Previous methods	176
6.5.2	Peak optimiser	176
6.5.3	Modified Hotan method	178
6.6	Results	179
6.6.1	EPOS	179
6.6.2	CPSR2	181
6.6.3	PSRIX	183
6.6.4	Comparing telescope data set distributions through KS test	183
6.7	Scintillation based analysis	183
6.8	Discussion	187
6.9	Conclusions	188
7	Summary and Future Prospects	191
7.1	Summary	191
7.2	Future work and prospects	192
7.2.1	Enhancements to the TRAPUM computing infrastructure	192
7.2.2	Pulsar searches of Terzan 5 using TRAPUM	194
7.2.3	Continuation of the MGPS survey	194
7.2.4	Expansion of candidate datasets and classifier models for SAP Data Intelligence	195
7.2.5	Extension of profile instability analyses and implications on timing	195
7.3	Preparing for the SKA	196

List of Figures

1.1	Pulsar beam model visualised	23
1.2	Flux beam map of PSR J1906+0746	24
1.3	The P-Pdot diagram	27
1.4	Example of evolution of scattering of pulses across observing frequency .	34
1.5	Various neutron star equations of state as a function of mass and radius	37
1.6	Detection limits across the gravitational wave spectrum	39
1.7	Standard formation channel for double neutron star systems	42
2.1	Instrumentation for pulsar data acquisition	48
2.2	Bandpass at L-Band for MeerKAT	50
2.3	Comparison between dispersed and dedispersed profiles	53
2.4	A visual illustration of the subbanding process	57
2.5	Comparison between red noise affected and dereddened timeseries	61
2.6	Effect of pulse width on power spread across frequency harmonics in the Fourier domain	62
2.7	A diagnostic plot generated after folding	64
2.8	Illustration of difference between a known pulsar and RFI with respect to the illumination of multiple beams	67
2.9	A visual representation of the outcome from the acceleration search tech- niques.	69
2.10	A standard workflow for an acceleration search pipeline based on time domain resampling.	72
2.11	A scatter plot of all known pulsars in the Galaxy overlaid according to the Galactic coordinates	82
3.1	Image of the MeerKAT radio telescope	89
3.2	Image of the FBFUSE and APSUSE racks	92
3.3	The summarised workflow of the Kubernetes based pulsar search pipeline implemented on the APSUSE Cluster	102
3.4	An entity relationship diagram of the TRAPUM database	104
3.5	A sample user interface to the TRAPUM internal webpage of a particular target source with MeerKAT	106
3.6	The tiling of the first Terzan 5 observation	114
3.7	The tiling pattern for the second Terzan 5 run with TRAPUM	116
3.8	Comparison of theoretical sensitivity with different backends and telescopes	117
3.9	A comparison of sensitivity of MeerTIME and TRAPUM systems via the detection of Terzan 5N	119
3.10	A comparison of Ter5M detections with an ephemeris fold and the fold generated by the search pipeline	121

4.1	The survey region layout for the latest and most sensitive Galactic plane surveys	127
4.2	The expected coverage of the MGPS-Sgr A* survey	129
4.3	SEFD for L-Band and S-Band receivers at MeerKAT	132
4.4	Theoretical sensitivity comparison across different galactic plane surveys	134
4.5	Number of pulsars expected from simulations for MGPS-L	137
4.6	Number of pulsars expected from simulations for MGPS-S	137
4.7	A snapshot of the pulsar candidate viewer	138
4.8	Field 1 of pilot survey for MGPS	140
4.9	Field 2 of pilot survey for MGPS	141
4.10	The dependence of the survey beam coverage factor on telescope elevation for the pilot survey pointings	143
4.11	Heatmap of MGPS coverage as a function of telescope elevation and beam overlap	145
4.12	The tiling pattern of the pilot survey pointing where PSR J0857-4424 was detected	147
5.1	Visual representation of a single layer and multi-layer perceptron	154
5.2	A visual representation of the confusion matrix in terms of true and predicted labels for pulsar candidate classification.	156
5.3	Schematic of pipeline developed with SAP Data Intelligence	161
5.4	Cumulative histogram of classifier model scores for known pulsar detections	165
5.5	A receiver operating characteristic (ROC) curve for each of the classifier models	166
6.1	Pulse profile of PSR J1022+1001 at 1.4 GHz	177
6.2	PO results for all three data sets across 20 years	180
6.3	Phase difference and amplitude ratio between peaks for EPOS pulse profile of PSR J1022+1001 as a function of gan ratio	182
6.4	Comparison of cumulative distribution functions for CPSR2 data between uncalibrated and calibrated data.	182
6.5	MHM results for all calibrated and uncalibrated data	184
6.6	MHM plot for the data set used by Hotan et al. (2004a)	185
6.7	Comparison of cumulative distribution functions of amplitude ratios between PSRIX, EPOS data and PSRIX data with white noise.	185
6.8	Comparison of cumulative distribution functions of amplitude ratios between PSRIX and CPSR2 data after calibration.	186
6.9	Results from applying the original Hotan et al. method on CPSR2 (uncalibrated) data with 5 min long integrations as used by Hotan et al.	189
6.10	Results from applying the original Hotan et al. method on CPSR2 (uncalibrated) data with 20-30 min long integrations.	190
6.11	Results from applying the original Hotan et al. method on CPSR2 (calibrated) data with 20-30 min long integrations.	190

List of Tables

3.1	An overview of the technical specifications of the MeerKAT telescope . . .	90
3.2	A list of hardware specifications for the FBFUSE and APSUSE cluster .	93
3.3	List of structural properties of the Terzan 5 globular cluster	109
3.4	A summary of the search parameters applied on Terzan 5	118
3.5	List of redetected Terzan 5 pulsars from the acceleration searches conducted	124
4.1	The observation parameters of the three MGPS sub-survey region	128
4.2	Distribution of input parameters for pulsar population synthesis	135
4.3	Summary of known pulsar detections from the pilot survey	146
5.1	Description of different classifier models used with SAP Data Intelligence	163
5.2	Area under the curve values for all classifier models	166
6.1	Different data sets used as well as their year spans for this analysis. (N) indicates the number of epochs used for analysis	175

Recurring Acronyms in this thesis

ADC	Analog-to-Digital converter
AEI	Albert Einstein Institute
AI	Artificial Intelligence
AMQP	Advanced Message Queuing Protocol
ANN	Artificial Neural Network
APSUSE	Accelerated Pulsar Search User Supplied Equipment
AXP	Anomalous X-ray pulsar
BLUSE	Breakthrough Listen User Supplied Equipment
BW	Bandwidth
CA	Configuration Authority
CAM	Control and Monitor
CBF	Correlator/Beamformer
CNN	Convolutional Neural Network
COTS	Commercial off-the-shelf
CPU	Central Processing Unit
CRAFTS	Commensal Radio Astronomy FAST Survey
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CUDA	Compute Unified Device Architecture
DISS	Diffractional interstellar scintillation
DM	Dispersion measure
DNS	Double Neutron Star
EoS	Equation of State
EPOS	Effelsberg Pulsar Observing System
EPTA	European Pulsar Timing Array
FBFUSE	Filterbank BeamFormer User Supplied Equipment
FDAS	Fourier Domain Acceleration Search
FDMT	Fast Dispersion Measure Transform
FFA	Fast Folding Algorithm
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Arrays
FITS	Flexible Image Transport System
FRB	Fast radio burst
FoV	Field of View
FWHM	Full Width Half Maximum
GBNCC	Green Bank North Celestial Cap
GC	Globular cluster
GH-VFDT	Gaussian Hellinger Very Fast Decision Tree
GPPS	Galactic Plane Pulsar Snapshot survey
GPU	Graphical Processing Unit
GR	General Relativity
GRB	Gamma-ray burst

GSM	Global System for Mobile Communication
GW	Gravitational waves
HMXB	High mass X-ray Binary
HTRU	High Time Resolution Universe
HTRU South Lowlat	The High Time Resolution Universe South Low Latitude pulsar survey
IF	Intermediate Frequency
IPTA	International Pulsar Timing Array
IQRM	Inter-Quartile Range mitigation
ISM	Interstellar medium
JSON	JavaScript Object Notation
KAPB	Karoo Array Processor Building
LO	Local Oscillator
LoS	Line of sight
LOTAAS	LOFAR Tied-Array All-Sky Survey
LNA	Low-Noise amplifiers
LMC	Large Magellanic Cloud
LMXB	Low mass X-ray Binary
MGPS	MPIfR Galactic Plane Survey
MHM	Modified Hotan Method
ML	Machine Learning
MOI	Moment of Inertia
MPIfR	Max-Planck-Institute für Radioastronomie
MSP	Millisecond pulsar
NANOGrav	North American Nanohertz Observatory for Gravitational Wave
PALFA	Pulsar Arecibo L-band Feed Array survey
PER	Pipeline Efficiency Ratio
PFB	Polyphase Filterbank
PICS	Pulsar Image Classification System
PMPS	Parkes Multibeam Pulsar survey
PO	Peak Optimiser
PPTA	Parkes Pulsar Timing Array
PSR	Pulsar
PTA	Pulsar Timing Array
PTUSE	Pulsar Timing User Supplied Equipment
PWNe	Pulsar Wind Nebulae
RA	Right ascension
RAM	Random Access Memory
RF	Radio frequency
RFI	Radio frequency interference
RISS	Refractive interstellar scintillation
RLO	Roche Lobe Overflow
RM	Rotation measure
RRAT	Radio rotational transient
SARAO	South African Radio Astronomy Observatory

SBCF	Survey Beam Coverage Factor
SDP	Science Data Processor
SEFD	System equivalent flux density
SGR	Soft gamma repeater
Sgr A*	Sagittarius A*
SKA	Square Kilometre Array
SKARAB	Square Kilometre Array Reconfigurable Application Board
SMC	Small Magellanic Cloud
SNR	Supernova Remnant
SSB	Solar System Barycentre
STDP	Standard Deviation per phase
SQL	Structured Query Language
S/N	Signal to noise ratio
TAB	Tied-array Beam
TDAS	Time Domain Acceleration Search
TFR	Time and Frequency Reference
TRAPUM	Transients and Pulsars with MeerKAT
ToA	Time of arrival
TUSE	Transient User Supplied Equipment
USE	User Supplied Equipment
VEGAS	Versatile GBT Astronomical Spectrometer
XML	Extensible Markup Language

Telescope Acronyms

Arecibo	Arecibo observatory 305-m William E. Gordon radio telescope
ASKAP	Australian Square Kilometer Array Pathfinder
CHIME	Canadian Hydrogen Intensity Mapping Experiment
Effelsberg	Effelsberg 100-m radio telescope
FAST	Five-hundred-metre Aperture Spherical radio Telescope
Fermi-LAT	Fermi Gamma-ray Large Area Space Telescope
GBT	Robert C. Byrd Green Bank Telescope
GMRT	Giant Metrewave Radio Telescope
H.E.S.S	High Energy Stereoscopic System
LOFAR	The Low Frequency Array
MeerKAT	MeerKAT radio telescope
NICER	Neutron Star Interior Composition Explorer
Parkes	Parkes 64-m radio telescope
UTMOST	Upgraded Molonglo observatory synthesis telescope
VLA	Karl G. Jansky very large array
WSRT	Westerbork synthesis radio telescope

Numerical Constants

$$\begin{aligned}\pi &= 3.14156 \\ 1 \text{ rad} &= 57.296 \text{ degrees} \\ e &= 2.7183 \\ \log e &= 0.4343 = \ln(10)^{-1}\end{aligned}$$

Physical Constants

$$\begin{aligned}\text{Speed of light} & c = 2.9979 \times 10^{10} \text{ cm sec}^{-1} \\ \text{Gravitational constant} & G = 6.670 \times 10^{-8} \text{ dynes cm}^2 \text{ gr}^{-1} \\ \text{Planck's constant} & h = 6.626 \times 10^{-27} \text{ erg sec} \\ \text{Coulomb's constant} & k = 1 \\ \text{Boltzmann's constant} & k_B = 1.381 \times 10^{-16} \text{ erg deg}^{-1} \\ \text{Stefan-Boltzmann constant} & \sigma = 5.6704 \times 10^{-5} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ K}^{-4} \\ \text{Electron mass} & m_e = 9.110 \times 10^{-28} \text{ gr} \\ \text{Proton mass} & m_p = 1.673 \times 10^{-24} \text{ gr}\end{aligned}$$

Astronomical Constants

$$\begin{aligned}\text{Astronomical unit (1 AU)} &= 1.496 \times 10^{13} \text{ cm} \\ \text{Parsec (1 pc)} &= 3.086 \times 10^{18} \text{ cm} \\ \text{Julian light year (1 ly)} &= 9.460730472 \times 10^{17} \text{ cm} \\ \text{Julian year (1 yr)} &= 3.15576 \times 10^7 \text{ sec} \\ \text{Solar mass (1 } M_{\odot}) &= 1.989 \times 10^{27} \text{ g} \\ \text{Solar radius (1 } R_{\odot}) &= 6.960 \times 10^{10} \text{ cm} \\ \text{Solar luminosity (1 } L_{\odot}) &= 3.9 \times 10^{33} \text{ erg s}^{-1} \\ \text{Absolute magnitude of the Sun (} M_V) &= 4.77 \\ \text{Apparent magnitude of the Sun (} m_V) &= -26.7 \\ \text{Effective temperature of the Sun (} T_{\text{eff}}) &= 5770 \text{ K}\end{aligned}$$

Introduction to Pulsars

Contents

1.1	A brief history	19
1.2	Pulsar properties	20
1.2.1	Formation	20
1.2.2	Structure and composition	21
1.2.3	Electrodynamics and emission	21
1.2.4	Polarisation	23
1.2.5	Spin evolution	24
1.2.6	Magnetic Field	24
1.2.7	Characteristic age	25
1.2.8	Sub-populations	26
1.3	Effects of the Interstellar Medium	32
1.3.1	Dispersion	32
1.3.2	Scattering	33
1.3.3	Scintillation	34
1.3.4	Faraday Rotation	35
1.4	Motivation for pulsar searching	36
1.4.1	Improving the understanding of the neutron star structure	36
1.4.2	Testing general relativity and alternative theories of gravity	37
1.4.3	Detecting gravitational waves using pulsar timing arrays	38
1.4.4	Probing the Galactic centre region	40
1.4.5	Improving Galactic pulsar population models	40
1.4.6	Understanding different millisecond pulsar evolution mechanisms	41
1.4.7	Discovering sources beyond the current pulsar census	43
1.5	Thesis outline	43

1.1 A brief history

The discovery of the first pulsar in 1967 was preceded by a significant body of work that sought to explain the life-cycle of a star and the factors that determine its fate. It was known that a classical star exists in a state of hydro-static equilibrium where the radiation and gas pressure is counteracted by gravity (Eddington, 1920). The

discovery of the companion of Sirius A led to the theory of existence of quantum stars, where electron degeneracy pressure balances gravity. These stars came to be known as white dwarfs (Fowler, 1926). Baade & Zwicky (1934) first coined the term supernova, estimating that an energy on the scale of 10^{53} ergs could be released from the bulk annihilation of a star. They also hypothesised that the remnant of this explosion would be a star made of neutrons, just a few years after the discovery of the neutron (Chadwick, 1932). Hund (1936) had suggested that if the nuclei and electrons were to combine to form neutrons, the reaction would be highly endothermic. Landau (1938) built on this idea and suggested that massive stars with immense self-gravity could compensate for the loss in internal energy if neutronisation were to occur. Similar to the calculation made by Chandrasekhar (1931) for the maximum mass of a white dwarf, Oppenheimer & Volkoff (1939) calculated an equation of state for a neutron star and also included a constraint on the maximum mass a neutron star could have.

Post World War II, neutron star research did not pick up until the 1960s. Hewish & Okoye (1965) discovered an unfamiliar bright source in the direction of the Crab supernova remnant. Shklovskii (1968) studied X ray and optical observations of Scorpius X1 and hinted that emission is most likely from a neutron star undergoing accretion. Pacini (1967) suggested that a rapidly rotating neutron star with massive magnetic fields should have an electromagnetic signature and powers the Crab remnant. Concrete evidence for this theory was established when Jocelyn Bell discovered the first pulsar (Hewish et al., 1968). Despite this discovery, the link between pulsars and neutron stars was still a matter of debate (Hewish et al., 1968; Gold, 1968; Large et al., 1968a; Staelin & Reifenstein, 1968). The pulsations from the discovery were initially thought to be oscillating white dwarfs. The discovery of periodic pulsations in a 33 ms interval from the Crab pulsar (Lovell & Craft, 1968) provided the crucial evidence that pulsars are most likely highly compact neutron stars. Furthermore, the secular increase in the periodicity of pulses observed with time indicated that the pulsations manifest as a consequence of rotation rather than oscillation.

1.2 Pulsar properties

1.2.1 Formation

The mass of a typical main sequence star usually ranges from $0.1-100 M_{\odot}$. The end of the life-cycle of a star has different outcomes depending on the progenitor mass. Less massive stars follow a self-consistent mechanism for most of their lifetime showing ideal gas-like behaviour. The temperature and pressure of the system are coupled and help balance the self-gravity. Eventually, the nuclear fuel exhausts and the temperature generated is not high enough to form elements beyond Carbon via nuclear fusion. At this stage, the core is degenerate and obeys Fermi-Dirac statistics. The self-gravity of the system is now balanced by electron degeneracy pressure, thus forming He or CO white dwarfs. The self-gravity of massive stars ($> 20 M_{\odot}$) dominates at the end of the lifetime to such an extent that neither classical pressure nor any form of degeneracy pressure is able to counter the formation of a singularity.

The current consensus is that an intermediate range of stellar masses ($\sim 8\text{-}20 M_{\odot}$) provide the right conditions to form a neutron star (e.g. [van den Heuvel & Habets, 1984](#)). These stars provide the fuel and temperature to form iron based cores. Since the binding energy per nucleon is highest for iron, more massive elements can only be formed through an unfavourable endothermic process. At temperatures close to 10^{10} K, neutrinos are formed and escape with the majority of the thermal energy. This cools down the core leaving it degenerate. The mass of this iron based white dwarf increases with time and when the density of the core reaches approximately $10^{11} \text{ g cm}^{-3}$, neutronisation via inverse beta-decay occurs. The number of electrons reduce, leading to a decrease in electron degeneracy pressure. This positive feedback triggers a dynamical collapse of the system within milliseconds only to be halted by the neutron degeneracy pressure. The remnant is now a neutron star. The collapse also triggers a massive supernova explosion that produces a burst of neutrinos. This burst is also adequate to provide the energy required to form nickel and cobalt from iron. A recent review by [Cerdeira-Duran & Elias-Rosa \(2018\)](#) discusses our current understanding of core-collapse supernovae and neutron star formation.

1.2.2 Structure and composition

The neutron star formed is not uniform in composition. The outer crust is thought to be a solid iron lattice which is about a kilometre in thickness. This crust is believed to encase a core of superfluid neutrons and superconducting protons that show properties of a type II super conductor (e.g. [Baym et al., 1969](#)). An integral part of neutron star research has been to determine the relation between macroscopic quantities like temperature, pressure and density better known as the polytropic equation-of-state (EoS). While [Oppenheimer & Volkoff \(1939\)](#) made the first estimate of the EoS for neutron stars, several subsequent studies have attempted to model the properties of the inner layers (see [Sauls \(2019\)](#) for a review). Although a vast set of theories exist, there is an open debate about what the true EoS of a neutron star is. From the experimental point of view, precise mass measurements of pulsars using pulsar timing (e.g. [Ransom et al., 2014](#)) as well as optical spectroscopy ([Antoniadis et al., 2013](#)) have provided improved constraints on a multitude of EoS theories (see Section 1.4.1). Recently, the Neutron Star Interior Composition Explorer (NICER; ([Arzoumanian et al., 2012](#))) was launched with the aim of improving the understanding of the neutron star interior using X-ray timing and spectroscopy (see further discussion in Section 1.4.1).

1.2.3 Electrodynamics and emission

Pulsars are rapidly rotating neutron stars that predominantly show non-thermal emission. The equivalent blackbody temperature of the pulsar (10^{23} K) is orders of magnitude higher than the temperature at which self-absorption would take place (maximum of 10^{12} to 10^{13} K). This suggests that the emission is coherent (similar to masers) leading to orders of magnitude increase in the radiated intensity.

One of the earliest models proposed explaining the electrodynamics of pulsars is

the Goldreich-Julian model (Goldreich & Julian, 1969). From classical electromagnetic theory, it is known that a rotating magnetised conducting object would produce an equivalent electric field to counter balance the Lorentz force experienced by the charged particles of the object. This can be expressed as

$$\mathbf{E} + \frac{1}{c}(\boldsymbol{\Omega} \times \mathbf{r}) \times \mathbf{B} = 0 \quad (1.1)$$

where \mathbf{B} is the magnetic field of the pulsar, $\boldsymbol{\Omega}$ is the angular velocity of a point \mathbf{r} within the rotating sphere and \mathbf{E} is the total electric field generated.

Given the enormous magnetic field of pulsars (see Section 1.2.6), the equivalent electric field produced is of the order of 10^{12} V. The electrostatic force is sufficient to remove charged particles from the outer crust and accelerate them to ultra-relativistic speeds away from the surface of the neutron star. However, these accelerated particles do not move randomly and are confined by the magnetic field lines. Owing to the strength of the magnetic field, any charges which are gyrating about the magnetic field would quickly lose energy and drop down to the lowest Landau level. Thus, the charges now move along the curved field lines. The charges are accelerated along the field lines and produce radiation known as curvature radiation. High-energy photons produced by curvature radiation interact with the magnetic field and lower-energy photons to produce electron-positron pairs that radiate more high-energy photons resulting in a cascading process. The charges that are trapped, move along the closed field lines of the neutron star (see Figure 1.1).

The pulsar magnetosphere co-rotates with the neutron star but has a limited extent. At a certain critical distance from the neutron star, the co-rotation speed exceeds the speed of light and this distance is termed as the light cylinder radius (see Figure 1.1). Charges moving along the field lines which cross the light cylinder would escape from the pulsar. The periodic pulses observed from a pulsar have narrow widths and range from a spin frequency of 0.05-700 Hz. This indicates that the emission region is compact. The consensus is that the broadband emission is generated above the polar caps in the co-rotating magnetosphere of plasma and is beamed along the magnetic poles. Given that the magnetic axis and rotation axis need not be aligned with one another, if the beamed emission happens to fall in the line of sight of Earth, we observe this source as a radio pulsar.

Several emission models have been proposed to explain the pulsar emission properties (see Harding (2017) and references therein), but we do not fully understand the complete mechanism. Besides this, the different shapes of pulse profiles observed from different known pulsars (e.g. see Fig. 1 in Malov & Malofeev (2010)) have given rise to two main schools of thought. One interpretation suggests that the pulsar beam consists of a core and multiple conal components (Rankin, 1983). This has been shown to account for several pulse profile shapes having multiple components (e.g. Mitra & Deshpande, 1999). On the contrary, the other interpretation suggests the emission to be patchy across the beam i.e. random patches within the pulsar beam are active (Lyne & Manchester, 1988). Pulse profile studies have also given credence to the latter theory (e.g. Han & Manchester, 2001) including the recent work by Desvignes et al. (2019)

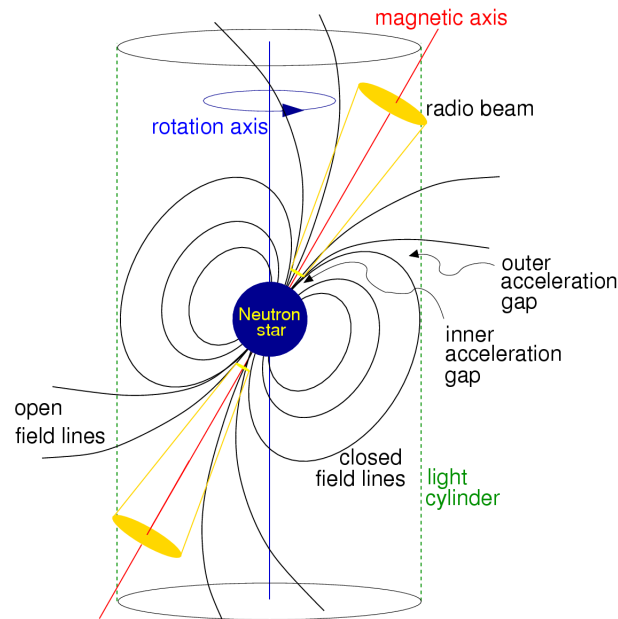


Figure 1.1: An artistic visualisation of the magnetic dipole model. Figure taken from Lorimer & Kramer (2012)

where the pulsar emission is traced across different latitudes of the projected beam (see Figure 1.2).

1.2.4 Polarisation

Polarisation refers to the orientation of electromagnetic radiation and can be classified into different categories viz. linear, circular and elliptical. Pulsars are known to show a high degree of linear polarisation (e.g. Johnston & Kerr, 2017) sometimes close to 100 percent. Additionally, the plane of polarisation of the pulse gives an insight into the geometry of the magnetic field lines (Radhakrishnan & Cooke, 1969) and also provides a mechanism to estimate the orientation of the magnetic axis with respect to the rotation axis. Some degree of circular polarisation from pulsars has also been observed. Some theories suggest that the magnetospheric plasma acts as a medium to convert linear to circular polarisation (see e.g. Lyubarskii & Petrova (1998) and references therein). However, the origin of circular polarisation is still an open question.

Pulsar polarisation also helps to understand phenomena extrinsic to the pulsar and its magnetosphere. It can serve as a probe to derive the magnetic field properties of the interstellar medium (Han et al., 1999, 2006, 2018). From the instrumental aspect, pulsars whose polarisation properties are well known can be used for polarisation calibration of the telescope. This calibration is necessary to correct for any leakage of polarised power owing to cross-coupling between receiver elements in the telescope

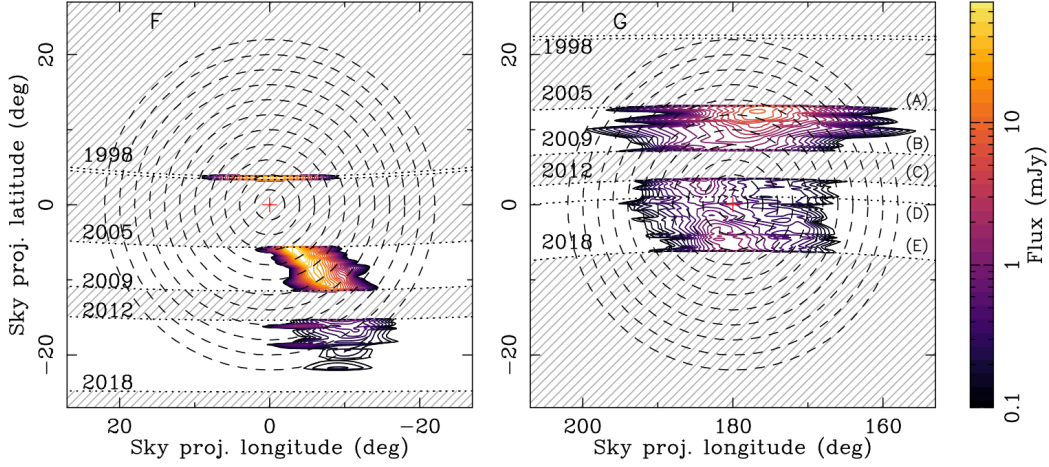


Figure 1.2: An actual map of the flux emission across the pulsar beam reconstructed for PSR J1906+0746. The two maps represent the conal emission from both magnetic poles which happen to cross the line of sight at Earth. The dotted lines represent the line of sight (LOS) across the cone for a particular year. Figure taken from [Desvignes et al. \(2019\)](#)

signal chain that could inadvertently impact accurate polarisation studies of pulsars as well as precision timing experiments ([Hamaker et al., 1996](#); [Britton, 2000](#); [van Straten, 2004, 2006, 2013](#)). A use case for this issue is discussed in Chapter 6.

1.2.5 Spin evolution

Pulsars born from core-collapse supernovae are typically born with a spin period of a few tens of milliseconds. They eventually spin down due to the loss of rotational kinetic energy given by ([Lorimer & Kramer, 2012](#)):

$$E = \frac{1}{2}I\Omega^2 \quad (1.2)$$

where I is the moment of inertia of the pulsar and Ω is the angular frequency ($\Omega = \frac{2\pi}{P}$ where P is the spin period.)

The power generated from this rotational kinetic energy (also termed as spin down luminosity) can be expressed as

$$\dot{E} \equiv -\frac{dE_{\text{rot}}}{dt} = -\frac{d(I\Omega^2/2)}{dt} = -I\Omega\dot{\Omega} = 4\pi^2 I \dot{P} P^{-3} \quad (1.3)$$

1.2.6 Magnetic Field

Owing to conservation of the magnetic flux before and after a supernova explosion (BR^2 is constant where B is the magnetic field and R is the radius of the neutron

star), a surge in the magnetic field strength is usually expected for the neutron star. [Ginzburg \(1964\)](#) calculated this number to be about 10^{12} G.

In order to model this magnetic field, we assume the pulsar to be a giant precessing bar magnet whose radiation power can be given as ([Jackson, 1962](#))

$$\dot{E}_{\text{dipole}} = \frac{2}{3c^3} |\mathbf{m}|^2 \Omega^4 \sin^2 \alpha \text{ erg s}^{-1} \quad (1.4)$$

where \mathbf{m} is the magnetic dipole moment, Ω is the spin frequency and α is the angle between the magnetic moment and the spin axis.

By equating equations 1.3 and 1.4 and utilising the fact that the magnetic field B is proportional to m/r^3 we can express the magnetic field at the surface (B_S) as ([Lorimer & Kramer, 2012](#)):

$$B_S \equiv B(r = R) = \sqrt{\frac{3c^3}{8\pi^2} \frac{I}{R^6 \sin^2 \alpha} P \dot{P}} \quad (1.5)$$

where I is the moment of inertia and R is the radius of the pulsar. Assuming certain values for I , R and α , Equation 1.5 shows that we can thus estimate the magnetic field using two observable parameters namely the spin period (P) and spin period derivative (\dot{P}). This relation serves as an important criteria to sub-categorise pulsars based on magnetic field strengths (see discussion on magnetars in section 1.2.8.1).

1.2.7 Characteristic age

Equation 1.4 is valid for an ideal scenario where dipole radiation is the only source of energy loss for a pulsar. A generalisation of this expression is to write a power law expressed as:

$$\dot{\nu} = -K\nu^n \quad (1.6)$$

where Ω is expressed in terms of the rotational frequency ν , K is a proportionality constant and n is known as the braking index. The age of a pulsar can be estimated by solving for the differential equation 1.6. It can be shown that the final expression for the age (T) is written as ([Lorimer & Kramer, 2012](#)) :

$$T = \frac{P}{(n-1)\dot{P}} \left[1 - \left(\frac{P_0}{P} \right)^{n-1} \right] \quad (1.7)$$

where ν is replaced by the spin period P and P_0 stands for the spin period at birth. Assuming that P_0 is much shorter than the present day observed spin period and the only medium of loss of energy is magnetic dipole emission ($n=3$), we can express the characteristic age (τ_c) of the pulsar as:

$$\tau_c \equiv \frac{P}{2\dot{P}} \simeq 15.8 \text{Myr} \left(\frac{P}{\text{s}} \right) \left(\frac{\dot{P}}{10^{-15}} \right)^{-1} \quad (1.8)$$

An important note here is that the expression obtained above is an approximation and does not hold for all pulsars. It is not a valid estimate for the age of young pulsars whose lives as a radio pulsar began after a recent supernova explosion (e.g. Crab). The approximation of $P_0 \ll P$ is invalid in this scenario. As will be discussed in the following subsection, some neutron stars are recycled and spun up to millisecond rotation periods with a much lower spin down rate than young pulsars. However, Equation 1.8 can account for the age post the recycling process. There also exist a class of pulsars that are not rotationally powered (see magnetars in Section 1.2.8.1) for which the age calculation does not hold. Besides this, there are significant number of pulsars where a period double derivative is non-zero hence invalidating the assumption that the period derivative is constant.

1.2.8 Sub-populations

Pulsars can be classified into various classes based on several criteria. Discussed below are a few classification schemes and their corresponding categories.

1.2.8.1 By formation and evolution

The pulsar formation mechanism discussed in Section 1.2.1 explains a fraction of the total number of pulsars visible in radio frequencies. Discussed below are canonical pulsars as well as pulsars formed via other evolution channels that lead to unique properties.

- *Canonical pulsars*

The majority of known pulsar discoveries (including the first discovery in 1967) have spin periods of the order of a few seconds and period derivatives approx. 10^{-15} s/s. They have magnetic fields of the order of 10^{12} G. These pulsars are conjectured to be born directly from core-collapse supernovae and are typically found to be isolated. The evolutions in the spin period and spin period derivative are secular and can be modelled in general. The characteristic ages of canonical pulsars are typically of the order 10^7 years. Through the course of this time, these pulsars spin down owing to magnetic dipole braking and at a certain critical point the dynamo mechanism produces a surface voltage difference that is not enough for cascade pair production. At this stage, the neutron star no longer functions as a radio pulsar. This is represented by the death line in the P-Pdot diagram (See Figure 1.3).

The applications of these pulsars are multitudinous. The pulsar emission mechanism can be studied in more detail through profile modelling (e.g. Rankin, 1993). The ensemble of these pulsars has particularly helped in population synthesis studies which give an insight into the conditions within the Galaxy that nurture the birth of these systems (e.g. Faucher-Giguère & Kaspi, 2005).

- *Recycled Pulsars*

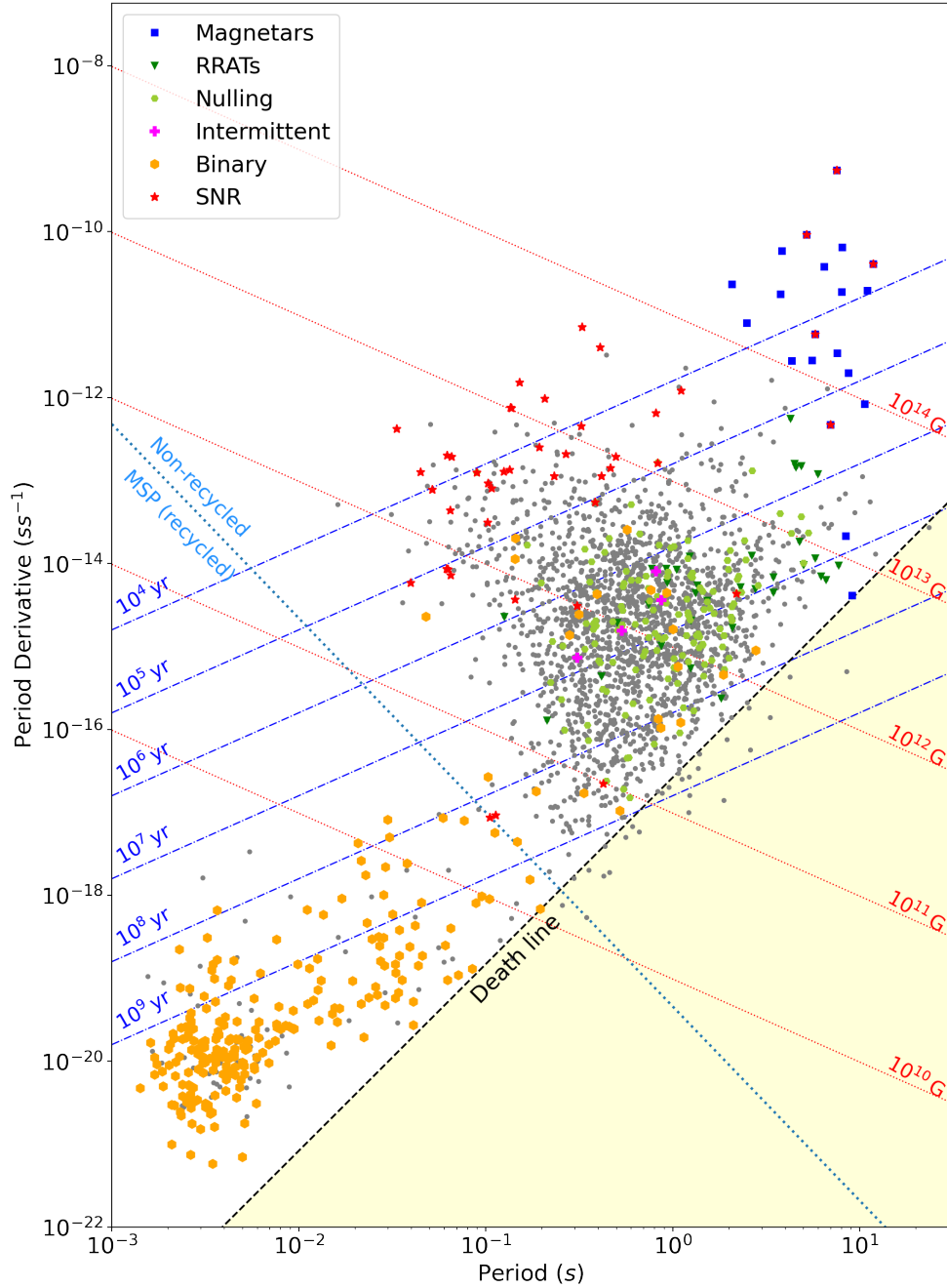


Figure 1.3: The P-Pdot diagram of the current set of all known pulsars (data taken from the ATNF catalogue (Manchester et al., 2005)). All the known pulsars are indicated as grey filled circles. Within these, each class of pulsars are demarcated by unique identifiers in terms of shape and colours. The lines of constant surface magnetic field and characteristic age are shown as the criss-crossing lines. An empirical definition for MSPs as given by Lee et al. (2012) is also plotted. The line demarcating the white and pale shades represents the death line which is model dependent. The line here is given by Equation 4 in Zhang et al. (2000). Pulsars belonging to globular clusters are significantly affected by the acceleration within the cluster. This affects the true intrinsic spin period derivative value and is hence not plotted here.

The discovery of the Hulse-Taylor binary pulsar (Hulse & Taylor, 1975b) opened up a paradoxical problem. The short spin period of 59 ms for this pulsar suggested that it must be young whereas the measured magnetic field was about 10^{10} G which was two orders of magnitude less than the value derived for known pulsars at the time. This suggested a different mechanism for pulsars to be formed.

Recycled pulsars are neutron stars in a binary system that accrete matter from a companion star which has exceeded its Roche Lobe (e.g. Davidson & Ostriker, 1973). During accretion, the infalling matter experiences the gravitational potential energy of the neutron star and emits radiation in X-rays. The matter being accreted also carries angular momentum which begins to spin up the neutron star and dampen the magnetic field. At a later point, the rotating dynamo is fast enough to restart the radio emission mechanism (hence the term recycled). The discovery of transitional millisecond pulsars provided the missing link for this theory (Archibald et al., 2009). These are pulsars which are observed to switch between accretion and radio emission phases (see e.g. Jaodand et al. (2018) and references therein). The recycled pulsars can be further classified into low mass X-ray binaries (LMXBs) and high mass X-ray binaries (HMXBs) where the low and high is determined by the companion mass. A detailed discussion on these classes can be found in Bhattacharya & van den Heuvel (1991) and Tauris & van den Heuvel (2006).

In HMXBs, the pulsars are spun up to a few tens to hundreds of milliseconds before the end of the life-cycle of the companion (the companion being a massive O or B star). If the binary does not survive the explosion, it would lead to two runaway neutron stars with a high kick velocities. However, if the binary survives the supernova explosion of the companion, double neutron star binaries and neutron star-black hole binaries can be formed. Here, one of the neutron stars is a recycled faster spinning pulsar whereas the other is a neutron star which is a product of the core-collapse supernova (see a detailed overview in Tauris & van den Heuvel (2006)). While this theory is applicable to the Hulse-Taylor pulsar, studies of the the Double Pulsar system (Burgay et al., 2003) (where pulsar A has a spin period of 22.69 ms and Pulsar B rotates with a spin period pf 2.77 s) provides evidence of an alternative mechanism where the second supernova is an outcome of electron capture on a O-Ne-Mg core (Tauris et al., 2012; Ferdman et al., 2013). Recently, Tauris et al. (2017) suggested an iron-stripped core-collapse mechanism instead.

- *Millisecond pulsars*

Millisecond pulsars (MSPs) as the name suggests are pulsars whose spin periods range in milliseconds (see e.g. empirical definition in Lee et al. (2012)). These pulsars occupy the bottom left corner of the P-Pdot diagram shown in Figure 1.3. As a consequence of their low spin period derivative (order of 10^{-19} and below), they act as precise clocks and are their stability is only second to modern day atomic clocks. Their remarkable rotational stability allows for precise measure-

ments of system parameters like the mass (e.g. [Kramer et al., 2006b](#)) as well as Post Keplerian parameters and gravity effects in binary systems (e.g. [Venkatraman Krishnan et al., 2020a](#)). The magnetic field of these MSPs is roughly of the order of 10^8 G. The consensus on MSP formation is that they are an outcome of recycling of the neutron star in LMXBs. Owing to the long life-cycle of the companion star (which is typically less massive than the neutron star) in LMXBs, the neutron star can accrete matter for a timescale of 10^7 years leading to spin periods of a few milliseconds ([Alpar et al., 1982](#)). The companion eventually forms a white dwarf system leading to a neutron star-white dwarf (NS-WD) binary. However, a significant number of isolated MSPs have been found that has led to unexplained questions (see details in Section 1.4.6).

- *Magnetars*

There exists a class of pulsars which show extremely strong magnetic fields of the order of $10^{14} - 10^{15}$ G. They spin once every 2–12 s and have larger than average spin-down rates compared to canonical pulsars, thus suggesting an enormous magnetic field strength. These objects known as magnetars are primarily powered by the decay of these magnetic fields ([Duncan & Thompson, 1992](#)). They have a characteristic emission in X-rays as well as gamma rays hence giving the terminology of Soft Gamma ray repeaters (SGRs) and Anomalous X-ray pulsars (AXPs). Magnetars are also known to produce transient events like short bursts (e.g. [Lin et al., 2013](#)), giant flares (e.g. [Ajello et al., 2021](#)) and quasi-periodic oscillations (e.g. [D’Angelo, 2013](#)). The widely accepted model for magnetars are that the enormous magnetic fields are unstable leading to stresses on the surface that could trigger these observational outbursts (e.g. [Thompson & Duncan, 1996](#)). While 30 magnetars are known, radio emission has been found in 7 systems so far ([Olausen & Kaspi, 2014](#))¹. Some of these sources show properties of rotation powered as well magnetic field powered pulsars posing open questions of their formation and evolution (e.g. [Camilo et al., 2006, 2007](#); [Levin et al., 2010](#); [Eatough et al., 2013b](#); [Lower et al., 2020](#)). A detailed review of magnetars is given in ([Kaspi & Beloborodov, 2017](#)).

1.2.8.2 By peculiar pulse emission features

- *Rapidly Rotating Transients*

[McLaughlin et al. \(2006\)](#) discovered a class of objects showing sporadic bursts of emission unlike the regular emission shown by rotation powered pulsars. Owing to this property, these objects have a higher probability of discovery if standard algorithms for pulsar searching incorporate the detection of single bright pulses rather than a persistence of the signal in time (see Chapter 2 for details on search techniques). These objects are known as Rapidly Rotating Transients (RRATs) and have spin parameters similar to canonical pulsars (as indicated by Figure 1.3).

¹<http://www.physics.mcgill.ca/~pulsar/magnetar/main.html>

There are currently more than 100 such sources known². The irregular nature of RRATs is still a matter of debate. Some theories suggest that pulsars that are nearly approaching the death line cannot generate the right conditions for sustained radio emission (Zhang et al., 2007) and thus emit sporadically. Cordes & Shannon (2008) suggested that in-falling debris from asteroids could come in the line of sight and interrupt the emission from the light cone. Some studies also suggest that some RRATs could be sources emitting continuous pulsed emission distributed over a wide range of energies and are observed as sporadic due to large distances (Weltevrede et al., 2006). Thus, it is still unclear that RRATs form a unique population of radio pulsars (e.g. Keane et al., 2011).

- *Nulling pulsars*

Unlike the sporadic nature of RRATs, there exists pulsars that show standard periodic pulsed emission for majority of the time, but the emission stops suddenly for several pulse periods. This abrupt interruption of emission is a phenomenon known as nulling. This feature was first discovered by Backer (1970) and there are currently over 200 radio pulsars that null (Konar & Deka, 2019). The nulling fraction (fraction of total pulses where no emission is detected) and null length (the duration of an instance of nulling) are the two main measures of characterising the nulling properties of a pulsar. Several studies have been conducted to understand underlying patterns of the nulling parameters for multiple pulsars and infer if behaviour of the null occurrences is random or not (e.g. Ritchings, 1976; Janssen & van Leeuwen, 2004; Kloumann & Rankin, 2010; Gajjar et al., 2012). However, no conclusive evidence has been provided favouring one over the other. Recently, Konar & Deka (2019) have analysed a large sample of nulling pulsars and shown evidence for no correlation between the nulling parameters and intrinsic pulsar parameters like spin period and magnetic field. Overall, pulsar nulling continues to be a poorly understood phenomenon.

- *Intermittent pulsars*

Kramer et al. (2006a) observed a strange phenomenon associated with PSR B1931+24. They observed a feature similar to nulling where the pulsar signal is lost. However, they observed that the timescale of the sporadic emission lasts tens of days which is much higher than nulling timescales. The other unique feature observed was the decrease in the spin down of the pulsar during the phase with no radio emission. Given their transitional nature, it is most likely that these systems form a large fraction of yet to be discovered pulsars. Although there have been subsequent discoveries of similar behaving pulsars (e.g. Lorimer et al., 2012; Camilo et al., 2012; Lyne et al., 2017) as well as mechanisms proposing the reasons behind the rapid cut off in radio emission (Kramer et al., 2006a; Mottez et al., 2013a,b; Li et al., 2014), a fundamental understanding of intermittency is an open question.

²<http://astro.phys.wvu.edu/rratalog/>

1.2.8.3 By association

- *Binary pulsars*

These contain the broad class of pulsars which are gravitationally bound with another source. The companion object could be a main sequence star (e.g. PSR J1903+0327; (Champion et al., 2008)), white dwarf (e.g. PSR J1738+0333; (Antoniadis et al., 2012)) or a neutron star (e.g. PSR J1518+4904; (Nice et al., 1996)). Given that the lifetimes of main sequence stars are typically low compared to compact degenerate stars, majority of the systems observed contain recycled pulsars which have acquired pulsar like properties due to accretion from the companion (as explained in the section on recycled pulsars). Binary systems have a wide variety of applications that are discussed in detail in Section 1.4. The binary motion of these pulsars leads to modulation effects that require a different approach while implementing pulsar search techniques. This is discussed in further detail in Chapter 2.

- *Pulsars associated with Supernova remnants*

Within canonical pulsars, there exists a subset of young pulsars which are a few thousand years old post the supernova explosion. On this timescale, the supernova remnant (SNR) is still expanding and not yet come to equilibrium with the surrounding medium. The prime examples for this are the Crab and Vela pulsars. Detailed study of the SNR along with the central pulsar engine can also be critical for better understanding the conditions leading to supernova explosions (e.g. Heger et al., 2003). They also help better constrain the galactic neutron star formation rate (Keane & Kramer, 2008). As an ensemble population, they can provide constraints on fundamental pulsar properties (e.g. Malov, 2021).

- *Pulsars in Globular clusters*

Globular clusters (GCs) are a dense collection of ($\sim 10^6$) gravitationally bound stars. The predominantly old stellar population in GCs provides an ideal environment for neutron star formation. The dense cores of these clusters allow for gravitational dynamics to have a significant impact unlike typical environments in the Galactic field leading to exchange encounters and formation of binary systems with recycled pulsars (e.g. Verbunt & Freire, 2014). This has been proven and confirmed time and again with more than 90 percent of GC pulsar discoveries found to be MSPs³. Complementary to the large number of MSPs discovered, several X-ray binaries have also been detected (see Heinke (2010) and references therein) which as explained in Section 1.2.8.1 are a key transition stage for recycling.

Discoveries of pulsars in globular clusters have had a significant influence on multiple scientific applications. Firstly, they have helped better understanding fundamental properties of neutron stars like constraining the equation of state

³see up to date list of globular cluster pulsars in <https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html>

(e.g. [Hessels et al., 2006](#)) and the physics of accretion (e.g. [Papitto et al., 2013](#)). Recently, they have also helped probe the magnetic environment of the GC as well as the Galaxy (e.g. [Abbate et al., 2020](#)) via Faraday rotation measurements (see Section 1.3.4). Additionally, the discovery of unique pulsars in these systems have been symbiotic with improved techniques for pulsar searching (e.g. [Andersen & Ransom, 2018](#)). A detailed overview of the pulsar discoveries in globular clusters via targeted surveys and their significance is discussed in Chapter 2.

1.3 Effects of the Interstellar Medium

The pulsing signal that is detected via telescopes carries a signature of its propagation through the interstellar medium (ISM). Listed below are phenomena that modify the intrinsic properties of the received pulses.

1.3.1 Dispersion

[Hewish et al. \(1968\)](#) observed that the pulsar signal was broadband and multiple observing frequencies of the received pulse did not reach the telescope at the same time. This is a well known effect of dispersion through the ISM. The ISM is predominantly composed of free electrons that can be modelled as a cold plasma. A property of this medium is a frequency dependent refractive index and the group velocity (v_g) within the medium can be written as :

$$v_g = c \sqrt{1 - \left(\frac{f_p}{f}\right)^2} \quad (1.9)$$

where f_p , denoting the plasma frequency, can be further expressed as

$$f_p = \sqrt{\frac{e^2 n_e}{\pi m_e}} \simeq 8.5 \left(\frac{n_e}{\text{cm}^{-3}}\right)^{1/2} \text{ kHz} \quad (1.10)$$

The time delay between a signal travelling in pure vacuum and the plasma of the ISM can be expressed as

$$t = \left(\int_0^d \frac{dl}{v_g}\right) - \frac{d}{c} \quad (1.11)$$

We can substitute for v_g using equation 1.9 . Further, the relation between f_p and f can be expanded by a Taylor series and equation 1.11 can be rewritten as ([Suresh & Cordes, 2019](#)):

$$t = \frac{e^2}{2\pi m_e c} \int n_e dr \frac{1}{f^2} \pm \frac{e^3}{2\pi^2 (m_e c)^2} \int n_e B dr \frac{1}{f^3} + \frac{3e^4}{8\pi^2 m_e^2 c} \int \frac{N}{n^2 dr} \frac{1}{f^4} + \text{higher order terms} \quad (1.12)$$

However, the contribution of terms with a $1/f^3$ dependence and onward can be safely neglected assuming that $f_p \ll f$. This simplifies the expression further and

substituting for the values of c , m_e and e , we can express a time difference between a broadband signal in two observing frequencies f_1 and f_2 as

$$\Delta t \simeq k \text{DM} \times (f_1^{-2} - f_2^{-2}) \quad (1.13)$$

where $k = 4.15 \times 10^6$ ms and DM is known as the dispersion measure and represents a measure of the integrated column density of free electrons along the line of sight.

$$\text{DM} = \int_L n_e dl \quad (1.14)$$

where n_e is the free electron density and L is the path length between observer and pulsar.

Once the DM value is accurately measured, it is important to remove the dispersive effect on the signal which would otherwise smear out and lower the S/N of the received signal. The process of removing this effect is called dedispersion and is described in detail in Chapter 2.

Dispersion plays an important role in pulsar studies. DM values from pulsars can help build and improve the Galactic free electron density models (e.g. Cordes & Lazio, 2002; Yao et al., 2017). Given the spatial variability of the ISM, the measured DM value can fluctuate with time. Thus, a careful understanding of variations in DM is also important for precision pulsar timing experiments (e.g. Donner et al., 2020).

1.3.2 Scattering

Some pulse profiles are observed to have an exponential tail. This is an effect of the intrinsic pulsar signal propagating through the turbulent non-uniform ISM. The signal is then deflected through a multitude of different path lengths leading to the same signal arriving at different times with a reduced intensity.

A simple way to model this effect is to assume a thin scattering screen lying between pulsar and the observer (Scheuer, 1968; Williamson, 1972). The intensity varies as a function of the angular width θ_d of the disk and can be expressed as:

$$I(\theta)d\theta \propto \exp(-\theta^2/\theta_d^2) 2\pi\theta d\theta \quad (1.15)$$

Signals travelling at different angles (denoted by θ) with respect to the screen can be translated to a time delay. Hence, the intensity can be expressed as a function of time and a scattering timescale τ_s given by (Lorimer & Kramer, 2012):

$$I(t) \propto \exp(-t/\tau_s) \quad (1.16)$$

The dependence of the scattering timescale on the observing frequency (ν) is expressed as a power law:

$$\tau_s \propto \nu^{-\alpha} \quad (1.17)$$

The value of α is model dependent. In case of the thin screen model presented above, $\alpha=4$. If the ISM is assumed to show Kolmogorov turbulence, $\alpha=4.4$ (Romani et al., 1986). Various studies have also empirically derived α by modelling the average pulse

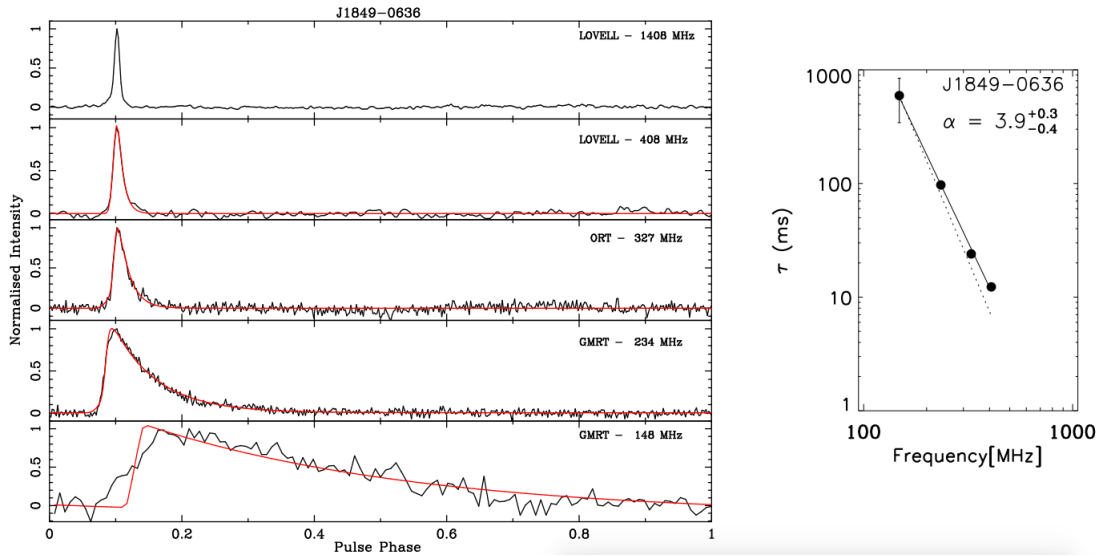


Figure 1.4: A sample plot showing the evolution of the scattered pulse across observing frequency for PSR J1849-0636. The red line represents the exponential decay model fit to the pulse profile. The plot on the right shows a fit to the scattering time scale as a function of frequency. Figure taken from (Krishnakumar et al., 2017)

profile of a sample of pulsars over multiple observing frequencies (e.g. Krishnakumar et al., 2017; Bansal et al., 2019; Oswald et al., 2021). The α values derived from these studies are typically lower than that expected from Kolmogorov inhomogeneities (i.e. below 4.4). Figure 1.4 represents one such example of a particular pulsar observed at different frequencies and whose frequency dependence of the scattering tail is modelled. The scattering timescale is also dependent on the electron density of the ISM along the line of sight, thus leading to efforts to empirically model the dependence of DM (e.g. Bhat et al., 2004).

Despite most pulsars showing a negative spectral index (i.e. higher luminosity at lower frequencies), higher observing frequencies (order of GHz) are preferred to study intrinsic properties of the pulsar emission in order to avoid scattering effects. The impact of scattering also plays a major role in pulsar survey strategies. Galactic plane surveys in particular observe regions with a high density of matter along the line of sight. This can smear the received pulses and thus reduce the sensitivity to new pulsars. For this reason, most galactic plane surveys in recent times are conducted at L-Band or above (see Chapter 2)

1.3.3 Scintillation

The random fluctuations in electron density of the ISM coupled with scattering leads to temporal and frequency dependent variations in the observed intensity. This phenomenon is termed as interstellar scintillation and is commonly observed in pulsar observations. Depending on the scale of the scattering screen and the scattering magni-

tude, scintillation can be classified into diffractive and refractive interstellar scintillation (DISS and RISS).

DISS as the term suggests, is equivalent to a diffraction pattern produced through the small scale irregularities of the ISM. The limiting condition for phase coherence translates to a corresponding decorrelation bandwidth known as the scintillation bandwidth. It can be shown that (Lorimer & Kramer, 2012):

$$\Delta f \propto f^4 \quad (1.18)$$

where Δf is the scintillation bandwidth and f is the observing bandwidth. RISS is an outcome of large scale electron density irregularities in the ISM and is generally broadband in frequency (e.g. Gupta et al., 1994). They tend to produce low amplitude modulation in the pulsar flux over time scales of months to years. The scintillation bandwidth is also inversely proportional to the pulsar distance (Backer, 1975). Since DM is an indirect measure of distance, low DM pulsars tend to show prominent changes in intensity in comparison to high DM pulsars. This is a primary reason for observing known nearby globular clusters (e.g. NGC 104) multiple times over short integration times in order to search for new pulsars. The rapid changes in intensity could potentially yield an observing epoch where a *scintle* can increase the received flux and potentially discover previously unknown pulsars. On the contrary, high DM pulsars have decorrelation bandwidths much lower than the typical channel bandwidths used in telescopes. Such pulsars are a useful probe for studying refractive scintillation.

1.3.4 Faraday Rotation

Sir Michael Faraday demonstrated that a light beam propagating through a magnetic field would experience a rotation in the plane of polarisation. This phenomenon which is now termed as Faraday Rotation can be explained as circular birefringence of a medium that leads to a difference in refractive index between a right and left circularly polarised wave. The difference in refractive index leads to a difference in propagation speeds and the phase difference induces a rotation on an incoming linearly polarised wavefront. An incoming pulsar signal travelling through the magneto-ionic ISM also undergoes Faraday Rotation owing to the highly linearly polarised nature of the pulsar radiation that is impacted by the Galactic magnetic field. To generalise, an electromagnetic wave of frequency ν travelling through a distance d of the magnetised ISM has a phase lag with respect to an infinite-frequency (non-rotated) wave. This can be expressed as (Lorimer & Kramer, 2012):

$$\Delta\Psi = -kd \quad (1.19)$$

where $k = 2\pi\nu$ is the wavenumber. Further, the circular polarisation dependent refraction index n is expressed as:

$$n(\nu) = \sqrt{1 - \frac{\nu_p^2}{\nu^2} \mp \frac{\nu_p^2 \nu_B}{\nu^3}} \quad (1.20)$$

where ν_B is the cyclotron frequency with respect to the galactic magnetic field. The “−” sign in \mp represents the expression for a left handed circularly polarised wave and “+” stands for the right hand circular polarisation. The expression for ν_B for a given galactic magnetic field parallel to the line of sight B_{\parallel} is given as:

$$\nu_B = \frac{eB_{\parallel}}{2\pi m_e c} \quad (1.21)$$

In order to calculate the differential phase rotation induced, we can rewrite equation 1.19 as

$$\Delta\Psi_{\text{Faraday}} = \int_0^d (k_R - k_L) dl \quad (1.22)$$

where k_R and k_L represent the wavenumbers for right and left circularly polarised waves. Assuming that $\nu \gg \nu_B$ (which is mostly valid for typical pulsar observations), equation 1.22 can be written as

$$\Delta\Psi_{\text{Faraday}} = \frac{e^3}{\pi m_e^2 c^2 f^2} \int_0^d n_e B_{\parallel} dl \quad (1.23)$$

1.4 Motivation for pulsar searching

Pulsars have proven to be excellent probes for a variety of scientific phenomena. Listed below are some of the fundamental and critical open questions that drive the need for new discoveries.

1.4.1 Improving the understanding of the neutron star structure

The true composition of neutron stars (as explained in Section 1.2.2) is still unknown. The large densities (order of 10^{14} g cm^{−3}) of these neutron stars are not reproducible in current terrestrial laboratories. Hence, the best option available is to determine the internal properties of pulsars based on astronomical observations that help calculate the mass, radius and spin rate of the neutron star. While there exists a vast set of equations of state (EoS) proposed, discoveries of fast rotating (spin frequency > 700 Hz) (e.g. [Hessels et al., 2006](#); [Bassa et al., 2017b](#)) and massive pulsars ($M > 2 M_{\odot}$) (e.g. [Antoniadis et al., 2013](#); [Fonseca et al., 2021](#)) have helped in ruling out several existing EoS ([Lattimer & Prakash, 2001](#)) as shown in Figure 1.5. The discovery of sub-millisecond pulsars with a much higher rotation rate compared to the current set of millisecond pulsars can significantly change the current approach in modelling the EoS of neutron stars ([Haskell et al., 2018](#)).

Over the years, pulsar timing has provided precise mass measurements of 48 pulsars⁴ but no equivalent precision over the radius of the neutron star. A primary objective of the Neutron star Interior Composition ExploreR (NICER) mission ([Arzoumanian et al., 2012](#)) operated by NASA is to also provide measurements of the radius of neutron stars. Recently, [Riley et al. \(2019\)](#) reported the first precise radius measurement of

⁴https://www3.mpifr-bonn.mpg.de/staff/pfreire/NS_masses.html

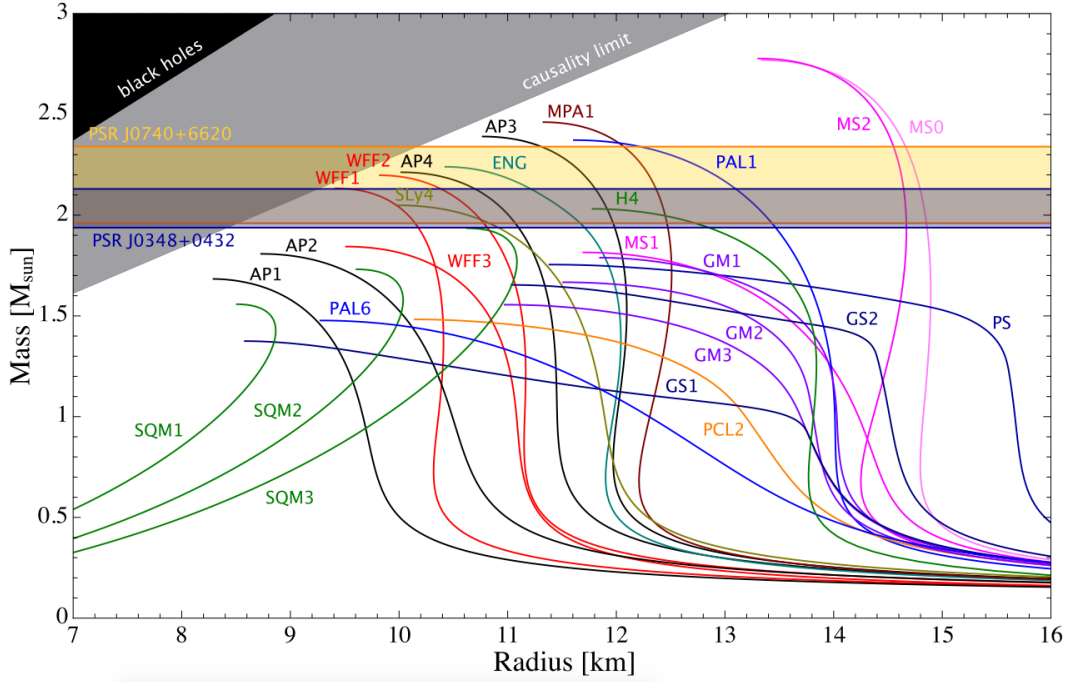


Figure 1.5: Different equations of state highlighting the relation between the mass and radius of a neutron star. The horizontal bands represent the $2\text{-}\sigma$ error bars for PSR J0348+0432 (Antoniadis et al., 2013) and PSR J0740+6620 (Cromartie et al., 2019) which currently holds the record for being the most massive pulsar known. Figure taken from Hu et al. (2020).

PSR J0030+0451 using NICER data as $R = 12.71^{+1.14}_{-1.19}$ km . Similar such radius measurements in the future for other pulsars would help further eliminate many EoSs as shown in Figure 1.5. Mass and radius measurements would also provide an accurate measure of the moment of inertia (MOI) of neutron stars. MOI estimates can also be provided by compact binary pulsars via pulsar timing. For example, Hu et al. (2020) showed through simulations that the MOI for the Double Pulsar can be measured to 11 % accuracy at best in a timescale of ten years with the MeerKAT and SKA telescopes. This motivates the discoveries of compact pulsar binary systems with tighter orbits than the Double Pulsar.

1.4.2 Testing general relativity and alternative theories of gravity

Ever since Albert Einstein first proposed the theory of general relativity (GR) (Einstein, 1915), many tests have been conducted to determine its validity across a wide spectrum of gravitational field strengths. The Pound-Rebka experiment (Pound & Rebka, 1960) confirmed the predictions of GR by measuring the relative gravitational redshift between two sources situated at different altitudes. The observed value matched the expected value to within a 10 % limit. Other precision experiments include laser inte-

ferometry (e.g. [Turyshev et al., 2004](#)) and lunar ranging (e.g. [Müller et al., 2019](#)) that are capable of measuring a wide variety of fundamental effects produced from GR.

The discovery of the Hulse-Taylor binary pulsar PSR B1913+16 ([Hulse & Taylor, 1975b](#)) established a new field of using pulsars for fundamental tests of gravity. The regular monitoring of this binary pulsar provided the first indirect evidence for gravitational waves and this work was awarded the Nobel prize in 1993. A seminal breakthrough came with the discovery of the Double Pulsar PSR J0737-3039A/B ([Burgay et al., 2003](#); [Lyne et al., 2004](#)) which allowed for a multitude of applications including the test for general relativity (GR) in the strong field limit ([Kramer et al., 2006b](#)). The Double Pulsar currently provides multiple stringent tests of GR in the strong field with 99.95 % compatibility between theory and experiment. The recent discovery of the most accelerated binary pulsar currently known ([Cameron et al., 2018](#)) has provided further motivation to discover more such relativistic systems. This system offers the possibility to detect different effects of GR including Lense-Thirring precession ([Lense & Thirring, 1918](#); [Venkatraman Krishnan et al., 2020a](#)) and relativistic deformation of the orbit (see e.g. [Damour & Deruelle, 1985](#)) with a much higher significance in comparison to the Hulse-Taylor pulsar and the Double Pulsar. Similarly, the discovery of a triple system PSR J0337+1715, consisting of a millisecond pulsar and two white dwarfs ([Ransom et al., 2014](#)) has allowed for the most stringent test of the strong equivalence principle ([Archibald et al., 2018](#); [Voisin et al., 2020](#)) with the violation parameter constrained to the order of 10^{-6} at a 95 % confidence level.

In addition to GR, relativistic binary systems also help constrain alternative theories of gravity. This ranges from scalar-tensor gravity theories like Brans-Dicke (e.g. [Will & Zaglauer, 1989](#); [Freire et al., 2012](#)) to constraints on the mass of the graviton ([Shao et al., 2020](#)). The potential discovery of a pulsar-black hole binary offers multiple possibilities including stringent tests of general relativity and a direct measurement of the black hole spin, thus testing the Cosmic Censorship Conjecture and ‘no hair’ theorem ([Wex & Kopeikin, 1999](#); [Liu et al., 2014](#)).

1.4.3 Detecting gravitational waves using pulsar timing arrays

Following the discovery of MSPs and their remarkably stable nature in emitting timed pulses, [Foster & Backer \(1990\)](#) conceived the idea of using a collection of such pulsars as a timing array to detect gravitational waves (GW). GW passing by Earth would disrupt the time of arrival of the pulses from different parts of the sky and the angular separation between the set of pulsars would produce a correlated signature. The recent discovery of gravitational waves by the LIGO consortium ([Abbott et al., 2016](#)) including a kilonova event due to a double neutron star merger inspiral ([Abbott et al., 2017](#)) and a possible neutron star-black hole binary merger ([Abbott et al., 2020](#)) has opened a promising new window to multi-messenger astronomy. Pulsar timing arrays (PTAs) complement the LIGO interferometer and offer exclusive sensitivity to the nanohertz spectrum of GW (see [Figure 1.6](#)). This spectrum offers insight into the class of merging supermassive black holes and the primordial GW background which is a relic of the early universe (e.g. [Chen et al., 2019, 2020](#)).

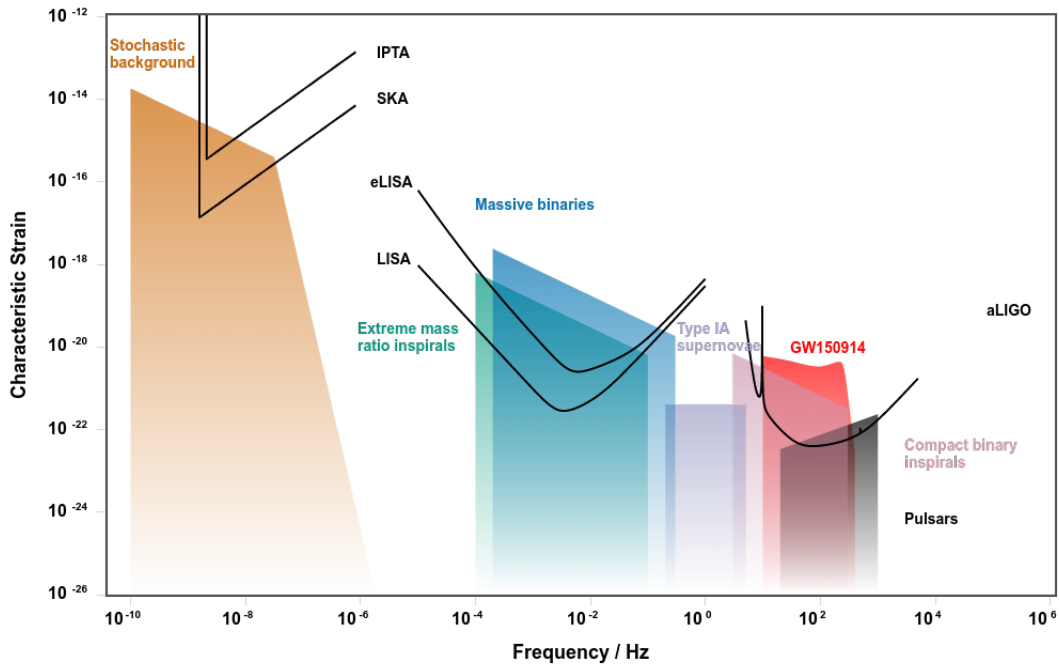


Figure 1.6: A representation of detection limits for different kinds of gravitational wave sources across the spectrum. This plot was generated using <http://gwplotter.com/>.

Over the last decade, a multitude of collaborations have been working on precision pulsar timing experiments using a collection of MSPs. The first such collaboration conceived was the Parkes Pulsar Timing array (Manchester, 2008) which aimed at using pulsar timing data from about 20 MSPs to place limits on the gravitational wave background. Since then, other pulsar timing arrays have emerged namely, the European Pulsar Timing Array (EPTA)(Stappers et al., 2006), North American Nanohertz Observatory for Gravitational Waves(NANOGrav)⁵ and recently the Indian Pulsar Timing Array (Paul et al., 2019) to name a few. Each of these collaborations come under the umbrella of the International Pulsar Timing Array (IPTA)(Hobbs et al., 2010). The IPTA regularly releases the data to the public in the quest to detect low-frequency gravitational waves (e.g. Perera et al., 2019). The data sets produced from the PTA collaborations have increased in sensitivity over time and in turn reduced the upper limits of a potential stochastic gravitational wave background signal detection. Recently, NANOGrav reported evidence of a stochastic process with a correlated signature across different pulsars (Arzoumanian et al., 2020). Although the astrophysical origin of this signal is debatable, it motivates the discovery of new MSPs which can be incorporated into the PTAs and thus increase the sensitivity to detect nanohertz gravitational wave sources in the future.

Apart from detecting gravitational waves, the property of MSPs being excellent time keepers has found applications in other fields. They could serve as a reference

⁵<http://www.nanograv.org/>

system for deep space navigation (Becker et al., 2018). The precision achievable with timing MSPs has also proven to be an excellent tool to improve measurements of masses of planets and asteroids within the solar system (Caballero et al., 2018).

1.4.4 Probing the Galactic centre region

The small number of pulsars discovered in the Galactic centre region so far (6 known pulsars in a 70 pc radius around Sagittarius A*; Johnston et al., 2006; Deneva et al., 2009; Eatough et al., 2013b) raises questions on the current estimates of neutron star birth rates around this region (Wharton et al., 2012). It could also indicate that the propagation effects are a major hindrance to the detection of such systems. Despite a poor yield so far, the numerous scientific possibilities coupled with improved sensitivity from latest telescopes in the Southern hemisphere like MeerKAT motivate the continuation of searches around the Galactic centre. Discovery of a pulsar in a tight orbit with Sagittarius A* would be an ideal probe for understanding the gravitational influence of supermassive black holes as well as the environment surrounding Sagittarius A* (see Bower et al. (2019) and references therein). Although the current set of pulsars in the Galactic centre (except the Galactic centre magnetar: PSR J1745-2900) are relatively distant from Sagittarius A* (> 0.1 deg.), more such discoveries can find a wide range of use cases. For example, previous discoveries have allowed for a better understanding of the magneto-ionic environment around the Galactic centre region (Desvignes et al., 2018). Additionally, studying their spin down rates has helped constrain the gravitational potential at the centre of the Galaxy (Perera et al., 2019). A sub-survey focused on the Galactic centre with MeerKAT is described in Chapter 4.

1.4.5 Improving Galactic pulsar population models

Over the years, new pulsar discoveries have aided our understanding of the underlying distribution of the ensemble population. The main caveat with pulsar surveys is the limited sensitivity in terms of restricted bandwidth, poor time and frequency resolution and a bounded sky coverage. This inherently leads to selection biases and thus provides an incomplete picture of the true scenario. Given that every new population study has improved upon previous limitations due to sample sizes (Lyne et al., 1985; Lorimer et al., 2006), the continual search for pulsars is necessary to resolve existing conflicts and open questions. These questions revolve around finding out the true number of neutron stars/pulsars and their luminosity and spatial distributions in the Galaxy. The population models derived from these numbers can enable a vast number of use cases (Lorimer et al., 2019). For example, they would improve the constraints on the characterisation of various pulsar properties as discussed in Section 1.2 (e.g. Johnston & Kerr, 2017; Johnston & Karastergiou, 2019) including constraints on the supernova rate in the Galaxy. A corollary to improving the population synthesis studies is the modelling of the millisecond pulsar population (e.g. Lorimer et al., 2015). However, it is worth noting that the current studies are biased by MSPs which are relatively close to Earth in comparison to the scale of the Galaxy. This drives the need for discovering MSPs

beyond the current known population of 400⁶ at the time of writing. Besides this, improved population models accompanied by advanced technological capabilities continue to refine the strategies to search for new pulsars within the Galaxy (see Chapter 2).

1.4.6 Understanding different millisecond pulsar evolution mechanisms

The recycling process that aids in MSP formation (as discussed in 1.2.8.1) suggests that MSPs would be found in binary systems. However, several isolated MSPs are known including the first MSP ever discovered (PSR B1937+21; [Backer et al., 1982](#)). The formation mechanism of isolated MSPs does not have a convincing explanation yet and has motivated a wide range of theories to explain possible evolution scenarios. A prominent explanation is that the wind from the pulsar ablates the companion with time ([Fruchter et al., 1988](#)). The discoveries of multiple eclipsing pulsar systems (also termed as spiders) with low mass companions ($M \ll 0.1M_{\odot}$; also termed as black widows) (e.g. [Roberts, 2013](#)) have supported this mechanism. Another train of thought suggests that isolated MSPs are a consequence of exchange interactions from hierarchical triple systems (e.g. [Freire et al., 2011](#)). The discovery of the triple system ([Ransom et al., 2014](#)) has added substance to this argument.

While MSP evolution is a well studied problem, new discoveries could open new avenues. For example, transitional MSPs have provided deep insight into accretion physics through multi-wavelength studies ([Stappers et al., 2014](#)) as well as magnetohydrodynamical simulations ([Parfrey & Tchekhovskoy, 2017](#)). Adding to existing populations of double neutron star (DNS) systems and neutron star-white dwarf binaries can help probe the various channels of formation of compact binaries as well as better understand the supernova mechanism (e.g. [Tauris et al., 2017](#)). A recent discovery of an ultra-stripped supernova with evidence of a Helium envelope ([De et al., 2018](#)) has provided substantial credence to the standard channel of double neutron star formation (see Figure 1.7). A detailed description of this scenario can be found in [Tauris et al. \(2017\)](#). An example of a contrasting evolution history is the binary pulsar PSR J0453+1559 which is known to have a large mass asymmetry. [Martinez et al. \(2015\)](#) initially suggested that the pulsar companion is a low mass neutron star, but [Tauris & Janka \(2019\)](#) applied the electron capture supernova mechanism model ([Jones et al., 2016](#)) which yielded the companion to be a high mass white dwarf. Discovering more such systems could potentially resolve such conflicts.

Besides this, there exist discoveries whose properties make them outliers in the current understanding of formation and evolution scenarios (as discussed in Section 1.2.8.1). A prime example is the existence of a variety of millisecond pulsars in eccentric binaries ([Champion et al., 2008](#); [Barr et al., 2013](#); [Deneva et al., 2013](#); [Camilo et al., 2015](#); [Knispel et al., 2015](#); [Antoniadis et al., 2016](#); [Barr et al., 2017](#)). This is unusual given that the recycling process and tidal circularisation is expected to yield circular binaries. While formation of such systems in globular clusters is not surprising (owing to much larger densities and high encounter rates), these pulsars are anomalous

⁶<http://astro.phys.wvu.edu/GalacticMSPs/GalacticMSPs.txt>

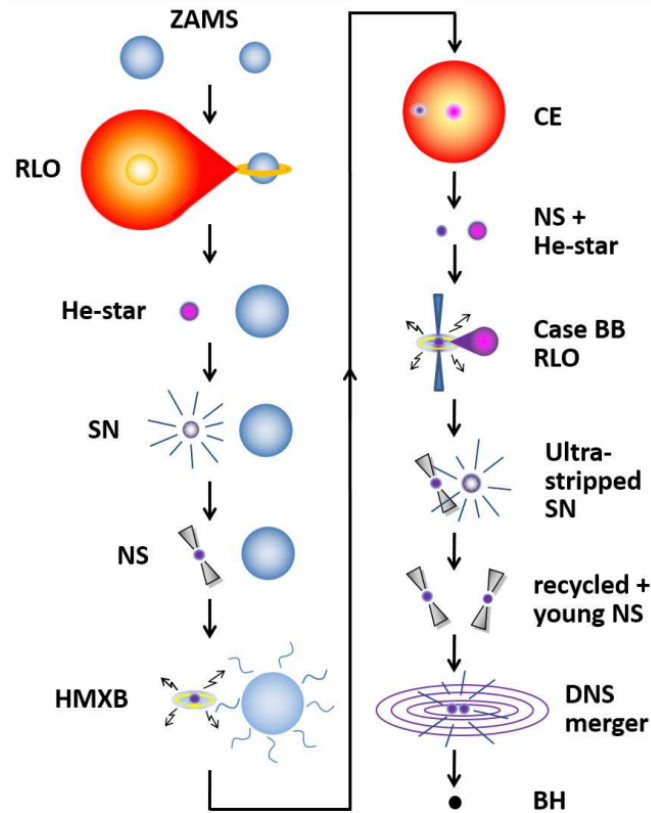


Figure 1.7: Sequence of events leading up to the formation of a double neutron star system and an eventual inspiral. The process begins from two zero age main sequence stars (ZAMS) in a binary system. One of the stars begins to fill its Roche Lobe and subsequently transfers material to the companion star. This is the Roche Lobe overflow (RLO). The former star eventually undergoes a supernova (SN) explosion. If the binary survives this explosion, the system enters the high mass X-ray binary (HMXB) phase. The companion star sheds its outer layers forming a common envelope (CE) and eventually forming a Helium star. A second RLO phase begins where the primary star accretes matter from the newly formed He-star leading to a second supernova explosion (so called ultra stripped supernova as reported to be observed by [De et al. \(2018\)](#)). If the system survives a second explosion, a double neutron star system is formed where one is recycled and the other is a young neutron star (similar to the Double Pulsar). The system loses energy via gravitational wave emission (as demonstrated with PSR B1913+16) eventually leading to a double neutron star (DNS) merger (reported as the first kilonova event by [Abbott et al. \(2017\)](#)). A more detailed explanation can be found in [Tauris et al. \(2017\)](#). Figure taken from [Tauris et al. \(2017\)](#).

given that they are found in the Galactic field. This has led to multiple theories explaining eccentric millisecond pulsar formation (Freire et al., 2011; Freire & Tauris, 2013; Antoniadis, 2014; Jiang et al., 2015) with no theory standing out from the rest currently.

1.4.7 Discovering sources beyond the current pulsar census

As discussed in Section 1.2.8, pulsars can be clustered into different sub-categories that range from MSPs and magnetars to sporadic emitters like RRATs and intermittent pulsars. Boosting the number of discoveries in each of these classes can help improve the current models used to explain the formation and evolution of such systems. Apart from increasing the population of known sources, searching for pulsar-like signals can potentially reveal new types of astrophysical sources. This is apparent from the discovery of fast radio bursts (FRBs) (Lorimer et al., 2007) which were a consequence of aiming to find pulsars via single pulses. Fast radio bursts are short duration bursts ranging a few milliseconds. Although the progenitors of FRBs are still a matter of debate, some recent results indicate a strong likelihood of FRBs to originate from magnetar outbursts (CHIME/FRB Collaboration et al., 2020; Bochenek et al., 2020). Other enigmatic potential discoveries include pulsar-black hole binaries, a binary system with two millisecond pulsars, sub millisecond pulsars and discovery of higher order hierarchical systems with four objects or more.

1.5 Thesis outline

This thesis deals with new strategies and techniques that have consequences on pulsar searching as well as timing. The focus on the pulsar searching side is to encapsulate modern technological capabilities for large scale pulsar surveys with interferometers via pulsar surveys conducted with the MeerKAT telescope. The focus on the pulsar timing aspect is the improved characterisation of pulse profile stability by studying a specific pulsar belonging to pulsar timing arrays.

- In Chapter 2, I present the pulsar search methodology and an overview of previous and ongoing pulsar surveys.
- In Chapter 3, I describe a high speed autonomous pulsar search system design for the MeerKAT telescope. I also present a specific case of using the MeerKAT system to search for millisecond pulsars in globular cluster Terzan 5 as part of the Transients and Pulsars with MeerKAT (TRAPUM) survey
- In Chapter 4, I present the MPIfR Galactic Plane Survey, detailing the system configuration, processing setup and initial results from a pilot survey.
- In Chapter 5, I present a scalable and flexible system for pulsar candidate classification in collaboration with the multinational technological giant SAP.

- In Chapter 6, I give a brief introduction to concepts in pulsar timing methodology. I then examine the profile instability of the millisecond pulsar PSR J1022+1001 in detail and present robust techniques to determine if the integrated profile of a pulsar is stable over long time scales.
- In Chapter 7, I summarise the findings of this thesis and provide a road-map of future research plans that are built upon the results obtained.

Finding pulsars

In Chapter 1, I summarised the properties of pulsars and the need for finding new pulsars. The process of radio pulsar searching involves specialised hardware and a pot-pourri of techniques implemented on software with the motive of detecting a previously unknown pulsar signal. While certain parts to this procedure have stood the test of time and continue to be utilised in one form or another, an array of new issues have arisen particularly from man-made interference and large data volumes of data flowing in, primarily owing to the large number of pixels on sky. This chapter gives an in-depth analysis into the various steps that go into a pulsar searching setup, a multitude of search techniques catering to different needs and an overview of accomplishments from previous and ongoing pulsar surveys.

Contents

2.1	Radio telescope instrumentation for pulsar data acquisition . .	46
2.2	Pulsar search procedure	49
2.2.1	Data storage	49
2.2.2	Interference mitigation techniques	49
2.2.3	Dedispersion	52
2.2.4	Barycentering	58
2.2.5	Discrete Fourier Transform	58
2.2.6	Red noise removal	60
2.2.7	Harmonic Summing	60
2.2.8	Folding	63
2.2.9	Candidate Sifting	65
2.2.10	False alarm probability and minimum S/N	66
2.2.11	Acceleration searching	68
2.2.12	Other search techniques	75
2.3	Pulsar Software	78
2.4	Overview of Pulsar Surveys	80
2.4.1	Large area surveys	80
2.4.2	Targeted surveys	83
2.5	Pulsar surveys with interferometers	84

2.1 Radio telescope instrumentation for pulsar data acquisition

A radio telescope can be comprehended as a system that helps track and monitor astronomical sources via reception of radiation in the radio frequency end of the electromagnetic spectrum. The power received from such far-field sources is weak and hence radio astronomers resort to using Jansky (Jy) for measuring the flux of an astronomical source. To put this in context, 1 Jy in standard SI units is given as:

$$1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1} \quad (2.1)$$

The faintest pulsars currently known are few μJy (see e.g. [Manchester et al., 2005](#)), thus highlighting the weakness of the signals that are received. In order to maximise the chances for new pulsar discoveries, telescopes need to ensure maximum sensitivity through:

1. A robust receiver design ensuring low noise amplification and minimised losses to radio frequency interference (RFI) and spillage.
2. High time resolution that can help discover the fastest spinning pulsars as well as enable analysis of structure within the pulsed emission.
3. Large bandwidth front-end with a high frequency resolution back-end to ensure maximum received power from the weak and highly dispersed pulsar signals.
4. Storage of high fidelity data through conversion into optimised data formats that can be archived and reprocessed using a multitude of techniques over a long period of time.

A standard telescope chain used for pulsar searches can be broadly categorised into (see Figure 2.1):

2.1.0.1 Front-end

1. *Radio Antenna*

This apparatus is responsible for converting the incident electromagnetic radiation into an analog Radio Frequency (RF) signal. The radio antenna apparatus is designed in different ways depending on the observing frequency. Telescopes observing at radio frequencies up to 500 MHz typically consist of an array of dipoles (e.g. LOFAR) whereas higher observing frequency (1 GHz and above) antennas are typically designed with a parabolic reflector which helps focus the incoming radio radiation onto a primary focal point (e.g. Effelsberg radio telescope). The incident wave front is then converted into a voltage signal at the feed. The feed is also equipped to record two orthogonal polarisation modes that help measure the linear or circular polarisation respectively of the incoming signal.

2. *Low-Noise Amplifiers*

Given that astronomical signals are weak, the RF signal requires an amplification process ensuring a minimal injection of noise. Low-Noise amplifiers (LNAs) are specifically designed for this purpose and are located in close proximity to the receiver. This reduces the effect of possible attenuation or noise corruption of the RF signal along the telescope chain. LNAs are cryogenically cooled thus reducing the thermal noise in the signal chain and ensuring minimal loss in sensitivity.

3. *Bandpass filtering*

Although the RF receiver is sensitive across a finite bandwidth, the attenuation at the edges of the band may not be enough to remove man-made interference signals at frequencies outside this range. Thus, a bandpass filter around the central frequency of interest is applied to mitigate the impact of out-of-bound sources that could tamper the RF signal.

4. *Mixer*

The amplified RF signal undergoes a down-conversion to a lower intermediate frequency (IF) after passing through a mixer. First, a local oscillator (LO) produces a sinusoidal signal (ν_{LO}) that is multiplied with the incoming RF signal (ν_{RF}). This produces two frequencies which are the sum and difference of the two input signals ($\nu_{RF} \pm \nu_{LO}$). The beat frequency ($\nu_{RF} - \nu_{LO}$) is retained and the other frequency is rejected by a second bandpass filter. This step can help reduce transmission losses that are prominent at higher frequencies. Another practical advantage to down-conversion is that a low Nyquist sampling rate would be required to prevent aliasing during analog-to-digital conversion (see the next Section).

2.1.0.2 Back-end

1. *Analog-to-digital conversion*

The down-converted signal is transformed to a digital format using an Analog-to-Digital converter (ADC). The sampling rate is given as:

$$t_{samp} = \frac{1}{2B} \quad (2.2)$$

where B is the observing bandwidth set by the Nyquist-Shannon theorem (Nyquist, 1928; Shannon, 1949) in order to avoid aliasing effects. The number of bits per sample is also determined in this step. Greater the number of bits, greater is the size and dynamic range of the converted data. Powerful RFI signals would saturate the ADC thus showing up on the extreme end of the dynamic range. This could be a useful feature for RFI mitigation techniques (see more details in 2.2.2).

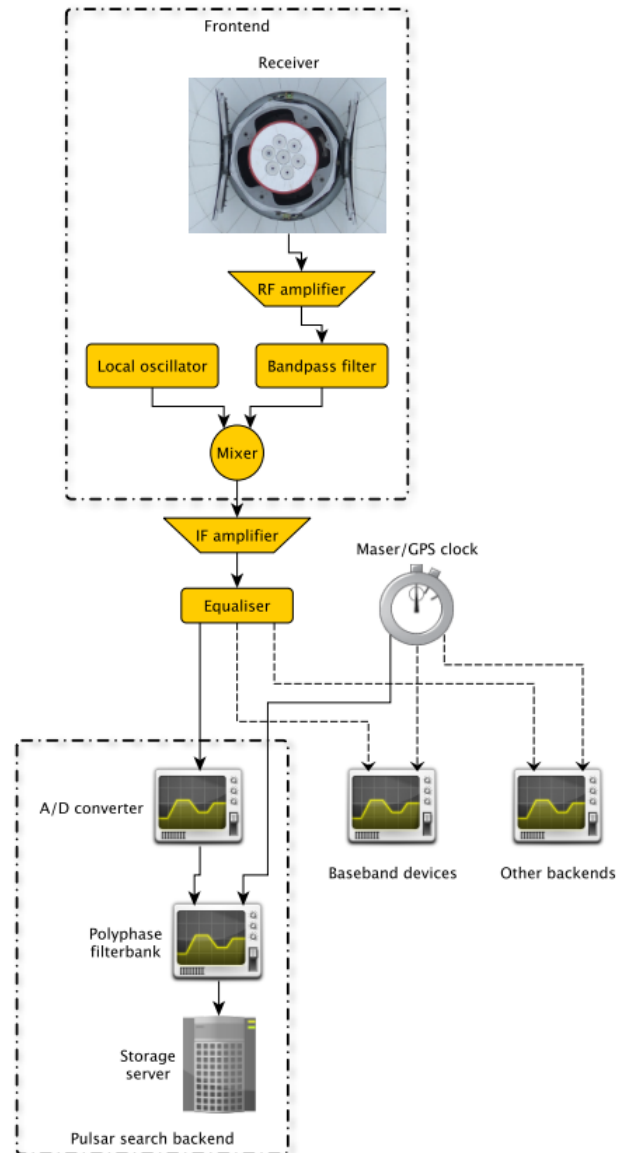


Figure 2.1: A schematic of a generic pulsar observing and recording setup. Figure taken from (Barr, 2013).

2. *Polyphase filterbanks*

A pulsar spectrometer better known as a polyphase filterbank (PFB) splits up the power across the observing frequency bandwidth into a fixed number of spectral channels. This is done by applying the Fast Fourier Transform (Cooley & Tukey, 1965) via software or through a more customised solution using Field Programmable Gate Arrays (FPGAs). The polyphase in PFB refers to polyphase filtering techniques implemented to minimise spectral leakage between channels¹. The time and frequency resolved chunks of data are stored as files on disk which is used for further processing.

2.2 Pulsar search procedure

Once the filterbank data from an observation is recorded on disk, several methods are applied to increase the probability of finding a previously unknown pulsar signal. Described below are each of these steps in detail.

2.2.1 Data storage

The output files from a polyphase filterbank contain information in the time and observing frequency phase space. However, this information has to be augmented with observation parameters like sampling time, frequency bandwidth, sky coordinates of the observation pointing to name a few. A standard data format with an appropriate header to each data file recorded resolves this issue. The common data format used for storing pulsar search data is the SIGPROC filterbank format². The filterbank format contains a binary header containing a dictionary of keyword-value pairs that store the most relevant metadata pertaining to the recorded file. This is followed by n-bit values which are ordered such that the first N_{chan} (i.e. number of frequency channels) samples represent the first time sample per frequency channel and so on. While this format has a few shortcomings compared to the much more elaborate Flexible Image Transport System (FITS) format³, it provides the appropriate amount of information relevant for implementing searching techniques. Data analysis pertaining to pulsar searches in this thesis makes use of the filterbank formatted data unless specified otherwise.

2.2.2 Interference mitigation techniques

The advancement in communication systems in recent times have led to extensive usage of RF bands as a mode for transmission. This includes television tower signals, GSM bands for mobile communication, satellite signals, military radar and numerous other communication protocols which operate at frequencies that overlap with the

¹More information can be found in https://casper.ssl.berkeley.edu/wiki/The_Polyphase_Filter_Bank_Technique

²More details on this can be found in <http://sigproc.sourceforge.net/sigproc.pdf>

³Detailed description of the format can be found in https://www.atnf.csiro.au/research/pulsar/psrfits_definition/Psrfits.html

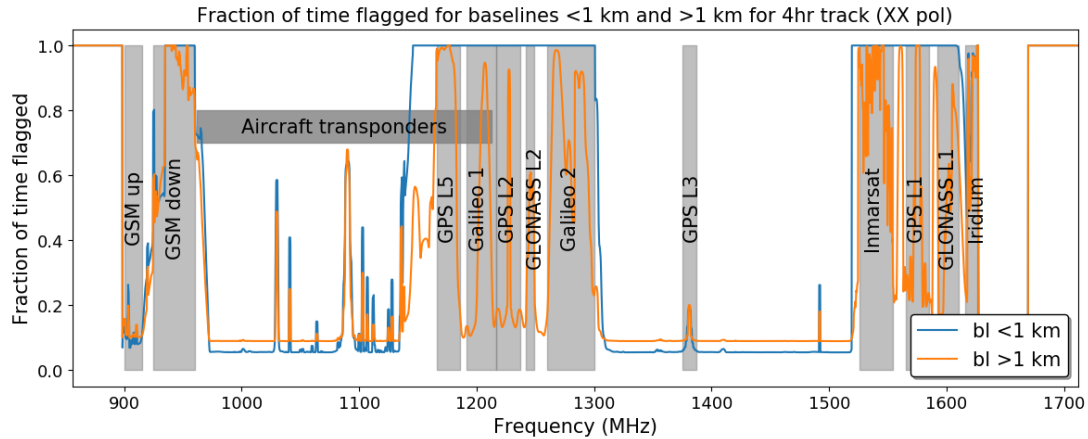


Figure 2.2: The RFI situation at L-Band at MeerKAT monitored over a 30 hour observation. The blue line shows flagging of channels on core baselines (<1 km) that includes a static mask and an outlier detection algorithm. The orange line shows flagging for the longer baselines. Only a single polarisation (XX) is shown. The corresponding sources of RFI are also highlighted corresponding to their frequency of transmission. Credits: SARA0

bands of scientific interest for the radio astronomy community (see Figure 2.2). Pulsar observations generally operate in the MHz to GHz range and bear a major brunt of the man-made interference. It is thus an essential task for current day observing setups to adapt and cope with RFI signals in the observing bandwidth of interest via robust excision techniques. Summarised below are some set of techniques applicable to pulsar observations:

- *Frequency Domain mitigation*

One of the diagnostics used during pulsar observations is the average power across time samples associated with each frequency channel. This is referred to as the *bandpass* and any channel beyond a power threshold is flagged and not included in further processing. This threshold can be decided based on different statistical metrics. A common metric used is a running median with a fixed channel window across the bandpass which is unbiased by a few outlier samples. The corrupt channels once identified could be replaced either with zeros, white noise or any other user defined value. In the case where sources of RFI operating at certain frequencies are well known, the corresponding channels that overlap with the interference signals can be ignored.

- *Time domain mitigation*

Given that man-made signals are terrestrial, they do not undergo dispersion (see Section 1.3.1) unlike a received pulsar signal. This feature can be exploited to identify high valued samples from a timeseries generated without correcting for dispersion (also known as a zero-DM timeseries). The high-valued samples

can also be replaced by zeros, or injected with samples drawn from a Gaussian distribution having a mean and standard deviation similar to neighbouring time samples. However, this technique could also be a disadvantage when detecting signals from fast radio bursts or giant pulses from pulsars.

There also exist hybrid techniques where information in the frequency and time space are used for RFI mitigation. A prime example is the zero-DM filter (Eatough et al., 2009) where the mean of all frequency channels corresponding to a time sample is subtracted from every frequency channel in the time sample. This is particularly useful in removing broadband interference signals.

- Multibeam excision

Multibeam receivers (e.g. 13-beam Parkes multibeam or the 7-beam Effelsberg receiver) not only offer increased sky coverage without compromising the sensitivity, but can also help better distinguish RFI from a pulsar signal. While a pulsar signal is detected in one or few neighbouring beams, RFI signals being terrestrial in origin and much stronger than astronomical signals, would therefore show up in almost all beams even though they are far apart.

There are simple as well as complex ways for applying multibeam excision techniques to clean the data. One way is to identify a time sample that shows up in a fixed number of beams above a certain threshold (typically 3-sigma or 4-sigma above noise). The probability for such a detection is given by p^n where p is the probability for a threshold based detection in one beam and n is the number of beams selected. The same analysis can also be applied in the Fourier domain where Fourier bins showing power above a threshold in multiple beams can be replaced with Gaussian noise. These techniques have been previously implemented for many pulsar surveys conducted with a multibeam receiver (see e.g. Barr et al., 2013). Another approach to the problem is by analysis based on spatial filtering. Kocz et al. (2010) and Kocz et al. (2012) demonstrated the usage of a covariance matrix to represent data of different beams and using eigen value decomposition to determine a subspace where RFI and pulsars can be demarcated from one another. Although this technique has proven to be useful for finding new pulsars (Kocz et al., 2012), the computational expense escalates as the number of beams increase. Given that current and future telescopes are aiming for multiple beam capabilities, a simple and scalable technique for multibeam excision is essential. In Chapter 3 we discuss an application of multibeam excision using spatial filtering used for pulsar candidate sifting for the default pulsar search pipeline with MeerKAT.

Another way to ascertain the true nature of the incoming signal in a multibeam setup, is to carefully characterise the beam pattern making sure that any side lobe based power is also accounted for. The amount of illumination of each beam due to RFI is also dependent on the distance between the telescope and the source of interference. This distance can in turn be expressed as a function

of observing wavelength which would determine if the RFI source is far-field or near-field (equivalent to the Fraunhofer and Fresnel zones). This implies that every beam is not always uniformly illuminated by an RFI source. Although modelling the expected illumination pattern at different distances is an accurate characterisation, the computational expense needs to be kept in mind before implementing on a large scale.

2.2.3 Dedispersion

A broadband pulsar signal when propagating through the ISM is dispersed due to the frequency dependent refractive index of the medium (as discussed in Section 1.3.1). This is quantified by the integrated column density of electrons along the line of sight or the dispersion measure (DM). In context with pulsar searching, removal of the dispersion effect is essential for improving the signal-to-noise ratio and detect a pulsar (see Figure 2.3). The dispersion measure smearing τ_{DM} across a finite bandwidth (BW in MHz) at a centre frequency of ν in GHz is proportional to the DM and can be quantified as (e.g. Cordes, 2002):

$$\tau_{DM} = 8.3 \times DM \times \nu^{-3} \times BW \text{ (in seconds)} \quad (2.3)$$

2.2.3.1 Types

- *Incoherent dedispersion*

Equation 2.3 highlights the need for the removal of dispersion. Owing to the polyphase filterbank, spectral information can be stored on a channel-by-channel basis. This feature can be utilised to calculate the time delay per frequency channel and apply a corresponding time shift. This can be expressed as an adapted version of equation 1.13 (Barr, 2013):

$$\Delta bin_i = \left\lceil 4.15 \times 10^3 \left(\frac{\text{sec}}{t_{\text{samp}}} \right) \left(\frac{\text{DM}}{\text{pccm}^{-3}} \right) \left(\frac{\text{MHz}}{f_i^2} - \frac{\text{MHz}}{f_{\text{top}}^2} \right) \right\rceil \quad (2.4)$$

where Δbin_i represents the integer number of time samples which needs to be shifted at frequency f_i with reference to the frequency at the top of the band (f_{top}) and $\lceil x \rceil$ is the nearest neighbour integer to x .

Once the delays per frequency channel are used to shift the data along the time-axis, the frequency axis of the filterbank data can be collapsed to output a *dedispersed timeseries*. The process of implementing equation 2.4 is also known as *incoherent dedispersion*. This technique is incoherent since the dispersion within a frequency channel is not removed. Despite this flaw, the simplicity of the algorithm enables enhanced computation speed and quick processing over large datasets. This is crucial when searching over a large range of DM trial values for an unknown pulsar signal.

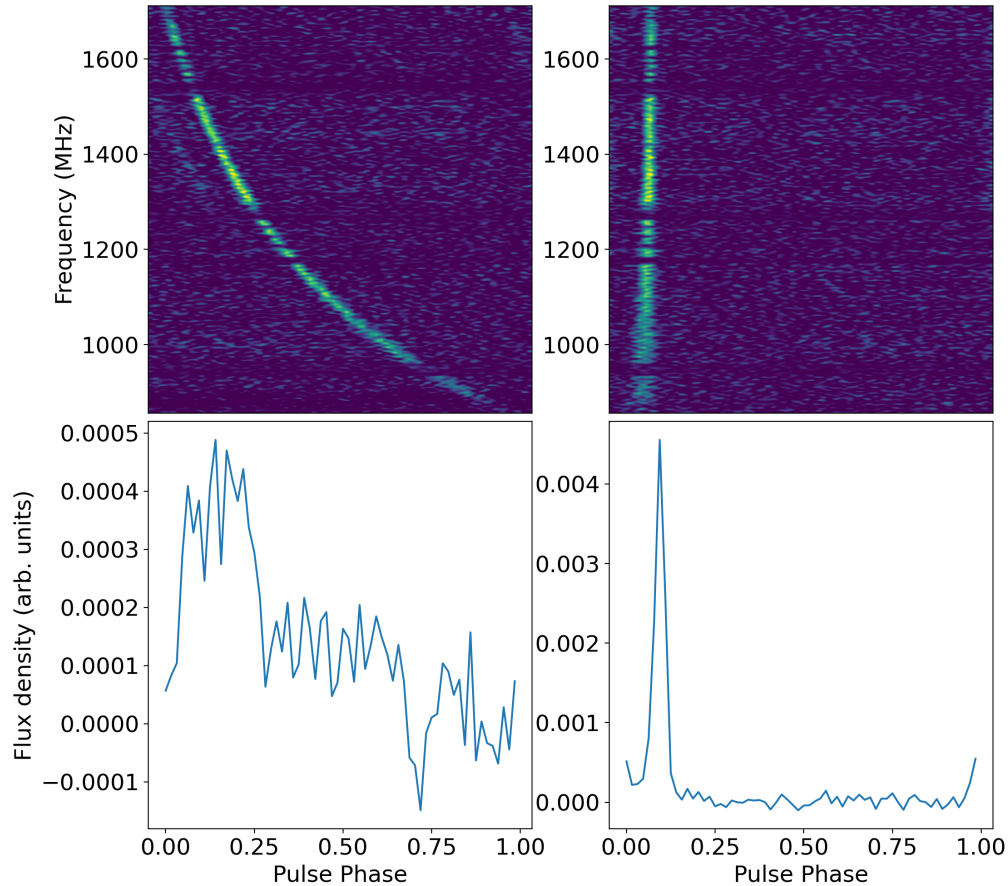


Figure 2.3: Example of the dispersion effect demonstrated using an observation of PSR J1748-2446I at 1.4 GHz with the MeerKAT radio telescope. *Left:* The pulsed signal has a quadratic dispersive delay feature thus leading to the signal arriving earlier in time for the higher frequencies. The integrated profile formed by summing across frequency smears out the power (bottom panel). *Right:* the same observation after dedispersion shows the pulsed signal to be aligned across frequency and hence restore the true pulse profile (bottom panel). Some frequency bands have been flagged due to RFI.

- *Coherent dedispersion*

In order to completely compensate for dispersive smearing, an alternative to incoherent dedispersion is required. This can be achieved by applying a deconvolution to the complex voltage signal in the telescope chain prior to the detection of power. The deconvolution is generally applied on data that is transformed into the Fourier domain. The phases of the complex Fourier signal are rotated based on the lag caused by the DM. The transfer function model can be expressed as (e.g. [Hankins & Rickett, 1975](#)):

$$H(f_0 + f) = e^{+i \frac{2\pi\mathcal{D}}{(f+f_0)f_0^2} DMf^2} \quad (2.5)$$

where f_0 represents the central observing frequency, f represents the Fourier frequency of interest and \mathcal{D} is the dispersion constant.

The drawback of this technique is the computational complexity required for retrieving a signal with an unknown DM value. Hence, it is generally not implemented for blind surveys but can be an advantage for targeted pulsar surveys on sources like globular clusters with an already known DM (e.g. [Ridolfi et al., 2021](#)). The robustness of coherent dedispersion is highly beneficial for pulsar timing purposes where the DM value is known apriori and high S/N pulses are key to high precision timing. Several telescopes have thus incorporated this feature for pulsar timing backends (see e.g. [Hotan et al., 2005](#); [De & Gupta, 2016](#); [Lazarus et al., 2016](#)).

2.2.3.2 DM search optimisation

A combination of appropriate algorithms and strategies can ease the burden of determining the DM of an unknown pulsar. Described below are some key aspects associated with the dedispersion process (we assume incoherent dedispersion henceforth) which serves as a balance between maximising probability of a detection and ensuring high computation speed.

- *Constraining the range*

A limited number of DM trials are used while searching. This is owing to processing time limitations and drop in sensitivity for high DM values. While electron density models (e.g. [Cordes & Lazio, 2002](#); [Yao et al., 2017](#)) can give an estimate of the expected DM along the line of sight (for a given distance), these values do not always comply with the parameters associated with known pulsars. Blind searches are typically bounded by the maximum DM at which the time and frequency resolution are not able to correct for smearing beyond a certain threshold. The DM cap chosen is also observing frequency and bandwidth dependent (as seen in equation 2.3). On the other hand, specific targets like globular clusters which contain an already known set of pulsars, have a well constrained range of DM values to search from (typically using the central DM value among known pulsars as a reference).

- *Determining the step size between DM trials*

Besides the DM range, the gap between two consecutive DM trials is also a crucial factor to calculate. An extremely small step size can drastically increase computation time. On the other hand, an extremely large step size could significantly reduce the sensitivity to a pulsar whose DM is midway between two neighbouring DM trials. The optimised DM step is decided by factoring various effects that could reduce the S/N of the actual pulse received. The observed width of the pulse (W_{tot}) is influenced by multiple independent factors that smear out the true width of the pulse (W_{int}) and can be expressed as (e.g. [Levin, 2012](#)):

$$W_{\text{tot}} = \sqrt{W_{\text{int}}^2 + \Delta t_{\text{scatter}}^2 + \Delta t_{\text{dispersion}}^2 + \Delta t_{\text{sampling}}^2 + \Delta t_{\Delta\text{DM}}^2} \quad (2.6)$$

where each of the smearing terms are discussed in detail below:

- $\Delta t_{\text{scatter}}$: As discussed in section 1.3.2, some pulsed signals show an exponential tail-like feature owing to mutipath propagation through the interstellar medium (ISM). The relation between the scattering timescale and the DM can be determined empirically. One such expression is given as ([Bhat et al., 2004](#)):

$$\log \Delta t_{\text{scatter}} = -6.46 + 0.154 \log(\text{DM}) + 1.07(\log \text{DM})^2 - 3.86 \log f \quad (2.7)$$

- $\Delta t_{\text{dispersion}}$: This is the intra-channel dispersive timescale.

- $\Delta t_{\text{sampling}}$: This is the intrinsic smearing associated with a finite sampling interval and is given by the sampling time used.

There are multiple ways to choose the step size based on the parameters expressed in equation 2.6. Given that $\Delta t_{\text{sampling}}$ is inevitable in the dedispersion setup, a logical consequence of choosing a DM step size is such that any smearing across the observed bandwidth does not exceed $\Delta t_{\text{sampling}}$. This can also be expressed as ([Barr, 2013](#)):

$$\Delta\text{DM} = 2.41 \times 10^{-4} \left(\frac{t_{\text{samp}}}{\text{sec}} \right) \left(\frac{\text{MHz}}{f_{\text{bottom}}^{-2}} - \frac{\text{MHz}}{f_{\text{top}}^2} \right) \text{pc cm}^{-3} \quad (2.8)$$

For a given set of observing parameters, there exists a DM at which the intra-channel smearing equals one sampling interval. This is known as the *diagonal DM*. DM values exceeding the diagonal DM, would smear the pulse beyond the given sampling time. This fact can be exploited to downsample the data in the time axis, since correcting for DM values above the diagonal DM would not require a fine time resolution.

A different approach for determining the DM step size is to set a cap on the tolerance threshold which is acceptable for the amount of smearing added on the intrinsic pulse width. This can be expressed as ([Morello et al., 2019](#)):

$$W_{\text{tot}} \leq (1 + \epsilon)W_{\text{eff}} \quad (2.9)$$

where W_{eff} is the effective pulse width based on all smearing factors except $\Delta t_{\Delta\text{DM}}$. It can be shown that every DM trial value subject to equation 2.9 can be approximated to the following expression (Levin, 2012):

$$DM = N_{\text{chan}}^2 \alpha DM' + \sqrt{16\alpha (\epsilon^2 - N_{\text{chan}}^2 \alpha) DM^2 + 16\alpha\beta (\epsilon^2 - 1) \left(\frac{\nu^3}{8.3B}\right)^2} \quad (2.10)$$

where N_{chan} is the total number of channels, DM' is the previous trial value, B is the full observing bandwidth, and $\alpha = 1/(16 + N_{\text{chan}}^2)$ and $\beta = \Delta t_{\text{scatter}}^2 + \Delta t_{\text{sampling}}^2 + W_{\text{int}}^2$ (in μs^2)

- *Subbanding*

Dedispersion can also be approached as a two-step process for sources where the DM is well confined (e.g. Magellanic clouds or globular clusters). First, the data are dedispersed within a set of subgroups of all channels. The shifts that are applied within each subgroup are calculated using equation 2.4 but with reference to the top most frequency within the frequency subband. This *subbanded* file can once again be dedispersed by applying a shift per frequency subband for DMs close to the DM applied in the first step (see Figure 2.4). The intermediate subbanded products are useful for long term archiving in storage systems with limited amount of space. The dedispersion computing time is also reduced for these subbanded files. We describe a use-case of subbanding for the TRAPUM survey in Chapter 3.

- *Improved algorithmic approach*

Dedispersion as an algorithm can be thought of as a $O(N_t N_\nu N_{\text{DM}})$ problem where N_t is the number of time samples, N_ν is the number of frequency channels and N_{DM} is the number of DM trials (e.g. Magro et al., 2011). A well established solution to reducing the complexity of such a problem is using a tree algorithm. Similar to the concept behind the Fast Fourier Transform (Cooley & Tukey, 1965), each individual computation can be simplified to a problem with two parts similar to branching from a tree. This algorithm was first proposed as a scheme for dedispersion by Taylor (1974). The scheme reduced the complexity from $O(N_t N_\nu N_{\text{DM}})$ to $O(N_t N_\nu \log N_\nu)$. The hierarchical structure of the algorithm also requires a base-two number of frequency channels. Variants of the tree-dedispersion algorithm have also been explored for optimal computational requirements. For example, Zackay & Ofek (2017) reported an algorithm termed as the Fast Dispersion Measure Transform (FDMT) which is computationally efficient without compromising on the sensitivity lost by brute-force incoherent dedispersion. Recently, the CHIME collaboration implemented a variant of tree dedispersion termed *bonsai*

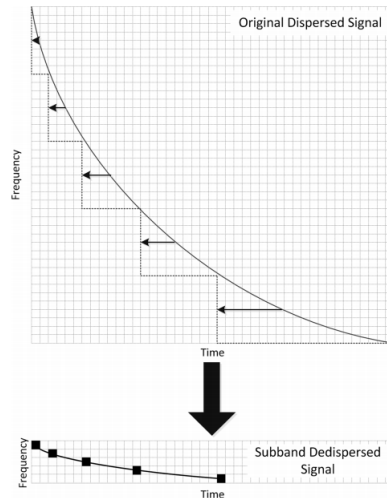


Figure 2.4: A visual illustration of the subbanding process. The frequency channels are aggregated into a set of frequency subbands where delays corresponding to a user defined nominal DM are applied for every channel in the subband. The final subbanded file is then searched around the previously chosen nominal value. Figure taken from [Magro et al. \(2011\)](#).

(K.M. Smith et al., in prep) where the problem is tackled in the $f^2 - t$ space rather than the $f - t$ space ([CHIME/FRB Collaboration et al., 2018](#)).

- *Implementation on advanced architectures*

Robust implementation of the dedispersion algorithm through different computer architectures has also been explored. Multicore CPU machines have been utilised to implement dedispersion via a parallel computing platform (e.g. [PRESTO Ransom, 2001, 2011](#)). [Barsdell et al. \(2010\)](#) delved into the intricacies of using multicore architectures and showed that the algorithm is best parallelised over DM and time dimensions along with sequential summing over frequency channels. They also demonstrated a use-case for using Graphical Processing Units (GPUs). GPUs unlike CPUs are composed of hundreds of cores capable of handling thousands of threads simultaneously, thus enhancing processing capability. [Dodson et al. \(2010\)](#) integrated such a GPU based dedispersion setup for a survey of transients. [Magro et al. \(2011\)](#) exploited the GPU architecture further and developed a brute-force as well as a subbanding based de-dispersion in CUDA⁴ which led to a performance speedup of about two orders of magnitude (between 50 and 200) compared to traditional pulsar software. [Barsdell et al. \(2012\)](#) also followed in similar lines to [Magro et al. \(2011\)](#) and analysed three versions of the GPU implementation for dedispersion, namely: direct, tree based and subbanding. They found similar or slightly faster (factor of ~ 2 -3 times improvement) speed in each version in comparison to [Magro et al. \(2011\)](#). [Bassa et al. \(2017a\)](#) implemented

⁴<https://docs.nvidia.com/cuda/index.html>

a routine (`cdmt`⁵) for coherent dedispersion on GPUs to perform semi-coherent dedispersion searches for pulsars with LOFAR. They demonstrated that the technique can be made feasible and is important at low frequency observations (< 200 MHz). They were also able to discover a new millisecond pulsar in the process [Bassa et al. \(2017b\)](#). Some of the latest pulsar search pipelines have incorporated a GPU setup for dedispersion (`DEDISP`⁶) and have helped discover pulsars from reprocessings of old survey data (e.g. [Morello et al., 2019](#)).

2.2.4 Barycentering

Prior to searching for periodicity after dedispersion, the reference frame for the observed data is changed to the Solar System barycentre (SSB). This is needed to account for Doppler shift induced changes to the pulse spin period due to the Earth’s rotation and revolution around the Sun. To compensate for this effect, a resampling of the dedispersed timeseries is done, thus changing the reference frame from topocentre to barycentre. The effect becomes more prominent as the length of the observation increases. The contribution due to the Earth’s motion can be studied carefully in followup observations of the discovered pulsar via pulsar timing (see introduction to Chapter 6). More details about the barycentering process can be found in [Lorimer & Kramer \(2012\)](#).

2.2.5 Discrete Fourier Transform

Once dedispersion is performed at a particular DM trial, the timeseries needs to be searched for periodic pulsations. Given that astronomical signals are weak, an unknown underlying periodicity in the timeseries is difficult to detect. Transforming this information from the time domain into the frequency domain is convenient and advantageous since a pulsar signal would show a spike in power confined to a frequency value representing the intrinsic spin of the pulsar. The Fourier Transform (FT) is thus a powerful technique that finds an application in this scenario. Given that the timeseries is finite and discrete, a Discrete Fourier Transform (DFT) is applied on the dedispersed timeseries. Assuming a uniformly sampled timeseries n_j , every k^{th} element of the DFT can be expressed as :

$$F_k = \sum_{j=0}^{N-1} n_j e^{-2\pi i j k / N} \quad (2.11)$$

where k is the equivalent Fourier frequency. Since the DFT output is complex, the equivalent power spectrum i.e. the power per Fourier bin can be calculated as

$$P_k = \text{Re}(F_k)^2 + \text{Im}(F_k)^2 \quad (2.12)$$

The most common implementation of the DFT is the Fast Fourier Transform algorithm ([Cooley & Tukey, 1965](#)). The algorithm simplifies the number of operations

⁵<https://github.com/cbassa/cdmt>

⁶<https://github.com/ewanbarr/dedisp>

from $O(N^2)$ to $O(N \log N)$ owing to its tree-based hierarchy. The algorithm is generally implemented in the Radix-2 format where a DFT of size N is broken down into two DFTs of size $N/2$ in a recursive manner until the smallest computational element becomes a DFT of size 2. Another feature of the DFT that helps speeding up computation is that a real valued signal with N samples would only require the first $N/2$ Fourier components to be computed owing to the symmetric nature of the DFT about the maximum frequency component. From a hardware-based perspective, modern FFT implementations have also been developed for GPUs for enhancing the processing speed. One such example is the the CUDA Fast Fourier Transform library abbreviated as CuFFT ⁷ which has been an integral part of pulsar search pipelines (e.g. Bassa et al., 2017a; Morello et al., 2019).

However, there exists a downside to the DFT implementation for search algorithms. Since the Fourier domain contains power in discrete frequency bins, pulsar spin frequency signals that do match the centre frequency of a bin would be degraded in signal-to-noise ratio (also termed as scalloping). In the worst-case scenario where the spin frequency lies exactly at the midpoint between two frequency bins, the degradation in Fourier power is close to 60 per cent (see e.g. Ransom et al., 2002). These complications can be solved by some of the techniques described below:

- Zero padding: The timeseries can be appended with zeros or a constant value in order to improve the resolution in the Fourier domain post the DFT operation. This has an added advantage that the relative strength between Fourier components is unaffected and shifted by a constant offset in the worst case scenario. However, the zero padding technique is not suitable for very long datasets where an increased FFT length increases the computation time. A corollary to this technique is to append data upto a point where the final timeseries has a length that can be expressed as a power-of-two. This becomes beneficial for the FFT implementation.
- Interpolation: In this technique, the S/N for a signal lying between frequency bins is boosted by using the information from the closest neighbouring bins to the frequency of interest. One such example is interbinning where the Fourier amplitude (\mathcal{A}_k) at half-integer frequencies can be approximated as (Ransom et al., 2002):

$$\mathcal{A}_{k+\frac{1}{2}} \simeq \frac{\pi}{4} (\mathcal{A}_k - \mathcal{A}_{k+1}) \quad (2.13)$$

An in-depth discussion of various such rectification schemes along with their advantages and drawbacks can be found in Ransom et al. (2002).

Although the above mentioned schemes can help recover the power, an underlying assumption associated with these schemes are that samples in the Fourier domain are not independent of one another. For practical purposes, it is assumed that this has a relatively insignificant effect on the statistics of the data (Lorimer & Kramer, 2012).

⁷<https://docs.nvidia.com/cuda/cufft/index.html>

2.2.6 Red noise removal

The dedispersed timeseries on close examination generally reveals long term fluctuations in the baseline of the data. When transformed to the Fourier domain, the power spectrum obtained shows a skewed profile dominated by low-frequency components (see Figure 2.5). This *red noise* is a consequence of instabilities due to a host of parameters within the signal chain including the power mains (~ 50 Hz frequency). Sensitivity to long period pulsar signals ($P > 1$ s) is degraded as their spectral power overlaps with the reddened area of the spectrum.

There are numerous techniques that can be employed to deredden the data. The lowest frequency component corresponding to the first bin in the power spectrum represents the DC offset. This information is immaterial for pulsar searching purposes and is hence masked out. A standard technique to whiten the data is to apply a running median filter with a certain window size along the timeseries. While it is computationally cheaper to use a running mean, the median offers the advantage to being agnostic to few outlier points (possibly time domain RFI bursts). The chosen window size is a trade off between a valid estimation of the median and unperturbed sensitivity to long period pulsars. The same technique can also be applied in the Fourier domain where the Fourier coefficients are scaled to ensure that the final power spectrum follows a chi-squared distribution with two degrees of freedom (Israel & Stella, 1996; Ransom et al., 2002). One such use-case was applied for a reprocessing of HTRU data (Morello et al., 2019) where the factor is calculated as:

$$B_i = A_i \sqrt{\frac{2 \ln 2}{\text{median}_{i-m \leq k \leq i+m} \{|A_k|^2\}}} \quad (2.14)$$

where A_i are the complex Fourier amplitudes of the dedispersed time series and B_i represent the amplitudes post dereddening. Additionally, the Fourier power around a particular bin i is estimated by taking a running median within a given window size $2m$. A small window size is preferred for the low frequency end of the power spectrum due to the steepness of the profile. As one approaches the higher frequencies, the spectrum flattens out allowing not only for increasing the window size without biasing the median calculation but also speeding up the dereddening process. Modern pulsar software have also integrated the dereddening task in GPUs by parallel processing each running median filter task per chunk of data (e.g. PEASOUP). This provides an additional step up to the processing speed which is crucial when dealing with high fidelity long datasets.

2.2.7 Harmonic Summing

By the property of the Fourier Transform, a train of pulses in the time domain would translate to a train of delta functions in the Fourier domain and modulated by the pulse shape. However, besides having an underlying periodicity, individual pulses from pulsars have a variety of shapes. Most pulse widths are also a small fraction of the total rotational period (i.e. a low duty cycle). This leads to an uneven distribution of power

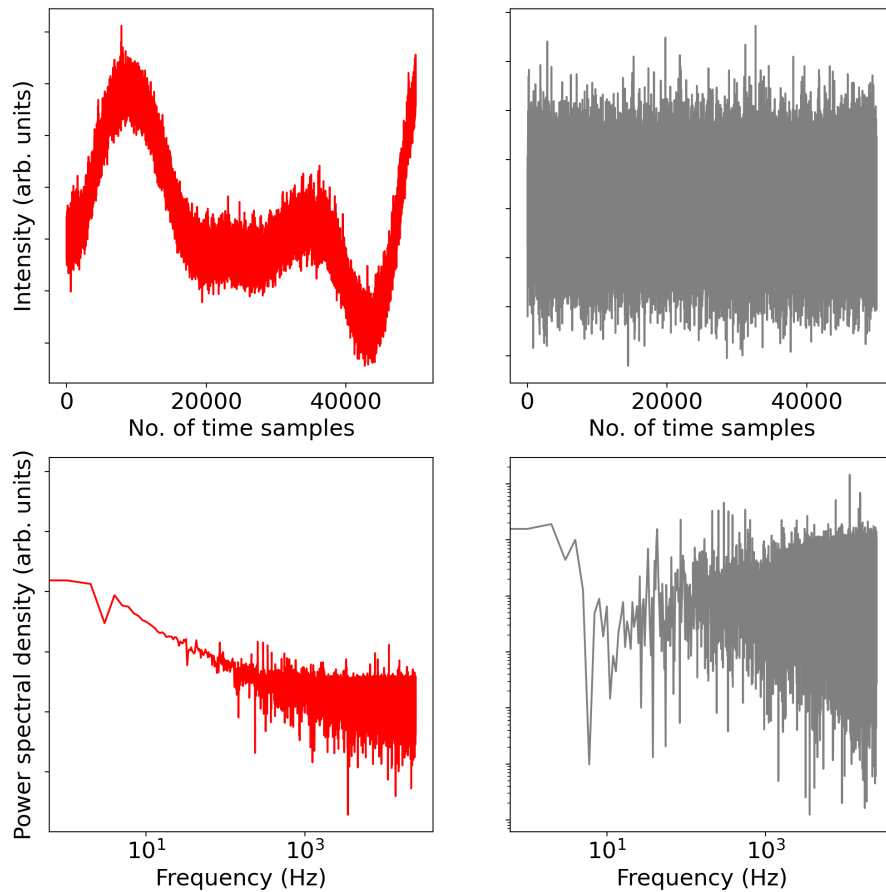


Figure 2.5: The impact of red noise on an observation with MeerKAT is shown here. *Left:* The top panel represents a zero DM timeseries. The timeseries shows variations in the baseline which is reflected in the relatively high power in the low frequency end of the Fourier spectrum (the bottom panel). *Right:* After dereddening, the timeseries behaves like white noise which is reflected in the power spectral density being uniform across spectral frequency.

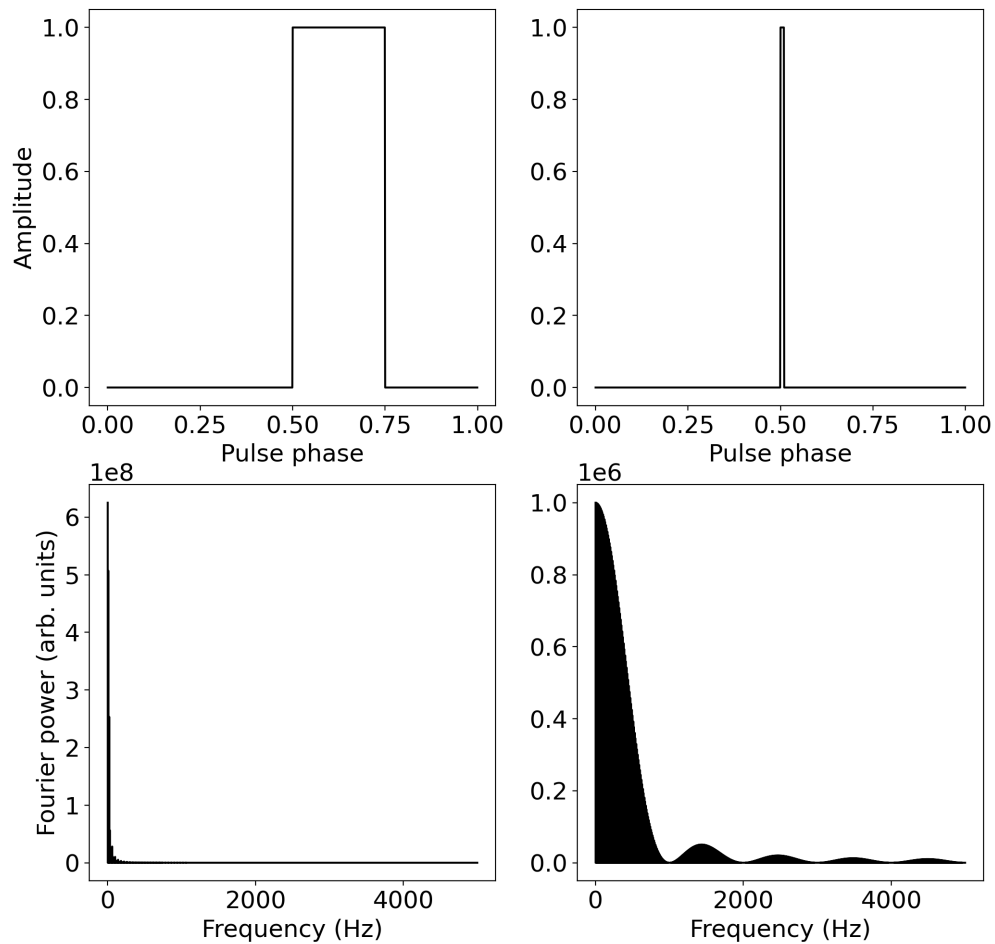


Figure 2.6: Effect of the width of a top-hat pulse train on the power across higher harmonics is demonstrated here. Left: A top-hat pulse with a duty cycle of 25 per cent is shown in the top panel and the Fourier spectrum of a train of these pulses is shown in the bottom panel. Most of the power is concentrated in the fundamental frequency. Right: A top-hat pulse with a duty cycle of 1 per cent is shown in the top panel and the Fourier spectrum of a train of these pulses is shown in the bottom panel. Unlike, the wider pulse, the Fourier power spreads to much higher harmonics than before.

of the pulsar signal among the fundamental spin frequency along with higher order harmonics (see Figure 2.6). This spread in power reduces the sensitivity to detecting pulsars if the fundamental frequency alone is taken into consideration. The most common solution to resolve this problem known as incoherent harmonic summing (Taylor & Huguenin, 1969). The incoherence stems from the fact that phase information is lost when intensities are computed. In this setup, the power spectrum is stretched by a factor of 2 thus creating modified versions of the original spectrum. These spectra are then stacked together thus boosting the signal from the pulsar. The empirical norm that is set for most pulsar searches is using upto 16 harmonics (e.g. Ransom et al., 2002). A detailed probabilistic approach to harmonic summing can be found in (Yu, 2018).

2.2.8 Folding

Once a potential peak in the Fourier spectrum is identified to have a false alarm probability lower than the set threshold, it is customary to follow up on the set of retrieved parameters to confirm if the pulsar signal is actually observed in the original data. A robust technique for confirming a detection is *folding*. Folding is the process of computing the average synchronous to the detected potential pulsar spin period signal. This procedure would improve the signal to noise ratio of a pulsar profile since the signal always adds up with every stack and the noise would be suppressed (assuming Gaussian statistics).

The pulsar candidate data cube (also known as a folded archive) obtained from folding contains information along three axes: observing frequency (ν), observing time (T) and rotational phase (ϕ). A standard way to visualise the datacube is to generate diagnostic plots showing projections of the data cube along a pair of axes (see figure 2.7). Given that most pulsars are broadband in nature and persistent in time (unless they emit sporadically like nulling and intermittent pulsars), the $\nu - \phi$ and $T - \phi$ projections in particular show an immediate signature of a possible new pulsar discovery. The folding routine is also typically designed to reduce the frequency and time resolution by summing over to generate a small number of time sub-integrations and frequency subbands. This reduces the size of the datacube and also makes it convenient for visualisation due to the increase in the brightness per unit of the projection plots.

Metadata information are also printed alongside the projection plots giving information about the relevant header parameters (e.g. source coordinates) and the input parameters used for folding (e.g. spin period, DM). Additionally, the folding procedure requires an optimisation scheme at the end of the stacking in order to fine-tune the set of parameters that yields the best S/N for the pulsar profile. The metrics used for optimising are also reflected in the diagnostic plot in order to ease decision making (see figure 2.7). For example, the best DM and period are estimated using the value with the highest reduced χ^2 value where χ^2 indicates how well the pulse profile fits with Gaussian white noise. The significance of the optimised pulse profile can be adapted from equation 2.17 and written as Barr (2013):

$$p(S/N > S/N_{\min}) = \frac{n_{\text{bins}}}{2} \operatorname{erfc}\left(\frac{S/N_{\min}}{\sqrt{2}}\right) \quad (2.15)$$

where n_{bins} are the number of bins in the folded profile and erfc is the error function.

With growing numbers in candidates produced as well as increased resolution in the filterbank outputs, folding routines are time-consuming. From the computational aspect, the folding routine cannot be translated directly to an embarrassingly parallel problem unlike dedispersion but can be implemented as a distributed computing routine. This is because each sub-division of the data put through the process would eventually need to be stitched with the other divisions, thus causing a lag in compute time. Pulsar software have been adapted for multi-core and multi-threaded functionalities on CPU (e.g. `prepfold` from `PRESTO`) as well as GPUs (e.g. `DSPSR`) that fold multiple candidate parameter files to generate several folded archives simultaneously for further visual inspection.

2.2.9 Candidate Sifting

Modern pulsar surveys can generate millions of possible pulsar candidates. Empirical evidence based on statistics from previous pulsar surveys have shown that most of these candidates show a heavy class imbalance tilted in favour of noise and RFI based candidates (Lyon et al., 2016). This demands a mechanism for sifting through the candidates and selecting a fraction of these candidates that are worth inspecting for new discoveries.

The decision-making strategy that goes into choosing appropriate candidates can be adapted on two levels. Folding can be a rate determining step for pulsar search pipelines and would hence benefit from a reduction in candidates put through the routine. Similarly, while candidate diagnostic plots are traditionally inspected by eye, a second layer of sifting schemes following folding can prove to be an added advantage to speed up and automate the inspection process. Discussed below are two broad set of schemes that can be applied preceding and succeeding the folding routine:

2.2.9.1 Pre-folding sifting

There are numerous sifting schemes that can be employed before folding. An initial cut is made based on a threshold for the S/N of the candidate. A standard cut is then applied to reject all candidates that are produced from a DM trial value of zero since a zero DM candidate indicates a terrestrial signal detection. This zero-DM cut can also be extended to a slightly higher value (say 3-4 pc cm^{-3}) assuming that previous surveys have already had the sensitivity to discover any pulsars within this DM range. A set of RFI frequencies from sources like satellites and mobile towers can also be used to cross match with the frequencies of the generated candidates. Given that the RFI signals are powerful and produce sharp features, candidates that match the higher order harmonics of the fundamental RFI frequency are also eliminated.

Despite applying multi-beam RFI excision techniques (as discussed in 2.2.2), a large number of RFI candidates are produced from the search owing to power in higher

order harmonic frequencies and spectral leakage. Surveys with multi-beam receivers thus need to employ spatial-filtering techniques to eliminate candidates. The idea is to group together candidates that are harmonically related and also keeping a tab of the beam that generated the candidate. Candidates that are RFI would most likely show a uniform illumination for all the beams (see Figure 2.8). This provides a unique distinguishing feature that can help eliminate RFI candidates. Although a true pulsar candidate is likely to be found in a beam closest to the true source coordinates with a certain S/N, the same candidate or its harmonics can still be detected in multiple beams with tapering S/N values (see bottom left plot in Figure 2.8). A possible solution here is to specifically choose the candidate in the beam with the highest S/N for folding (see Chapter 3 for an application of this scheme). A much more accurate and rigorous scheme for spatial filtering is to uniquely characterise the point spread function (PSF) of each beam and overlay the pattern of each beam to form a final superposed pattern. Candidates of similar periodicity can be grouped and their spatial profile can be cross matched with the beam pattern to help distinguish between pulsar and non-pulsar candidates.

Filtering out the known pulsars in the field is generally not a norm since any known pulsar detections can help benchmark the sensitivity of the end-to-end search pipeline. Although, some well known bright pulsars like PSR J0437-4715, Crab and Vela can be responsible for producing numerous harmonically related candidates. Similar to a set of RFI frequencies, one can thus also provide a list of bright known pulsars (with a cap on the number of harmonics) that are expected from the data recorded on a certain field of view.

2.2.9.2 Post folding sifting

The traditional approach to sniffing out possible new pulsar discoveries is humans visually inspecting a diagnostic plot associated with a folded archive. However, this task becomes tedious when dealing with candidate numbers ranging from several thousands to millions. The amount of human labour invested for such a task is detrimental from the aspect of time required as well susceptibility to human error. This has led to the surge of automated classifiers that provide the best possible accuracy and speed for ranking pulsar candidates. Besides using human-based scoring algorithms, the use of machine learning/artificial intelligence (ML/AI) have in particular provided a flexible platform to better model the pulsar candidate phase space (See Chapter 5 for a detailed overview of the techniques used). The reduced set of candidates which rank high among the ML classifiers used are then visually inspected and re-observed for confirmation.

2.2.10 False alarm probability and minimum S/N

It is important to understand the statistics that lie behind the Fourier bins, in order to estimate the significance of a signal amidst the noise. It is well known that the sum of squares of n independent variables with Gaussian probability distribution functions (PDFs) follows a χ^2 distribution with n degrees of freedom. Thus one can

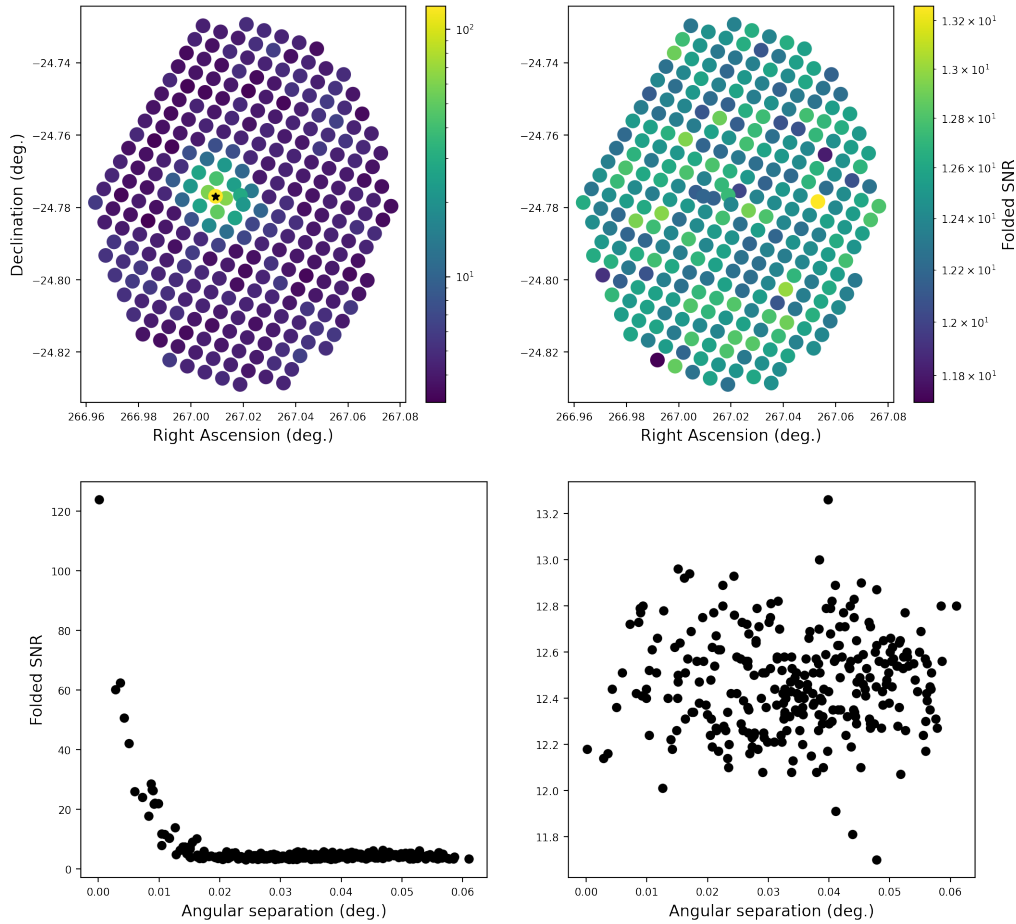


Figure 2.8: The difference between a known pulsar signal and RFI signal seen across multiple beams is demonstrated. Top panel represent the S/N per synthesised beam from the MeerKAT beamformer (see Chapter 3) as a function of beam position. The top left panel is obtained from folding the ephemeris of Terzan 5A on all beams for the first 20 seconds of data. The black star shows the expected position of the pulsar. The top right panel is a similar fold obtained by folding modulo a frequency corresponding to RFI. A note is that the filled circles are not representative of the true shape of the beam. The bottom panel represents the folded S/N as a function of angular separation in reference to a fixed beam position. The bottom left plot shows the function in reference to the position of the beam closest to Ter5A. The bottom right plot shows the function for a position chosen at random.

work out the probability of a pulsar candidate showing up as a consequence of a noise fluctuation better known as the false alarm probability. When dealing with Fourier power, there are two degrees of freedom (referring to the real and imaginary Fourier components). Given that harmonic summing adds another dimension to the degrees of freedom, the false alarm probability is given as (Ransom et al., 2002):

$$p_{\text{false}}(P > P_{\text{min}}) = \sum_{j=0}^{m-1} \frac{(P)^j}{j!} \exp(-P_{\text{min}}) \quad (2.16)$$

where m is the number of harmonic folds performed. Alternatively, the significance of a signal can also be expressed as a function of the signal-to-noise ratio (S/N). For example, a pulsar search done by estimating Fourier amplitudes rather than the power and using S/N instead of P_{min} can be expressed as (Lorimer & Kramer, 2012):

$$p_{\text{false}}(S/N > S/N_{\text{min}}) = \exp\left(-[\sigma_{\mathcal{A}}S/N_{\text{min}} + \overline{\mathcal{A}}]^2\right) \quad (2.17)$$

An important point to note here is that the S/N and P_{min} calculations would be affected by RFI, thus increasing the threshold levels from the estimated values. The threshold is also dependent on the total number of trials going into a search setup. While a periodicity search alone is mainly dependent on the number of DM trials and the number of Fourier components, adding more dimensions to the search space with acceleration searches (see 2.2.11) and other higher dimension searches (see section 2.2.12) would proportionally scale the total number of trials, thus pushing the threshold levels further up.

2.2.11 Acceleration searching

The searching procedure described so far has assumed that the pulsar is isolated. Thus, the spin period is treated to be constant during the course of an observation. While changes are expected in the spin period when the reference frame is not barycentered (as explained in section 2.2.4), a much more prominent modulation is observed for a pulsar in a binary system. Pulsars in binary systems can have orbital periods ranging from several years like PSR B1259–63 (Johnston et al., 1992) to as short as 75 min (PSR J1653–0158; Nieder et al., 2020). This implies that the signals obtained from binary pulsars if uncorrected for Doppler modulation, would be smeared in time as well as the Fourier domain during the course of hours of observation (see figure 2.9). One thus needs a different approach to restore the sensitivity since the previously discussed routines do not suffice.

Depending on the time-scale of the orbital period (P_b) compared to the observing time (T_{obs}), different routines are applied to better model the deleterious effects of binary motion. We discuss here the regime where $T_{\text{obs}} < P_{\text{orb}}/10$ (Ransom et al., 2003) and the binary effects can be simplified to a change in one particular dimension i.e. acceleration. The acceleration search technique forms a primary part of the thesis and is described here in detail.

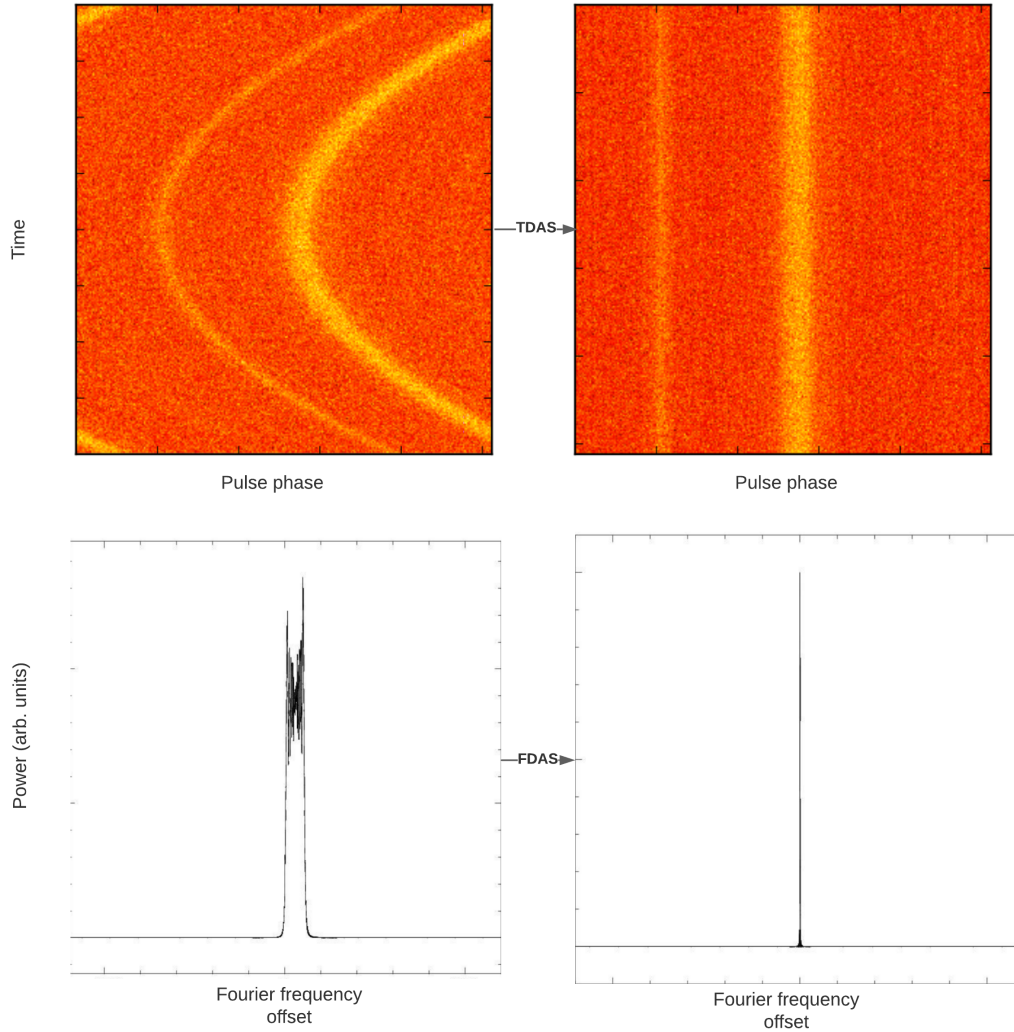


Figure 2.9: A visual representation of the outcome from the acceleration search techniques. *Top panel* : The effect of time domain resampling is shown. The left plot shows a simulated accelerated pulsar signal. The right plot shows the same signal when re-sampled based on the assuming a linear acceleration when correcting for the Doppler effect. Credits to original figure: E. Barr *Bottom panel*: The effect of Fourier domain acceleration search is demonstrated. The left plot shows the power spectral density of an accelerated signal. The power is spread across multiple neighbouring frequencies owing to the acceleration. The right plot shows the power spectral density once, the acceleration is taken care of. The power peaks at the intrinsic spin frequency of the pulsar. Credits to original figure: S. Ransom

2.2.11.1 Mathematical formulation

As mentioned previously, binary pulsars undergo Doppler modulation depending on the direction of orbital motion with respect to the line-of-sight of the observer on Earth. The relation between the observed spin period (P_{obs}) and the true spin period of the pulsar (P_{true}) can be expressed as:

$$P_{\text{obs}}(t) = P_{\text{true}} \left[1 + \frac{v_l(t)}{c} \right] \quad (2.18)$$

where v_l is the velocity of the pulsar along the line of sight. This velocity can in-turn be expressed as a Taylor series expansion (e.g. [Johnston & Kulkarni, 1991](#)):

$$v(t) = \left(\frac{dv}{dt} \right) t + \left(\frac{d^2v(t)}{dt^2} \right) \frac{t^2}{2} + \left(\frac{d^3v(t)}{dt^3} \right) \frac{t^3}{6} + \dots \quad (2.19)$$

If the modulation is assumed to be affected by a constant acceleration alone, the contribution due to derivatives higher than the dv/dt term are considered to be negligible thus simplifying equation 2.19 to

$$v(t) = at \quad (2.20)$$

where the acceleration a refers to the contribution from the radial acceleration and not the total acceleration of the pulsar.

2.2.11.2 Algorithms used

The contribution due to acceleration can be removed from the recorded filterbank data in different ways. The algorithms required to achieve this can be broadly classified based on the domain where the correction factor is applied i.e. along the time or Fourier domain. We discuss the main features of these techniques below.

1. Time domain acceleration search (TDAS):

The key idea in this approach is to re-sample a dedispersed timeseries to the inertial frame of the pulsar. By applying this technique, one should obtain a time series as if the pulsar has no orbital motion and thus obtain the pulses from its intrinsic rotation. This would also imply that the standard techniques following dedispersion can directly be applied to this modified timeseries. From the algorithm perspective, equations 2.18 and 2.20 can be adapted as ([Camilo et al., 2000](#)):

$$\tau(t) = \tau_0(1 + at/c) \quad (2.21)$$

where $\tau(t)$ is a time interval in the pulsar frame. The constant τ_0 is used for the purpose of normalisation and its value depends on the phase from which the re-sampling scheme is applied. For example, ([Camilo et al., 2000](#)) applied a condition where $\tau(t)$ is the same as the sampling interval without binary motion

(t_{samp}) at the centre of the observing length. This condition subject to equation 2.21 becomes:

$$\tau_0 = \frac{t_{\text{samp}}}{1 + aT/2c} \quad (2.22)$$

where T is the total length of the observation. The new timeseries can either be generated using a linear interpolation scheme (Middleditch & Kristian, 1984) or by adding and removing samples to account for the phase drift. The prime advantage of the time domain method is the computational ease of implementing and easy mapping from CPU (e.g. SIGPROC⁸) to GPU (e.g. PEASOUP⁹) based architectures that are required for realtime processing of the data. The technique has also been applied successfully for many pulsar surveys (Camilo et al., 2000; Manchester et al., 2001; Eatough et al., 2013a; Ng et al., 2015) that have yielded numerous binary pulsar discoveries including the most accelerated binary pulsar currently known (Cameron et al., 2018). The TDAS technique forms the crux of the searches done as part of this thesis (see Chapters 3 and 4). A summarised workflow of the entire chain incorporating TDAS is shown in Figure 2.10.

2. Fourier Domain Acceleration Search (FDAS):

While periodic pulsations would emerge as a peak in the Fourier spectrum, pulsations imposed with a linear acceleration lead to a range of observed spin periods leading to a spread in power over multiple spectral bins. Mathematically, the number of Fourier bins over which the power is spread (denoted as z) for a given acceleration (a) can be given as (e.g. Ransom et al., 2002):

$$z = t_{\text{obs}}^2 \frac{ah}{cP} \quad (2.23)$$

where t_{obs} is the observation length, h is the harmonic number and P is the spin period of the pulsar. Ransom et al. (2002) first utilised this feature to generate different templates corresponding to different values of z showing differing patterns in the Fourier spectrum. In order to find the true acceleration of the pulsar, a match filtering technique is applied where the Fourier spectrum of the uncorrected time series is cross-correlated with the template to collect the power back into a single frequency bin. The z -based template that best matches the spectrum can be used to retrieve the corresponding acceleration of the pulsar. While TDAS requires an FFT operation for every acceleration trial value used, FDAS is computationally much more efficient requiring the FFT operation to be applied once in the entire process. The FDAS technique forms the core of the PRESTO package which has been instrumental in discovering numerous pulsars over the last 2 decades (Ransom et al., 2005; Keith et al., 2011; Barr et al., 2013; Coenen et al., 2014, e.g.) including the latest globular cluster pulsar discoveries

⁸<http://sigproc.sourceforge.net/>

⁹<https://github.com/ewanbarr/peasoup>

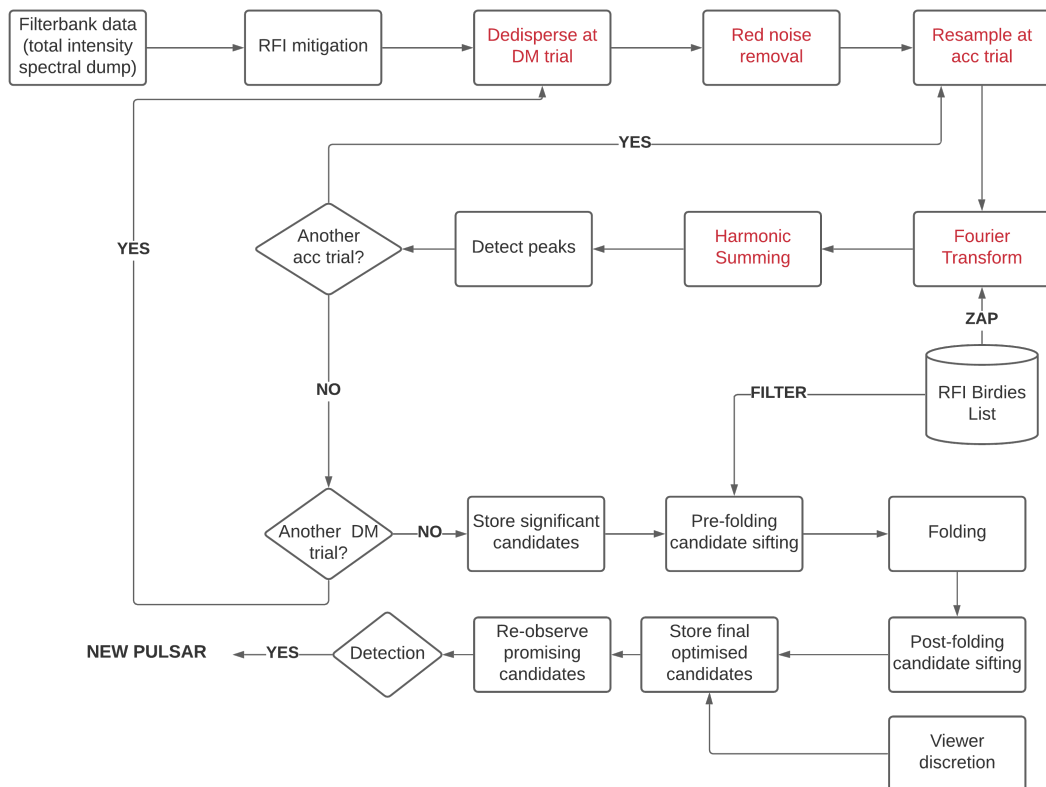


Figure 2.10: A standard workflow of the acceleration searching pipeline based on time domain resampling. This procedure forms the crux of all searches conducted as part of this thesis. Processes that have been transferred to Graphical Processing Units (GPUs) are highlighted in red.

from MeerKAT. The flagship discovery using this technique is the fastest spinning pulsar currently known (Hessels et al., 2006).

Similar to TDAS, the usage of GPUs has also been explored for FDAS. The accelsearch routine of PRESTO has been incorporated to run on GPUS¹⁰. Recently, Dimouli et al. (2018) have explored efficient ways of implementing the FDAS technique on GPUs as part of the AstroAccelerate suite of software. They particularly found a significant speed up when utilising the shared memory of GPUs that removes problems with overheads when using CuFFT based operations.

2.2.11.3 Optimisation

Similar to dedispersion, the acceleration search technique can also be optimised to ensure a balance between computing time and degradation in sensitivity. Described below are key strategy points to consider when planning an acceleration search.

- *Acceleration range*

The acceleration experienced by a binary pulsar is dependent on its orbital parameters. This implies that a limit can be imposed on the maximum acceleration value attainable for a fixed set of these parameters. For example, the maximum orbital acceleration (a_{\max}) in a circular edge-on orbit (i.e. the inclination i between the orbital plane and plane of the sky is 90 deg), using Kepler's third law can be expressed as (e.g. Ng, 2014):

$$|a_{\max}| = \left(\frac{2\pi}{P_{\text{orb}}} \right)^{4/3} (T_{\odot} f)^{1/3} c \quad (2.24)$$

where P_{orb} is the orbital period, T_{\odot} is a constant ($4.925490947 \mu\text{s}$), c is the speed of light and f is the mass function given by

$$f = \frac{(m_c \sin i)^3}{(m_p + m_c)^2} \quad (2.25)$$

where m_p is the pulsar mass and m_c is the mass of the binary companion. Based on a certain observation length (T_{obs}) which is subject to $T_{\text{obs}} < P_{\text{orb}}/10$, the acceleration bounds ($\pm a$) are fixed. Important point to note here is that this approximation does not bode well with eccentric binary systems where the acceleration at periastron reach very high values compared to other locations in the orbit. However given that this ultra-high acceleration phase lasts for a short time compared to the orbital period (as a consequence of Kepler's third law), the bounds of the acceleration search parameter space calculated from equation 2.24 are considered to be reasonable (e.g. Ng, 2014).

¹⁰https://github.com/jintaoluo/presto_on_gpu

- *Step size*

Once the bounds of the acceleration values are fixed, the number of trials between the selected range needs to be decided. Similar to conditions applied while calculating the step size for DM trials, the acceleration step size which is too short would drastically increase computational time. On the other end, a sparse step size would not yield the best detection even if corrected for the nearest possible acceleration value.

A simple argument to decide the step size from the FDAS point of view is to ensure that the number of Fourier bins drifted does not exceed one bin. This condition applied to equation 2.23 becomes:

$$\Delta a < \frac{cP}{ht_{\text{obs}}^2} \quad (2.26)$$

Similarly from the TDAS perspective, a step size is chosen such that an acceleration value lying midway between two trials would not smear the signal beyond a threshold. Since the S/N is at least limited by the sampling interval, the threshold factor can be determined based on a factor of the sampling interval. For example, [Eatough et al. \(2013a\)](#) set this threshold to be four times the sampling interval. Owing to the quadratic nature of the smearing across an observation for a certain acceleration, half of the observation would be smeared less than the sampling interval. If the reference for resampling is calculated with respect to the midpoint of the observation, the step size Δa can be shown to be ([Eatough et al., 2013a](#)):

$$\Delta a = \frac{64ct_{\text{samp}}}{T_{\text{obs}}^2} \quad (2.27)$$

Analogous to using a tolerance for the DM step size (as expressed in equation 2.9), the acceleration step size can also be decided by using the formulation:

$$W_{\text{tot}} = \text{acc}_{\text{tol}} W_{\text{eff}} \quad (2.28)$$

where acc_{tol} is the acceleration tolerance parameter which is set based on the extra smearing caused due to an error in the true acceleration leading to a wider pulse width of W_{tot} instead of W_{eff} due to already existing factors.

- *Partially coherent segmented search*

A caveat to the acceleration search technique mentioned before is the ratio between the orbital period and observing time needs to be ~ 10 ($T_{\text{int}} < P_{\text{orb}}/10$) for minimal loss in sensitivity. Given that the Doppler modulation is well approximated to an "acceleration-only" effect in this regime, compact binary systems take a hit on the sensitivity achievable in this phase space. For example, a binary system like PSR J1748-2446A which is in a 1.8 hour orbit cannot be fully

decoupled from its binary motion for a given acceleration and observing length exceeding 10 min. This can be tackled by splitting up a given observation length into multiple sections/segments where each segment could be searched as a separate entity (hence the term partially-coherent). The smaller the segment, the more compact a binary system for which an “acceleration-only” approximation holds true. Another advantage is that pulses that are degraded in one segment of the observation may be found with a higher S/N in a different segment thus giving an independent confirmation of the same pulsar candidate. This is plausible for systems where certain orbital phases are better suited for the linear acceleration regime than others. The main caveat however, is the loss in sensitivity by a factor of $\sqrt{T_{frac}}$ which is given as :

$$T_{frac} = T_{obs}/T_{segment} \quad (2.29)$$

where T_{obs} is the total observing length and $T_{segment}$ is the length of the particular segment of interest. This scheme has already been deployed for targeted (e.g. [Ransom et al., 2005](#)) as well as blind surveys (e.g. [Ng et al., 2015](#); [Cameron et al., 2018](#); [Cameron et al., 2020](#)) and yielded significant new discoveries.

2.2.12 Other search techniques

- **Phase modulation search:**

When the duration of a observation is much greater than the orbital period ($T_{obs} \gg P_b$) of an unknown binary pulsar, the pulse frequency observed is not only modulated by a Doppler shift but also a higher order effect that is a reflection of multiple orbits. Based on this feature, a search can be carried out where the modulation of the pulses on an orbital period envelope would show a unique spectrum in the Fourier domain ([Jouteux et al., 2002](#); [Ransom et al., 2003](#)). Performing a DFT once again on this Fourier Spectrum gives a direct estimate of the orbital period.

- **Slide and Stack search:**

A variant of the standard FDAS technique is to split up the observation into multiple segments, independently generate the Fourier spectrum per segment and account for the change in the peak spin frequency of the unknown binary pulsar when moving from one segment to another. The drift in the peak Fourier bin ($\dot{\nu}_{drift}$) per segment when assuming a linear acceleration regime can be given as adaptation of equation 2.23 ([Faulkner et al., 2004](#)):

$$\dot{\nu}_{drift} = \nu_0 a_1 T / (n_{seg} c) \text{ Hz} \quad (2.30)$$

where n_{seg} is the number of segments across a total observing length T . Each segment spectrum is slid in accordance with the drift and combined to form a ‘stacked’ spectrum. The main assumption in this technique is that the segment

length chosen is compatible with a regime where the binary motion of the pulsar has no weight. The main drawback of this technique is the loss in sensitivity owing to restriction in the individual lengths of data to form the spectra.

- **Dynamic Power spectrum search**

Binary pulsar search techniques discussed so far have dealt with the extremities of the relation between the orbital period and the observation length ($T_{obs} \gg P_b$ and $T_{obs} \ll P_b$). This calls for a gap to be filled in by a technique that is suited for a regime where T_{obs} is roughly of the same order as P_{orb} . The dynamic power spectrum search derives initial principles from the stack search by forming spectra over multiple segments of an observation. Instead of scrunching to form a combined spectrum, each spectrum from the respective segments is now a 1D subarray of a 2D spectrum formed between frequency and time. Any binary pulsar signal would be visible as a sinusoidal pattern in this 2D spectrum. The length of the sinusoid would depend on the orbital phase upto which the total observation length is covered. This scheme has yielded discoveries in globular cluster M62 (Chandler, 2003) and also been implemented for certain pulsar surveys (e.g. Pennucci, 2009).

- **Jerk search**

A step higher from the standard acceleration search is to also include the acceleration derivative term when accounting for the modulation observed for the line of sight velocity v_l . This implies an additional term to equation 2.20 and the expression is now written as:

$$v_l = at + \frac{jt^2}{2} \quad (2.31)$$

with j being the acceleration derivative better known as jerk. Similar to the acceleration search, the jerk parameter space can also be explored in the time as well as frequency domain. In the time domain, the resampling procedure expressed in equation 2.21 can be rewritten as :

$$\tau(t) = \tau_0(1 + at/c + jt^2/2c) \quad (2.32)$$

In order to optimise on the step size, the conditions applied for the linear acceleration regime can similarly be extended for the jerk term where the smearing of the pulse across an observation is cubic in nature (see e.g. Eatough et al., 2013a). Andersen & Ransom (2018) reinterpreted the jerk formulation for the Fourier domain where apart from a z parameter (as described in equation 2.23) they also account for $\dot{z} = w$ that is expressed as:

$$w = t_{obs}^3 \frac{jh}{cP} \quad (2.33)$$

The jerk search has demonstrated an improvement in sensitivity over the acceleration search through analytical (Bagchi et al., 2013) as well as practical (Andersen & Ransom, 2018) means for observing lengths that lie approximately between 5 and 15 percent of the orbital period of the pulsar. A major caveat for this technique is the trade-off with respect to the increase in the number of trials, thus increasing the computational load (e.g. Eatough et al., 2013a). This technique is thus currently restricted for targeted searches that can afford a compromise in the number of DM trials. However, given the similar methodology to the acceleration search, the algorithms are easily transferrable to GPUs which would aid in speeding up these searches in the future.

- **Keplerian orbit search:**

The parameter space for binary pulsar searching has been explored for multiple regimes so far. While the previously discussed jerk search technique increases the sensitivity to orbits that are approximately six times the size of an observation length, there exist a niche class of objects consisting of highly relativistic and compact binary systems which have significantly low orbital periods (a couple of hours utmost) compared to the ones that are currently detected. Although a handful of such systems have been found, noteworthy ones being the Double pulsar (Burgay et al., 2003) and Terzan 5A (Lyne et al., 1990), the scarcity of such detections in the pool of known binary pulsars, calls for a new level of sophistication to the existing search schemes. These are the coherent full Keplerian orbit searches where each of the five Keplerian orbital parameters adds a new dimension to the search. While this search technique compensates for all drawbacks from previously discussed binary pulsar search techniques, it comes as a cost of being computationally very demanding (Dhurandhar & Vecchio, 2001) and has thus found no practicality relevant to modern day pulsar searching.

Alternatively, there exist simplified versions of the full orbit search which have found applications. Einstein@Home is a volunteer distributed computing project (Anderson et al., 2006) that has provided the platform to develop algorithms for implementing a coherent search for circular binary orbits (e.g. Knispel, 2011; Allen et al., 2013) thus reducing the dimensions from five to three. This technique utilises a template bank based match-filtering technique that is particularly well known in gravitational wave signal processing. This technique has already found pulsars (e.g. Knispel et al., 2013) and includes the detection of the most compact binary system (Nieder et al., 2020).

- **Fast Folding Algorithm**

Given that any new pulsar candidate confirmation involves the process of folding (Section 2.2.8), a search technique can be devised based on using different trial spin periods to fold a given dedispersed timeseries. Although such a search is coherent and highly sensitive, the process is computationally expensive. An optimised way of handling numerous spin period trials, is the Fast Folding Algorithm (FFA) (Staelin, 1969). In FFA, one splits the data length into sections where for

a given reference spin period, samples from each of these sections can be summed in multiple ways to cover different spin period trials simultaneously. Inspired by the FFT, this summation process can be converted into $O(N \log N)$ problem ensuring that:

$$N/P_0 = 2^n \quad (2.34)$$

where N is the total number of samples, P_0 is the number of samples in a section (equivalent to the number of bins in a folded profile) and n is any integer.

Although the standard FFT based periodicity searches have ruled the roost for most pulsar surveys, they do not fit well with identifying slow spinning pulsars ($P > 1$ s) for different reasons. Firstly, the slow pulsar signals feature in the low frequency end of the Fourier spectrum that is usually corrupted by red noise. Secondly, the accumulation of power for multiple pulses for a fixed observing length is much higher for millisecond pulsars than slow pulsars. Therefore, as a consequence of balancing between computation time and sensitivity to slow pulsars, FFA based searches become relevant in this specialised scenario. A detailed analysis into the pros and cons of using the FFA as opposed to the FFT search can be found in [Cameron et al. \(2017\)](#) and [Morello et al. \(2020\)](#).

- **Single pulse searches**

All the searches discussed previously have particularly hinged on the persistence of pulses with time. But there exist classes of pulsars like RRATs and intermittent pulsars (as discussed in Chapter 1) that show sporadic emission. Discovering more such sources would thus require a different approach thus leading to the use of single pulse search schemes. The standard procedure used for detecting any transient individual pulses is to apply match filtering to identify statistically significant events in a dedispersed timeseries. Such searches have been previously employed on pulsar surveys (e.g. [Thornton et al., 2013](#)). A notable highlight of single pulse searching has been the emergence of the field of Fast Radio Bursts (FRBs). The scientific prospects of FRBs have provided a positive feedback in order to further advance single pulse search techniques and which have in turn pushed the number of new discoveries by an order of magnitude (e.g. [CHIME/FRB Collaboration et al., 2019](#)).

2.3 Pulsar Software

A large number of software packages have been developed and maintained through the years as part of the pulsar searching process. Discussed below are packages that have been used particularly for this thesis.

1. **DEDISP**

The DEDISP package is an implementation of the incoherent dedispersion algorithm for GPU based processing ([Barsdell et al., 2012](#)). As explained in section

2.2.3.2, the dedispersion routine has hugely benefited from the shift to the GPU architecture in terms of improved computation speed. The original package¹¹ has been modified to enable multi-GPU capability¹².

2. PEASOUP

PEASOUP is a pulsar search library implemented in C++/CUDA that is capable of running an acceleration search using time domain resampling. This package is compatible with a dedispersed timeseries produced from DEDISP. Prior to the acceleration search, the library also performs harmonic summing and red noise removal (as shown in equation 2.14) via GPUs. In order to not compromise on processing speed, the dereddening is performed once per DM trial. Although the standard package offers a tolerance based method (see equation 2.9) for deciding the DM step size, the package has been modified to optionally run a user defined list of trials. The acceleration step size is chosen based on a tolerance based method. More details on each of the unique features of the package can be found in (Morello et al., 2019).

3. PRESTO

PRESTO (Ransom, 2011) is a suite of pulsar search software that is capable of searching for binaries in the acceleration and jerk space using the Fourier domain. For this thesis, the `prepfold` routine in particular has been used for the folding procedure as part of the Terzan5 analysis (see Chapter 3).

4. PulsarX

PulsarX¹³ is a software package written in C++ which implements an efficient version of dedispersion and folding for multithreaded and multi-core CPU architectures. The `psrfold_fil` routine has been used as part of the folding for the MGPS (see Chapter 4). This routine adds in some unique features to the folding process that are discussed below:

- a) The folding routine uses time domain and frequency domain based RFI mitigation prior to folding
- b) The diagnostic plot generated reports a unique set of metadata that includes the maximum DM and the distance along the line of sight based on the YMW model (Yao et al., 2017).
- c) An optimisation is done in the period, DM and acceleration space and the best combination of these parameters are used to generate the final folded archive file in the PSRFITS format.

¹¹<https://github.com/ajameson/dedisp>

¹²<https://github.com/ewanbarr/dedisp>

¹³<https://github.com/yymen/PulsarX>

2.4 Overview of Pulsar Surveys

Following the discovery by [Hewish et al. \(1968\)](#), astronomers around the world were further motivated to find more pulsars. The back-ends of telescopes were adapted to conduct pulsar surveys ranging different observing frequencies and different regions of the galaxy. Discussed below is a summary of different types of surveys segregated based on observing regions and types of telescopes used¹⁴.

2.4.1 Large area surveys

Large area pulsar surveys have primarily banked on covering maximum sky using state-of-the-art instrumentation and innovative techniques as leverage for better sensitivity to detect new pulsars and maintain computing feasibility. The first significant leap in early pulsar surveys was the transition from using pen-chart recorder (e.g. [Turtle & Vaughan, 1968](#)) to using computerised methods for detection of single pulses in the early 1970s (e.g. [Davies & Large, 1970](#)). Robust search techniques ([Staelin, 1969](#); [Burns & Clark, 1969](#)) also played a crucial role in the surge of new pulsar discoveries. A highlight was the second survey with the Molonglo radio telescope ([Manchester et al., 1978](#)) covering the southern sky and a fraction of the northern sky ($\delta < +20^\circ$). This survey was able to double the known pulsar population and remained the most successful survey for two decades (in terms of number of discoveries).

Improvements in surveys since then have been an inter-balance between scientific and technological drivers. For example, the discovery of the first MSP ([Backer et al., 1982](#)) pushed subsequent surveys to employ high time resolution back-ends for telescopes (e.g. [Clifton & Lyne, 1986](#)). This was later complemented by increasing the observing bandwidth owing to the broadband nature of the pulsar signal. For example, the PKS70 survey ([Manchester et al., 1996](#)) was conducted with the Parkes radio telescope at declinations below 0° using 32 MHz bandwidth, a factor of 4 better than majority of the previous surveys. Subsequent surveys focused on enhancing the bandwidth as well as the corresponding frequency resolution. This helped reducing intra-channel dispersive smearing and hence pushed the upper limit of the DM value to be searched. A prime example here is the High Time Resolution Universe (HTRU) Survey ([Barr et al., 2013](#); [Keith et al., 2010](#)) which used the Effelsberg and Parkes telescopes to ensure maximal coverage in across both hemispheres. Furthermore, the usage of multibeam receivers (7-beam at Effelsberg and 13-beam at Parkes) increased the achievable individual pointing coverage by at least an order of magnitude. With equivalent improvements in storage and computing resources, the data collected from large surveys have been archived and undergone multiple reprocessings using different search techniques (e.g. [Ng et al., 2015](#)), resulting in numerous discoveries. This has provided valuable feedback to help address the limitations of existing techniques. Apart from obtaining high fidelity data, surveys have also focused on a niche parameter space that matches telescope specifications and pulsar searching criteria. For example, the

¹⁴For a detailed list of all pulsar surveys conducted so far refer to <http://www.jb.man.ac.uk/pulsar/surveys.html>

GBNCC survey (Stovall et al., 2014) is an all sky survey ($\delta > -40^\circ$) conducted at an observing frequency of 350 MHz with the Green Bank telescope that uses periodicity as well as single pulse search algorithms. This has led to a total of 190 discoveries including 24 being new RRATs¹⁵. The high dwell time of 1 hr per pointing of the LOTAAS survey (Sanidas et al., 2019) allowed the discovery of the slowest spinning pulsar currently known (Tan et al., 2018b). Recently, the CRAFTS survey (Li et al., 2018) was initiated using a drift scan mode to cover the visible sky from The Five-hundred-meter Aperture Spherical radio Telescope (FAST). This has already led to 125 new discoveries¹⁶, paving the way for many more. Overall, the success of such large area pulsar surveys continue to drive current and future projects like the Square Kilometre Array. This would help steadily increase the known pulsar population and in turn expand the scientific applications associated with the new findings.

2.4.1.1 Galactic Plane surveys

Early blind survey results (e.g. Wielebinski et al., 1969) suggested that a significant number of pulsars are located along the galactic plane. This was further complemented by early population models (Taylor & Manchester, 1977) that suggested about 10^5 pulsars in the Galaxy with a major fraction of them lying along the galactic plane. Following this, a major fraction of blind surveys have focused or partly incorporated regions along the galactic plane. These surveys have been fruitful in yielding a large number of discoveries including pulsars with unique properties (See Figure 2.11).

The first galactic plane specific survey conducted was AR1 (Hulse & Taylor, 1974) which led to the discovery of the first binary pulsar PSR B1913+16 (Hulse & Taylor, 1975a). This was conducted with the Arecibo telescope at 430 MHz with an 8 MHz bandwidth. A major factor for planning galactic plane surveys was the mitigation of signal smearing due to propagation effects from the interstellar medium like dispersion and scattering. These deleterious effects have a greater prominence when observing along the plane owing to the higher density of matter along the line of sight. Since early surveys like the first Molonglo survey (Large et al., 1968b) used a channel bandwidth of 2 MHz without the means of removing dispersion, they were much less sensitive to high dispersion measure (DM) pulsars. The impact of this effect was best reflected by the second Molonglo survey (Manchester et al., 1978). This survey used a 4MHz bandwidth split into 1 MHz per frequency channel. This allowed for a correction of dispersion and resulted in the discovery of 155 new pulsars. The scattering of integrated pulse profiles at low frequencies also encouraged the switch to higher frequencies like L-band (e.g. Clifton & Lyne, 1986; Clifton et al., 1992) as compared to 400 MHz used by telescopes like Arecibo (Hulse & Taylor, 1975b) and Lovell (Davies et al., 1977) in the 1970s. The 1990s saw galactic plane surveys yielding discoveries via Northern Hemisphere telescopes like Effelsberg (Lorimer et al., 2000) and Arecibo (Nice et al., 1993). However, a breakthrough in pulsar astronomy came in the late 1990s with the introduction of a multibeam receiver (13 beams) at Parkes having a large bandwidth

¹⁵<http://astro.phys.wvu.edu/GBNCC/>

¹⁶<https://crafts.bao.ac.cn/pulsar/>

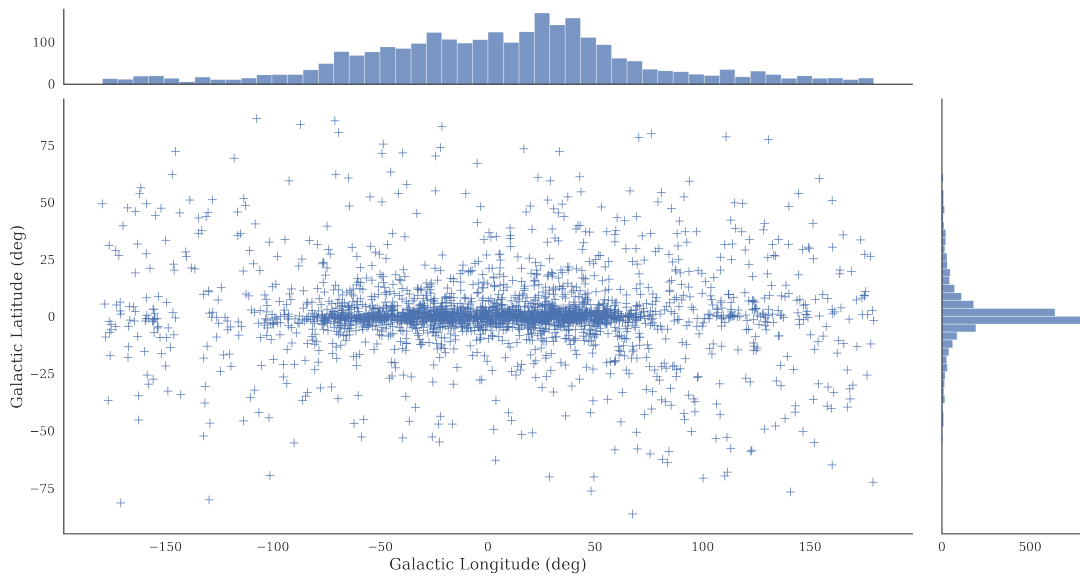


Figure 2.11: A scatter plot of all known pulsars in the Galaxy overlaid according to the Galactic coordinates. The margin plots on either axis are 1D histograms of the number of pulsars found along a fixed latitude or longitude. It is clear that the number of pulsars drop quickly away from the Galactic plane. A total of 2696 pulsars are shown here excluding those associated with Magellanic clouds and globular clusters. Data is taken from the ATNF catalogue (Manchester et al., 2005).

(288 MHz). The most prominent survey using this receiver was the Parkes Multibeam Pulsar Survey (Manchester et al., 2001). The number of discoveries in this survey exceeded 800 (Morris et al., 2002; Kramer et al., 2003; Hobbs et al., 2004; Faulkner et al., 2004; Lorimer et al., 2006). This rich data set allowed for multiple reprocessings with new techniques to extract previously missed pulsar signals (Keith et al., 2009; Eatough et al., 2010, 2013a; Knispel et al., 2013). While the HTRU survey (Keith et al., 2010; Barr et al., 2013) is aimed at maximum sky coverage, major parts of the survey have overlapping regions with the galactic plane. The HTRU South low-latitude survey in particular with a high time resolution ($64 \mu\text{s}$) and integration length (upto 72 min) has followed a similar path to PMPS (particularly in terms of the reprocessing strategy) and yielded numerous discoveries (Ng et al., 2015; Cameron et al., 2020) with the important discovery of the most accelerated binary pulsar currently known (Cameron et al., 2018).

Galactic plane surveys in the present and future aim to exploit new advancements in technological capability. This is not only in the form of improved telescope gain, multiple beams for larger sky coverage, high time and frequency resolution backends but also advanced search techniques and computing infrastructure that would be capable of handling massively high data rates and volumes. The latest ongoing survey from the Northern Hemisphere is the Galactic Plane Pulsar Snapshot survey (GPPS)¹⁷ con-

¹⁷<http://zmtt.bao.ac.cn/GPPS/>

ducted with FAST utilising a 19-beam receiver and a 400 MHz instantaneous bandwidth at L-Band (1.05-1.45 GHz). The equivalent counterpart in the Southern Hemisphere is the MPIfR Galactic Plane Survey (MGPS). A detailed description of this survey is given in Chapter 4.

2.4.2 Targeted surveys

A prominent disadvantage among blind surveys is the imbalance between sky coverage and the yield in terms of new discoveries per patch of sky covered in a single observing run. Numerous observation pointings are required to tile the entire sky. For example, the HTRU South low-latitude survey which is a subset of the HTRU south survey required 1200 pointings to cover the proposed sky coverage. Instead, targeted surveys that are aimed at particular regions which show evidence of an environment potentially harbouring neutron stars, would require a much lower number of pointings. These surveys have typically been aimed at globular clusters (GCs), gamma-ray sources, X-ray binaries and supernova remnants. Multiple pulsars associated with these regions have been found and these sources continue to be prioritised for radio pulsar searches. Summarised below is some seminal work through targeted surveys categorised based on the type of source observed.

2.4.2.1 Globular clusters

Globular clusters (GCs) have known to be a massive collection ($10^4 - 10^6$) of aged stars (10 Gyr or more), thus making them the ideal environment for the remnants of stars including neutron stars. Furthermore, the high interaction rate in these clusters allow the formation of a variety of different binary pulsar systems. In terms of the observing setup, GC observations offer a significant advantage. The beam width of a single dish telescope is usually enough to fit the entire extent of the cluster. This allows for long observations to be scheduled along with flexibility on the types of searches to be done with a specific focus on techniques to detect binary pulsars (see Chapter 3 for an example). At the time of writing, there are 223 known pulsars in 36 globular clusters¹⁸ and the road to these discoveries have produced multiple scientific and technical highlights.

The discovery of X-ray sources in multiple GCs (e.g. [Giacconi et al., 1974](#); [Clark et al., 1975](#); [Canizares & Neighbours, 1975](#)) coupled with these sources being associated with X-ray binaries ([Clark, 1975](#)) indicated a strong likelihood of finding recycled pulsars in these systems. The first MSP discovery ([Backer et al., 1982](#)) and associated evolution models ([Radhakrishnan & Srinivasan, 1982](#); [Alpar et al., 1982](#)) led to prioritising GC observations with radio telescopes (e.g. [Hamilton et al., 1985](#)). The discovery of the first radio pulsar located in a globular cluster ([Lyne et al., 1987](#)) sparked off a new dimension of enquiry for globular cluster as well as pulsar research. This discovery was an isolated millisecond pulsar (MSP) spinning at 3.05 ms in M28. Similar to the first MSP, this discovery also contradicted the existence of a binary companion associated with the MSP responsible for the recycling process. This motivated pulsar

¹⁸<https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html>

astronomers to prioritise more GC sources not only as potential pulsar factories but also to resolve the evolution mechanism of MSPs. The continued searches for pulsars in globular clusters began to steadily increase the known GC pulsar population in the late 80s and 90s (e.g. Lyne et al., 1988; Anderson et al., 1990). A significant discovery in this period was the discovery of 10 new MSPs in NGC 104 (also known as 47 Tucanae) by (Manchester et al., 1991) accounting for nearly 50% of the MSPs known at that time. Half of the discovered MSPs were in binary systems, thus adding weight to the recycling model for MSP formation. The early 2000s saw many more discoveries (e.g. Camilo et al., 2000; D’Amico et al., 2001; Possenti et al., 2003) in different GCs. New searching techniques (Ransom, 2001) began to play a key role in deep targeted searches of GCs. A significant breakthrough came with the Green Bank Telescope where 21 MSPs were found in globular cluster Terzan 5 (Ransom et al., 2005). A year later, the fastest pulsar known was discovered in the same GC (Hessels et al., 2006). This triggered extensive searches on Terzan 5 which were complemented by a swarm of new search techniques (more details in Chapter 3). The results of these searches have made Terzan 5 the record holder for hosting the most number of known pulsars in a GC (39). Other set of significant discoveries from GCs include highly eccentric binary systems ($e > 0.9$) with massive companions ($M > 1 M_{\odot}$) (e.g. DeCesar et al., 2015). These systems are most likely products of exchange encounters where a low mass companion of the MSP responsible for recycling is ejected and replaced by a more massive companion. Such systems are a consequence of dynamic interactive environments that have been unique to GCs (Verbunt & Freire, 2014).

The latest generation telescopes like FAST and MeerKAT have provided an unparalleled boost in sensitivity. Early results from GC surveys with these telescopes have shown promising signs including eight new MSPs found with MeerKAT (Ridolfi et al., 2021) and 29 new pulsars from 4 new globular clusters¹⁹. The growing numbers and associated science with GC pulsars have set up a strong foundation for future large scale globular cluster search projects (see Chapter 3 for the science case with TRAPUM).

2.4.2.2 Other sources

Besides globular clusters, some other targeted sources for pulsar surveys have included unassociated gamma-ray sources, nearby galaxies i.e. the Magellanic clouds and supernova remnants. More details on surveys with such targets and the corresponding strategy with the MeerKAT telescope is discussed in Chapter 3.

2.5 Pulsar surveys with interferometers

A majority of the surveys described earlier have been conducted with single dishes. There are multiple reasons for the dearth of pulsar surveys with interferometers. While some interferometers have served as a pulsar search instrument in the past, a primary reason for very few large scale pulsar surveys is the smaller size of the synthesised

¹⁹<https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html> also contains the list of the pulsars discovered using FAST as well as MeerKAT

beam²⁰. This would imply that the amount of time spent on observing escalates slowing down the survey. However, time-domain based pulsar searching still constitutes some of the core projects that utilise interferometers like Molonglo (Venkatraman Krishnan et al., 2020b), ASKAP (Murphy et al., 2013) and GMRT (Bhattacharyya et al., 2016). The latest addition to this are the TRAPUM (see Chapter 3) and MGPS (see Chapter 4) surveys with MeerKAT which form the core topics of this thesis.

Interferometers offer unique capabilities that can cater to pulsar surveys. These features are discussed below:

2.5.0.1 Fly’s eye and subarray modes

Interferometers typically work in unison where all dishes in the antenna array focus on observing a single patch of the sky. However, each individual dish could also act as an independent entity and observe patches of sky that are far apart. This mode of operation is also known as the Fly’s eye mode (inspired by the compound eye of the housefly). While the sensitivity can be reduced significantly, it is an advantage for conducting a quick scan of the maximum sky coverage possible. This mode has particularly been used by the Australian Square Kilometre Array Pathfinder (ASKAP) array to explore the transient sky (e.g. Siemion et al., 2012) and has also found success in discovering new fast radio bursts (e.g. Bannister et al., 2017). Recently, Qiu et al. (2019) were also able to detect single pulses originating from 11 known pulsars in this mode.

Building on the Fly’s-eye mode, a subset of dishes from the entire array could be used to survey a particular patch of the sky. In this case, coherent addition of signals from a subset of telescopes allows for a boost in sensitivity and can particularly improve the localisation of potential new pulsars. Thus the number of unique patches on sky observed simultaneously would be N/S where N is the total number of dishes in the array and S is the number of dishes within a sub-array. The sub-array mode strikes a balance between sky coverage and sensitivity and is thus a more preferable mode for pulsar observing. The VLA, GMRT as well as MeerKAT offer this capability.

2.5.0.2 Multi-beam Beamforming

With the sub-array and Fly’s eye mode, the maximum sensitivity of the interferometer is inevitably reduced. This does not bode well with pulsar searching where sensitivity is key to the possibility of new discoveries. However, practicality cannot be neglected in this scenario. As mentioned earlier, the usage of all dishes reduces the size of the synthesised beam (in comparison with the primary beam). This increases the observing time required to cover the same area of sky, thus reducing its feasibility.

One way to achieve the best of both worlds (sensitivity and angular resolution) is to form multiple synthesised beams that is able to tile a significant area of the sky. This can be done by phase coherent summing of the individual antennas. Appropriate

²⁰LOFAR has been an exception in this regard owing to its low observing frequency (140 MHz) thus having a wide field of view.

weights are used to steer the beam in a different direction. Instead of recording one such position, multiple such spatially close positions are recorded and processed. The main caveat to this setup is the surge in data rates. While this would have been a problem in the past, technology has caught up significantly to make such a setup feasible. The LOFAR telescope is an example where a multibeam infrastructure has been used for the LOTAAS survey (Sanidas et al., 2019). The survey utilises the core antennas (known as the Superterp) which enables a balance between the filling factor of the tied-array beams and the achievable sensitivity. The infrastructure uses analog and digital beamforming techniques to form a total of 219 tied-array beams and three incoherent beams per pointing (Stappers et al., 2011; van Haarlem et al., 2013; Sanidas et al., 2019). The beamforming was also transferred recently to a GPU-based infrastructure providing a step up in data handling and processing speeds (Broekema et al., 2018). This has been complemented by automated pulsar candidate classification schemes (Lyon et al., 2016; Tan et al., 2018a) that have helped combat the processing overhead.

Beamforming technology has also played a significant role in improving the capabilities of old interferometers through a total revamping of their observing infrastructure (front-end and back-end). For example, the Molonglo cross telescope built in the 1960s was revamped with a set of digital receivers and GPU computing units which provided 10 times the bandwidth and twice the field of view compared to the previous configuration (Bailes et al., 2017). This refurbished telescope (known by the acronym UTMOST) is also capable of producing 352 "fan-beams" within the primary beam and uses a real-time single pulse and standard pulsar searching pipeline (Venkatraman Krishnan et al., 2020b). Similarly, the Westerbork Synthesis Radio telescope underwent a renovation with the installation of focal plane arrays at the feed of the telescope leading to the Aperture Tile in Focus setup (APERTIF) (van Cappellen & Bakker, 2010). In this case, the feed consists of closely spaced individual elements that sample the focal plane of the receiver and is capable of producing up to 37 beams, increasing the field of view to 8 deg². While this setup has primarily been used for FRB searching, the design is easily transferable for pulsar searches.

2.5.0.3 Multi-beam coincidenting

Coherent addition of signals from a large number of dishes also implies that the power due to RFI is lowered. This is owing to a higher directionality in the coherent beams which yields a beam pattern that is less susceptible to multiple possible terrestrial RFI signals. One of the later stages in pulsar searching is removing false positive signals that are actually RFI but show up as pulsar signals. These can easily be discarded by implementing multi-beam spatial RFI filtering techniques (as discussed in section 2.2.2). An example of such a multi-beam coincidenting scheme is discussed in Chapter 3.

Pulsar Searching with MeerKAT

This chapter introduces the MeerKAT radio telescope and a large survey project focused on targeted searches for pulsars. As part of the computing infrastructure for conducting pulsar searches with MeerKAT, I co-developed the schema and tools for database management and integrated different processing pipelines (see Section 3.4) into the infrastructure to automate the search procedure. I also conducted an acceleration search on the globular cluster Terzan 5 (see Section 3.5).

Contents

3.1	Introduction to MeerKAT	88
3.2	System description	89
3.2.1	Correlator/Beamformer (CBF)	89
3.2.2	FBFUSE	91
3.2.3	APSUSE	92
3.3	Introduction to TRAPUM	94
3.4	An autonomous high-speed pulsar search pipeline	96
3.4.1	Background	96
3.4.2	Computing infrastructure	98
3.4.3	Processing pipelines	103
3.5	A case study with Terzan 5	108
3.5.1	Introduction to Terzan 5	108
3.5.2	Overview of radio pulsar searches with Ter5	110
3.5.3	Benefits of TRAPUM for Ter5 searches	111
3.5.4	Observations and setup	112
3.5.5	Sensitivity	113
3.5.6	Search strategy and processing	115
3.5.7	Results	120
3.5.8	Discussion	122
3.5.9	Conclusions	123

3.1 Introduction to MeerKAT

The MeerKAT radio telescope is currently the most sensitive radio telescope in the Southern Hemisphere. Located in the Karoo desert of the Northern Cape in South Africa, it consists of 64 dishes with each dish spanning 13.5 m in diameter (see Figure 3.1). Forty-two of these dishes are located within a 1 km diameter forming the inner core of the interferometer. The remaining dishes are spaced much further apart with the longest baseline being 8 km. Each dish has an offset-Gregorian design to prevent aperture blockage (see Table 3.1 for a list of specifications). Building on the current infrastructure, the MeerKAT Extension Project (MK+)¹ has been initiated recently, aiming to expand the total number of dishes to 84 with the longest baseline of 17 km. This would further enhance the sensitivity and imaging resolution of the interferometer.

The MeerKAT telescope processing chain consists of a mixture of commercial off-the-shelf (COTS) and state-of-the-art instrumentation offering high-fidelity data recording across a wide range of frequencies. The antennas are currently equipped with dual (linearly) polarised L-band (856-1712 MHz) and UHF-band (544-1088 MHz) receivers which are installed at the secondary focus² of the dish. The received voltage signals are directly sampled by 64 individual digitisers (also known as the D-engines) corresponding to each antenna without the use of a mixer after passing through an RF bandpass filter for the selected frequency band. The maximum instantaneous bandwidth is 856 MHz at L-Band and 544 MHz at UHF band implying a sampling rate of 1712 MHz and 1088 MHz, respectively. An accurate time standard is maintained during digitisation using a MASER-based Time and Frequency Reference (TFR) subsystem. The digitised data streams are relayed to the Karoo Array Processor Building (KAPB) via 40 Gigabit Ethernet (40GbE) fibre links running underground. The KAPB is constructed above as well as below ground level. The data is transported to a Faraday cage below ground called the KAPB Data Rack Area (KDRA). This shields the antennas from radio frequency interference (RFI) generated from the KAPB. Within the KDRA, incoming signals from the D-engine of each dish enter the Correlator/Beamformer (CBF) apparatus. The CBF consists of F-engines, X-engines and B-engines (see Section 3.2.1) as well as other user defined equipment catering to different science cases (see Section 3.2 for details).

Apart from the above mentioned instrumentation, MeerKAT is also equipped with supplementary subsystems and a power station that facilitate a smooth and continuous working of the entire setup. A Control and Monitoring (CAM) system and the Science Data Processor (SDP) run a range of processing pipelines to coordinate and scrutinise the entire workflow. A detailed explanation about the workings of the CAM and other components of the MeerKAT setup can be found in [Marais \(2015\)](#).

¹<https://www.mpifr-bonn.mpg.de/pressreleases/2020/9>

²MeerKAT will also be capable of observing at S-band (1.75-3.5 GHz) in the future. Observations with the S-Band receivers are currently in the commissioning phase. More details about the science case driven by this receiver is discussed in Chapter 4.

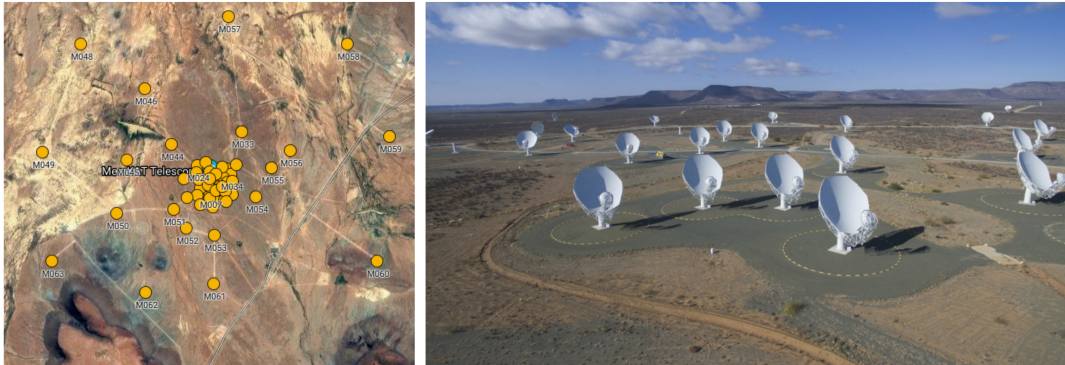


Figure 3.1: The left panel shows the Google Maps image of the MeerKAT telescope with the individual dishes highlighted. The right panel shows a collection of MeerKAT dishes in the inner core of the array (Credits: SKA South Africa).

3.2 System description

The process of scheduling an observation with MeerKAT to eventually producing a diagnostic plot (see Figure 2.7) involves a series of techniques for robust data reduction and accelerated computation. Described below are the steps and corresponding ancillary systems that reduce the data for implementation of pulsar searching pipelines.

3.2.1 Correlator/Beamformer (CBF)

The correlator/beamformer (CBF) as mentioned in Section 3.1 receives digitised streams of data from each individual antenna and is capable of producing different data products. Additionally, the CBF provides a network to enable efficient cross communication between different systems. The CBF consists of the following components:

- **F-engines:** This system behaves as a spectrometer offering frequency resolution over the observing bandwidth. It utilises a 16-tap Hann window polyphase filterbank (PFB) design to channelise the digitized bandwidth. The channeliser runs on a SKARAB³ board and is in turn split into two separate sub systems depending on the size of the output bandwidth. Science cases which require the full bandwidth of the receiver, such as time domain and continuum based studies utilise the wide-band F-engine. On the other hand, science cases which require a higher spectral resolution than offered by the wide-band F-engine (e.g. spectral line studies) use the narrow-band F-engine. The F-engine can currently output 1k, 4k and 32k number of channels for narrow-band as well as wide-band modes. The F-engine outputs are multiplexed ensuring that blocks of frequency data from all antennas are available simultaneously.
- **X-engines:** This setup is responsible for carrying out the cross-correlation of signals from each antenna. Similar to the F-engine, the X-engine is also divided

³<https://casper.ssl.berkeley.edu/wiki/SKARAB>

Table 3.1: An overview of the technical specifications of the MeerKAT telescope

Parameter	Value
Coordinates	-30°49'48.00" S 21°19'48.00" E
Number of antennas	64
Dish diameter	13.5 m
Minimum and maximum baseline	29 m and 7700 m
Operating observing frequencies	544-1088 MHz (UHF) 856-1712 MHz (L-Band)
Elevation and azimuth slew speed	1°/s and 2°/s
Maximum Gain	2.8 K/Jy
System temperature	18 K (L-Band) 23 K (UHF)

into X-engine narrow and X-engine wide in order to handle the data streams from F-engine narrow and F-engine wide respectively. Though the the setup currently operates on SKARAB, tests are ongoing to use a GPU-based replacement.

- **B-engines:** This is the beamformer that is used to generate a phased array beam after combining the channelised complex voltages from all antennas produced by the F-engines. There are 64 individual custom SKARAB boards that process channels based on a certain frequency subband of the entire bandwidth. The B-engines can generate up to four tied-array beams from up to four corresponding sub-array configurations amongst the 64 dishes.

Currently, the CBF network supports the following user supplied equipment (USE):

- BLUSE backend which is utilised for the Breakthrough Listen project (<https://breakthroughinitiatives.org/initiative/>). This setup directly utilises the channelised voltage data from the F-engine narrow as an input to dedicated GPU equipped hardware for beamforming and run processing pipelines to search for signals of interest (Gajjar et al., 2019).
- PTUSE which refers to the Pulsar Timing backend as part of the MeerTIME project (<http://www.meertime.org>). This setup ingests data from the B-engines into a ring-buffer from which the data is coherently dedispersed and eventually written out as fold-mode as well as search-mode archive files (PSRFITS format) using the known pulsar parameters. A detailed description of the entire setup along with early results can be found in Bailes et al. (2020)
- FBFUSE which is the multibeam beamformer utilised for the TRAPUM project (see Section 3.3). More details on this setup is discussed in Section 3.2.2.
- TUSE referring to the transient search cluster deployed as part of MeerTRAP (<https://www.meertrap.org/>). The MeerTRAP setup ingests beamformed

products from FBFUSE and runs searches in real time for fast transients. In case of an interesting detection, a transient buffer setup within FBFUSE is triggered to record upto 30 seconds of voltage data. The data can be correlated and imaged to help localise the source of the transient burst (Sanidas et al., 2017).

- APSUSE which is the dedicated pulsar searching cluster and serves as the processing hub for the TRAPUM project. The various specifications and functionalities of the cluster are described in Section 3.2.3.

The communication within the CBF is enabled by a Folded-Clos (also known as fat-tree) network. This network is multicast i.e. data published by one instrument can be subscribed by any other instrument within the apparatus. Additionally, the network also offers scalability which allows integration of new system components if needed in the future. More details about the network can be found in Slabber et al. (2018).

3.2.2 FBFUSE

The FBFUSE cluster is a high-performance multi-beam beamformer configured to ingest the full data rate output (upto 1.8 Tb/s for the entire cluster) from the F-engine wide machines and perform multiple beamforming operations in real-time. A list of specifications for the FBFUSE system is described in Table 3.2.

The F-engine output from the CBF network is distributed using the SPEAD⁴ protocol and is stored in a PSRDADA⁵ ring buffer. A software stack called MKRECV⁶ does the ring-buffer capture using the SPEAD2 library. This data is then reconstructed/transformed in memory. The number of tied-array beams to be synthesised can be set by the user but is limited to a multiple of 32. Based on results from extensive testing, the beamformer is capable of forming upto 960 beams for 856 MHz bandwidth and usage of all 64 antennas at 85 percent realtime. However the TRAPUM setup is currently capped to 480 beams to ensure stability of the entire system and production of reasonable data volumes that speeds up the computation time of consequent search pipelines. Once the number of beams are fixed, the tiling pattern for the observation pointing is calculated based on hexagonal packing of elliptical shaped beams. The corresponding delays required for forming each of these coherent beams are calculated using the Mosaic⁷ package. These calculations along with the transformed data stream serve as inputs for a GPU processing pipeline⁸. The processing job is executed in 64 simultaneous instances where each instance handles 1/64th of the total bandwidth. The data corresponding to each frequency subband from all the beams are dispatched to the CBF network and are accordingly multiplexed such that all frequencies from a beam are recombined. These stitched filterbank products are then available to be consumed by APSUSE and TUSE.

Besides the beamforming setup, FBFUSE also provides data that supports numerous other off-line operations. This is plausible from the transient buffer setup that is

⁴<https://casper.ssl.berkeley.edu/wiki/SPEAD>

⁵<http://psrdada.sourceforge.net/>

⁶<https://gitlab.mpifr-bonn.mpg.de/mhein/mkrecv/tree/master>

⁷<https://gitlab.mpifr-bonn.mpg.de/wchen/Beamforming/tree/master/mosaic>

⁸https://gitlab.mpcdf.mpg.de/mpifr-bdg/psrdada_cpp



Figure 3.2: A fish-eye view of the TRAPUM user supplied equipment (USE) at the KAPB. Top panel is the FBFUSE rack. Bottom panel shows APSUSE consisting of 7 racks from the left with the rest belonging to TUSE. (Credits: E. Barr)

capable of storing upto 30 seconds of data ingested from the F-engine. Similar to the X-engines and B-engines, the transient buffer data of FBFUSE can be used to produce correlated products and beamformed products respectively. This ensures compatibility with the requirements of other science cases as discussed earlier. Additionally, the buffer data can also enable coherent dedispersion and polarisation based analysis, thus making it ideal for following up on transient search triggers.

3.2.3 APSUSE

APSUSE is a dedicated high performance cluster that is equipped with hardware specifications (see Table 3.2) that offer a large storage space and a platform to deploy optimised software pipelines for accelerated computing. The setup consists of 8 capture nodes that capture data products from the CBF Ethernet network using the `SPEAD2` and `PSRDADA` libraries. The captured data are redistributed over an internal Infiniband network for storage. The setup also consists of 60 processing nodes with 24 CPU cores per node. Each node is additionally equipped with 2 Nvidia GTX 1080Ti GPUs, thus allowing for implementation of CPU and GPU based software pipelines. With a total storage capacity of 3.5 PB, the APSUSE cluster is capable of recording upto 20 straight hours uninterrupted (assuming maximum number of tied-array beams recorded). The cluster uses the `BeeGFS`⁹ distributed file system which enables read/write speeds (upto 50 GB/s) that match the performance required for capturing data from FBFUSE.

⁹<https://www.beegfs.io/c/>

Table 3.2: A list of hardware specifications for the FBFUSE and APSUSE cluster

Parameter	FBFUSE	APSUSE
Input data rate	1.8 Tb/s	280 Gb/s
Output data rate	280 Gb/s	-
Interface	40GbE, SPEAD protocol	40GbE, SPEAD protocol
Primary use case	Coherent and incoherent beam generation with time and frequency integrated Stokes I data products	Pulsar searching
Number of nodes	2 head nodes + 32 processing nodes	2 head nodes + 8 capture nodes + 60 processing nodes
Storage	3.5 PB (BeeGFS Filesystem)	
CPU	10-core Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz (2 per head node) 8-core Intel(R) Xeon(R) Gold 6134 CPU @ 3.20GHz (2 per processing node)	12-core Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz (2 per head node) 12-core Intel(R) Xeon(R) Gold 6136 CPU @ 3.00GHz (2 per capture node) 12-core Intel(R) Xeon(R) Silver 4116 CPU @ 2.10GHz (2 per processing node)
RAM	32 GB (head node) 384 GB (processing node)	32 GB (head node) 192 GB (capture node) 96 GB (processing node)
GPU	GeForce GTX 1080 Ti (2 per processing node)	GeForce GTX 1080 Ti (2 per processing node)

3.3 Introduction to TRAPUM

Of the 10 ongoing large survey projects with MeerKAT¹⁰, one project is focussed on searching for transients and pulsars through a targeted search strategy. The Transients and Pulsars with MeerKAT (TRAPUM) survey (Stappers & Kramer, 2018) aims to observe a variety of galactic as well as extragalactic sources in order to discover new radio pulsars and transients. The sub-categories of the targeted source types for the are classified into:

- **Unidentified gamma ray sources from the Fermi telescope:**

The aim here is to search for radio millisecond pulsars using a list of sources that do not currently yield a known object post cross-matching with the Fermi-LAT 4FGL catalogue (Abdollahi et al., 2020). This project is predicated on the success of previous such campaigns (e.g. Cognard et al., 2011; Keith et al., 2011; Camilo et al., 2015; Cromartie et al., 2016; Pleunis et al., 2017; Wang et al., 2018). The strategy with TRAPUM is two-pronged. One part is to conduct a blind search among sources selected using machine learning classifiers with input feature parameters based on the gamma-ray spectral emission properties and flux variability (e.g. Parkinson et al., 2016). The other part of the survey is targeted at candidates showing properties similar to the spider-class of pulsars. Spiders are a class of recycled eclipsing binary millisecond pulsars. The pulsar wind irradiates the companion leading to evaporation of material from the companion. This material obscures the pulsed emission at random orbital phases (e.g. Fruchter et al., 1988; Phinney et al., 1988). Depending on the size of the companion, the spiders are broadly divided into two categories depending on the mass of the companion namely black widows ($M_c \ll 0.1M_\odot$) and redbacks ($M_c \sim 0.2 - 0.4M_\odot$). More details on the properties of these objects can be found in Roberts (2013). For the TRAPUM survey, the spider candidates chosen have previously shown evidence for optical counterparts with spectroscopy indicating possible neutron star mass companions.

- **Supernova remnants (SNR), Pulsar Wind Nebulae (PWNe) and other TeV sources:**

These are sources that could potentially house young pulsars similar to the Crab pulsar. Several surveys targeting supernova remnants (SNRs) have been largely unsuccessful (e.g. Gorham et al., 1996; Roy et al., 2011; Straal & van Leeuwen, 2019). The main limitation from the observation aspect is the large field of view of these sources implying the need for multiple pointings across the remnant. The multibeam capability of MeerKAT offers a large field of view that can compensate for this. The selection of SNR sources are based on the catalogue from Green (2014) and PWNe sources/TeV sources from the H.E.S.S Galactic plane survey (H. E. S. S. Collaboration et al., 2018). Furthermore, the SNR/PWNe sources selected are such that their extent is compatible with the FoV of a TRAPUM

¹⁰<https://sites.google.com/a/ska.ac.za/public/meerkat/meerkat-large-survey-projects>

pointing and have no known pulsar association. The narrow tied-array beams ($\sim 14''$ with all antennas) of TRAPUM also help reduce the contribution from the background nebula in case of SNR/PWNe sources. Discovering young pulsars associated with these sources would not only improve the understanding of their formation history and energetics (e.g. Johnston et al., 2020) but also place constraints on the birth rate of such systems in the Galaxy (e.g. Keane & Kramer, 2008).

- **Globular Clusters:**

The overall success of previous targeted searches with GCs (as explained in Chapter 2) motivates expanded observations via TRAPUM. Some of the highlight discoveries in the past include eclipsing binary pulsars (e.g. Lyne et al., 1990; Manchester et al., 1991), the fastest spinning pulsar (J1748-2446ad; Hessels et al., 2006) and highly eccentric systems produced from exchange encounters (e.g. Prince et al., 1991; DeCesar et al., 2015; Ridolfi et al., 2019). The strategy to select GC targets for TRAPUM are:

- Clusters with known pulsars in order to make DM searches faster. A priority is given for those where the increased sensitivity offered by MeerKAT is more likely to yield large number of new pulsars (e.g. M28, Terzan 5).
- Southern clusters not where the sensitivity gain relative to the Green Bank Telescope is larger (e.g. M62).
- Core-collapsed globular clusters where the potential to find super-eccentric systems with MSPs and massive companions is high. These clusters offer the right environment where the same object is likely to undergo multiple exchange encounters (Verbunt & Freire, 2014). These clusters could also host binary systems with a recycled pulsar another recycled pulsar companion or a black-hole companion formed through exchange interactions (e.g. M30, NGC 6522, NGC 6624).

Besides the various applications of binary pulsars (as discussed in Chapter 1), new pulsar discoveries from GCs would help better understand their properties and environment (Pooley et al., 2003; Prager et al., 2017; Abbate et al., 2018) and can also probe the Galactic magnetic field as shown by Abbate et al. (2020).

- **Nearby galaxies namely the Small and Large Magellanic Clouds (SMC and LMC).**

Despite pulsar discoveries already made in the LMC (23; Ridley et al. (2013) and references therein) and SMC (7; Titus et al. (2019) and references therein), the improved sensitivity and resolution of MeerKAT boosts the possibility of finding more pulsars in these dwarf galaxies. The multibeam capability of TRAPUM allows a resolution that is suitable to probe specific parts of these galaxies. New discoveries can help determine the metallicity and rate of star formation (Antoniou & Zezas, 2016) in these galaxies.

Each of these subcategories contain multiple sources to be observed. In case of a discovery of a promising pulsar, observations of the particular source of interest would also be followed up using other telescopes where possible. This includes telescopes like Parkes, Lovell and Effelsberg. Discoveries in radio of unassociated Fermi sources would also be followed up in gamma rays. This would in-turn help build a dataset for future timing purposes. Additionally, a memorandum of understanding is setup between Albert Einstein Institute (AEI) and TRAPUM where electromagnetic counterparts for future detections of coalescing binaries can be followed up with TRAPUM. Apart from the specific science goals depending on the source of interest, discoveries made from TRAPUM would also serve the following generic science goals:

- They could be incorporated into the International Pulsar Timing Array project (<http://ipta4gw.org/>) which is aimed at discovering nanohertz gravitational waves through pulsar timing signatures by combining the latest data from the European Pulsar Timing Array (EPTA, [Desvignes et al., 2016](#)), the North American Nanohertz Observatory for Gravitational Waves (NANOGrav, [Arzoumanian et al., 2020](#)) and the Parkes Pulsar Timing array (PPTA, [Kerr et al., 2020](#)).
- They can increase the sample size for galactic neutron star population studies including their evolution and formation rates (e.g. [Bates et al., 2014](#))
- The targeted searches also give better chances to discover new or exotic objects of which relatively few are known (e.g. magnetars, RRATs, intermittent pulsars and other sub-populations discussed in Chapter 1).

A large scale survey like TRAPUM requires a robust end-to-end pipeline that is capable of processing data at enhanced speeds. Dedicated hardware and a corresponding software pipeline is essential to facilitate the TRAPUM workflow. In the following section, we describe the infrastructure behind an autonomous search pipeline.

3.4 An autonomous high-speed pulsar search pipeline

3.4.1 Background

When a TRAPUM observation is scheduled, a set of target parameters are supplied to the MeerKAT configuration authority (CA). The CA is a gateway responsible for cross communication between user defined inputs and telescope control. The user defines specifies the number of frequency channels, sampling time and the tied-array beam configuration to the CA. Following this, a unique schedule block identifier is generated for the planned observation. The data captured by FBFUSE during an observation is later streamed and stored in filterbank format on APSUSE (as explained in Sections 3.2.2 and 3.2.3). The data on APSUSE serves as the starting point for implementing the different pulsar search techniques. Given the sampling rate for ($76 \mu s$) and fine frequency resolution (4096 channels across 856 MHz at most) for TRAPUM, hardware accelerators are a must for processing data within a reasonable time constraint. A

software code would have a significant performance enhancement using an accelerator if it spends a lot of time doing a task that can be easily divided into hundreds (or more) of parallel tasks (i.e. the application can benefit from the massive parallelism offered by an accelerator). Also if a large number of algebraic operations are needed per gigabyte of data to process, it is worth transferring the data through the PCI-express bus. Algorithms with well defined memory access patterns and highly arithmetic intensity stand to receive the greatest performance boost from massively parallel architectures, while those that involve a significant amount of decision making may struggle to take advantage of the available processing power (Barsdell et al., 2010).

3.4.1.1 Constraints with Pulsar data and need for GPU

One of the standard accelerators used are Graphical Processing Units (GPUs). They possess a highly parallel structure thus making them more efficient than central processing units (CPUs) for algorithms that need to process large blocks of data in parallel. The GPU-accelerated computing offloads compute-intensive portions of the application to the GPU, while the remainder of the code still runs on the CPU. Applications thus run much faster from the perspective of a user.

Previously written software for pulsar searching purposes predominantly depended on improvement through increased computational power delivered by higher clock rates (e.g. SIGPROC¹¹). The advent of multi-core CPUs helped in further pushing the computation speed (e.g. PRESTO, Ransom, 2011). Currently, graphical processing units with massively parallel computing architectures are the ideal platform for processing astronomical data. Barsdell et al. (2010) provide a novel approach to transition from single core to multi core based on the analysis of algorithms. They discuss the pulsar dedispersion algorithm as an example where the task of computing the magnitude of time delay per channel is redistributed among various resources to provide a significant speed-up. However, they also stress the importance of algorithm level optimisation to precede an implementation level optimisation. A recent example is the Fast Dispersion Measure Transform (FDMT; Zackay & Ofek, 2017) which reduces the complexity of the dedispersion algorithm with minimal loss in sensitivity. Investment in such algorithms would be necessary for future processing setups.

Improved pulsar survey specifications are causing an exponential rise in pulsar candidate numbers and data volumes (Lyon et al., 2016). Key algorithms like the Fast Fourier Transform (FFT) require a lot of repetitive processes as well as several algebraic operations. Using a CPU with all its cores to process this information is about an order of magnitude slower than implementing the same optimised code with all available cores with GPU (e.g. Morello et al., 2019). While the algorithms and methods applied ever since the discovery of pulsars have proven to be robust, current volumes of data generated demand that the existing algorithms at the least be adapted for GPUs to enhance processing speed. Thus, the usage of libraries like CUDA are a necessity for modern day pulsar search pipelines.

¹¹<http://sigproc.sourceforge.net/>

3.4.2 Computing infrastructure

Given the constraint on the storage for the cluster, a systematic approach is required to ensure that a lack of disk space does not halt the cadence of observations. Thus an efficient bookkeeping system also forms an integral aspect of the setup required for pulsar searching. Furthermore, a setup is needed to be able to effectively manage resources on the cluster and accordingly deploy a set number of jobs for a given processing pipeline.

The data processing on the APSUSE cluster is facilitated by **Docker**¹² containers orchestrated by the **Kubernetes**¹³ orchestration platform. Each container receives inputs for the processing via messages delivered using the **RabbitMQ**¹⁴ message broker. A **MySQL** database service runs in the background that tracks and manages the observation data as well as the products of processing pipelines. These software and their relevant role in the pulsar searching infrastructure are described in detail below.

3.4.2.1 Open source software

A simple and cost-effective solution to building a data processing framework is to develop and benchmark using open source packages. Open source technology offers advantages in terms of transparency and a constant scope for improvement with assistance from the software community. Described below are the main open source tools used for the pulsar searching infrastructure. These tools are also of vital significance in modern day industrial computing applications.

- **Docker**

Docker is a container based virtualisation platform that helps configure an isolated self-consistent entity for executing and implementing user defined software. This platform as a service (PaaS) package can be utilised via an image i.e. a template built with dependencies based on a set of instructions. These list of instructions are given as an input file (termed as **Dockerfile**) which is read by the **Docker** engine service to build an image layer-by-layer. Once the image is built, any number of instances (termed as containers) can be run. This is equivalent to launching different virtual machine (VM) instances on a local computer. This scheme removes the hassle of repetitive installing and re-installing of packages when figuring out the required dependencies for a given software package to run machine-to-machine. A primary advantage of the container approach is task isolation. This implies that any process running on a particular container cannot affect processes running in another container or the host system (unless the container shares a common volume mount from the host system).

- **Kubernetes**

Kubernetes (abbreviated as K8s) is an orchestration system that allows automated deployment and management of applications that stem from a container-

¹²<https://www.docker.com>

¹³www.kubernetes.io

¹⁴<https://www.rabbitmq.com/>

isation configuration. Thus, containers from **Docker** images are compatible with K8s. Originally designed by Google, K8s is currently maintained by the Cloud Native Computing Foundation. Some of the highlight functionalities of K8s includes:

- *Availability of a service* : K8s offers containerised applications that can run as a service i.e. the applications can play a an interim role to exchange information for other containers running specific tasks. K8s can allocate a unique virtual IP for a running container to expose it to an external connection.
- *Load Balancing*: K8s ensures that once a container that is deployed in the framework, the deployment is stable with redistribution of the network traffic if needed.
- *Optional storage mounts*: K8s is capable of mounting a user defined storage space be it local or cloud based.
- *Automated rollout, scaling and rollback*: The number of deployed containers running at a time can be controlled. Accordingly, new containers would be launched and unwanted containers can be removed.
- *Automatic bin packing*:

K8s can be provided with a cluster of nodes that can be used to run containerised tasks (termed as *Pods*). These containers can also be allocated CPU and RAM requirements based on the user's request. Once these constraints are specified and the container is deployed, K8s automatically allocates the container to a corresponding node that can efficiently utilise the resources.

- *Self-healing*:

In case a container fails, K8s automatically triggers a restart. Container specifications that do not match available resources are immediately notified. These faulty pods are either replaced or killed depending on the resource compatibility and the load on the system. They are not available for usage until the pod recovers.

- *Password protection*

K8s has options to store sensitive information like passwords and SSH keys. Containers can cross communicate with password protected services within the cluster without exposing the login credentials.

The **kubeadm** toolbox allows for installation and management of the K8s platform. There are several types of K8s resource objects i.e. self-consistent job running entities that can be launched in the framework. For example, a deployment is a type of K8s object that sets up an application to be run with a user defined parameters. Similarly a service (mentioned above as one of the functionalities of K8s) is similar to a deployment but offers a pathway to communicate with other existing pods. A command line tool (namely **kubectl**) is used for deployment and management of such resource objects. Details on configuring these resource objects is found in the next section.

- **RabbitMQ**

RabbitMQ is a message-broker application that uses the Advanced Message Queuing Protocol (AMQP)¹⁵. This package offers a medium to send and receive messages in a structured fashion. Unique channels (termed as *queues*) are available through which different applications can cross-talk in a regularised manner. Additionally, the number of unique queues can be quick to scale, thus serving as a pathway for multiple networks. This is particularly important for multi-tasking between numerous applications. A Python implementation of the RabbitMQ protocols is used (termed as *pika*¹⁶) for the purpose of transmitting and receiving messages. *pika* contains all functionalities necessary to instantiate a message brokering setup.

- **MySQL**

MySQL is relational database management system (RDBMS) service. RDBMS helps structure and organise data in the form of tables with corresponding data types. The term ‘relational’ refers to functions that offer linking between tables based on certain data types that can serve as a mode of extracting data based on certain conditions. SQL (Structured Query Language) is a standard language used by programmers to communicate with an RDBMS in order to create, modify and retrieve data entries from an RDBMS. MySQL has added capabilities of handling multiple users and implement a database server on a host system of interest. This database can also be assigned external IP based access and allows for mirroring of the data in the database to different locations. The Python based *SQLAlchemy*¹⁷ package has been utilised to develop the necessary structure of the database as well as run complex queries to run specific tasks.

- **MongoDB**

Unlike MySQL, MongoDB¹⁸ is a non-relational database which stores data in the form JSON documents. These documents have self-defined guidelines internally but do not have to adhere to rigid structure specifications when input into the database. This scheme enhances scalability and diversifies the type of data that can be handled. Software development is also simplified since MongoDB documents are directly compatible with object-oriented programming languages. A MongoDB server runs as a K8s service and is used to update outputs from the acceleration search pipeline (see Section 3.4.3). This is done using *Pymongo*¹⁹ which is a Python toolkit for working with MongoDB.

¹⁵<https://www.amqp.org/about/what>

¹⁶<https://pika.readthedocs.io/en/stable/>

¹⁷<https://www.sqlalchemy.org/>

¹⁸<https://www.mongodb.com/>

¹⁹<https://pymongo.readthedocs.io/en/stable/>

3.4.2.2 Pipeline Workflow

The tools mentioned in the previous section serve as integral parts of the entire processing chain. The entire workflow of this chain is summarised in Figure 3.3.

Currently, 57 of the 60 nodes from the APSUSE cluster have been integrated for processing via K8s. The other 3 nodes serve as development nodes to test pipelines and run specific tasks. These 57 nodes behave together as a cloud service where users can supply their requirements to run the respective search pipelines. There are two main K8s services that run in the background during the processing. Firstly, a MySQL database server is launched as a K8s service (hereafter TRAPUM DB; see Figure 3.4). The database contains all necessary information ranging from the observation parameters like target specifications and pointings to input parameters for pipelines and corresponding data products that ingested and generated from each executed processing. Following every observation, a database loading script²⁰ is launched that creates a new entry in the TRAPUM DB. Furthermore, a separate CRONJOB runs in the background that creates a backup of the latest TRAPUM DB state on a daily basis. The second crucial service run with K8s is a RabbitMQ server through which necessary input parameters for a pipeline can be published to a unique queue. The corresponding pipeline deployment can consume these messages and launch the required tasks.

To ensure minimal effort from the perspective of the user, an internal webpage has been built on the foundation of the TRAPUM DB and maintained on the head node of the FBFUSE cluster (see Figure 3.5). This webpage grants access to all TRAPUM users to make a processing request. The user can choose the files from a particular observation epoch to be processed with a pipeline of interest along with defining specific input parameters to be utilised. Once a user sends in a processing request, a series of steps follow:

²⁰<https://github.com/MPIfR-BDG/trapum-db>

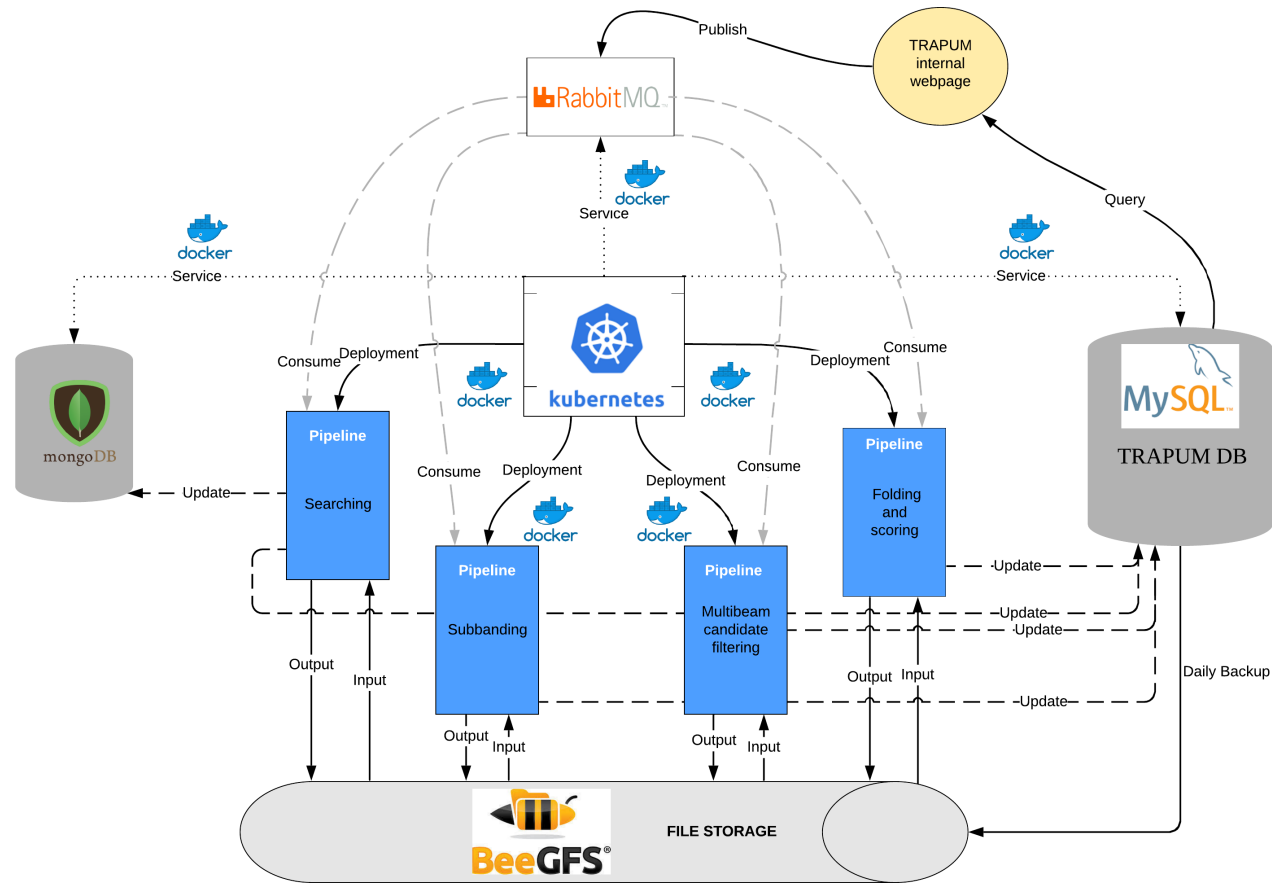


Figure 3.3: The summarised workflow of the Kubernetes based pipeline implemented on the APSUSE Cluster. The TRAPUM internal webpage serves as the user input for a cloud type service managed by Kubernetes. The Rabbit MQ server is populated with messages generated from querying the database and applying user inputs. The messages are consumed by specific pipeline deployments that run using Docker containers that are scalable. The data products generated from the pipelines are written to the BeeGFS file storage system and the corresponding information is relayed and updated to the database.

a) A new entry is instantiated in the Processings table of the TRAPUM DB which gets tagged with the necessary pipeline identifier.

b) A JSON formatted message is generated containing relevant information about user based parameters and is published to a particular RabbitMQ queue unique to the pipeline that is launched.

c) A pipeline deployment consumes this message on the other side and updates the TRAPUM DB at the start and end times of the execution. Once the task is complete, the TRAPUM DB is also updated with necessary data product entries that are created as a result of the pipeline execution.

The webpage is automatically updated with the latest state of the processing that was launched by the user. Once a processing is done, relevant details like the path to the data products are displayed for users to follow up on. In order to configure the pipeline requirements for the pipeline deployment, a YAML²¹ formatted file can be used as a configuration file for K8s. This YAML file contains information regarding the name of the `Docker` image required to run the pipeline as well as the respective script to be executed once a container is launched. The YAML file also allows specifications for CPU and ephemeral storage limits to be allocated for the deployment. Additionally, specific paths on the APSUSE filesystem can also be provided to the YAML file for reading and writing the required files. The overall infrastructure ensures flexibility in integrating any new pipeline into the system.

3.4.3 Processing pipelines

The multiple science working groups within the TRAPUM consortium implies a high demand for scheduling observations. Currently, TRAPUM observations can last upto four hours and configured with 288 tied array beams, 76 μ s sampling and 4096 frequency channels. However these numbers change depending on the targeted sources. The numbers are decided based on a balance between obtaining high fidelity data and the limiting writing speed to the file system. As a result, approximately 50 TB/hr of data is recorded on APSUSE. Thus, a primary limiting factor for restricting the observing cadence is the storage space available on the cluster. This implies that multiple resource efficient pipelines are necessary for data reduction and analysis. The results are crucial to loop back and delete unnecessary files and vacate storage space for subsequent observations. Described below are the core pipelines utilised for pulsar searching and data reduction via the K8s infrastructure:

3.4.3.1 Subbanding pipeline

The subbanding pipeline has been incorporated for reduction of data from an observation of a target with a well constrained prior on the DM (e.g. globular clusters with known pulsars). The pipeline performs the following routines:

²¹<https://yaml.org/>

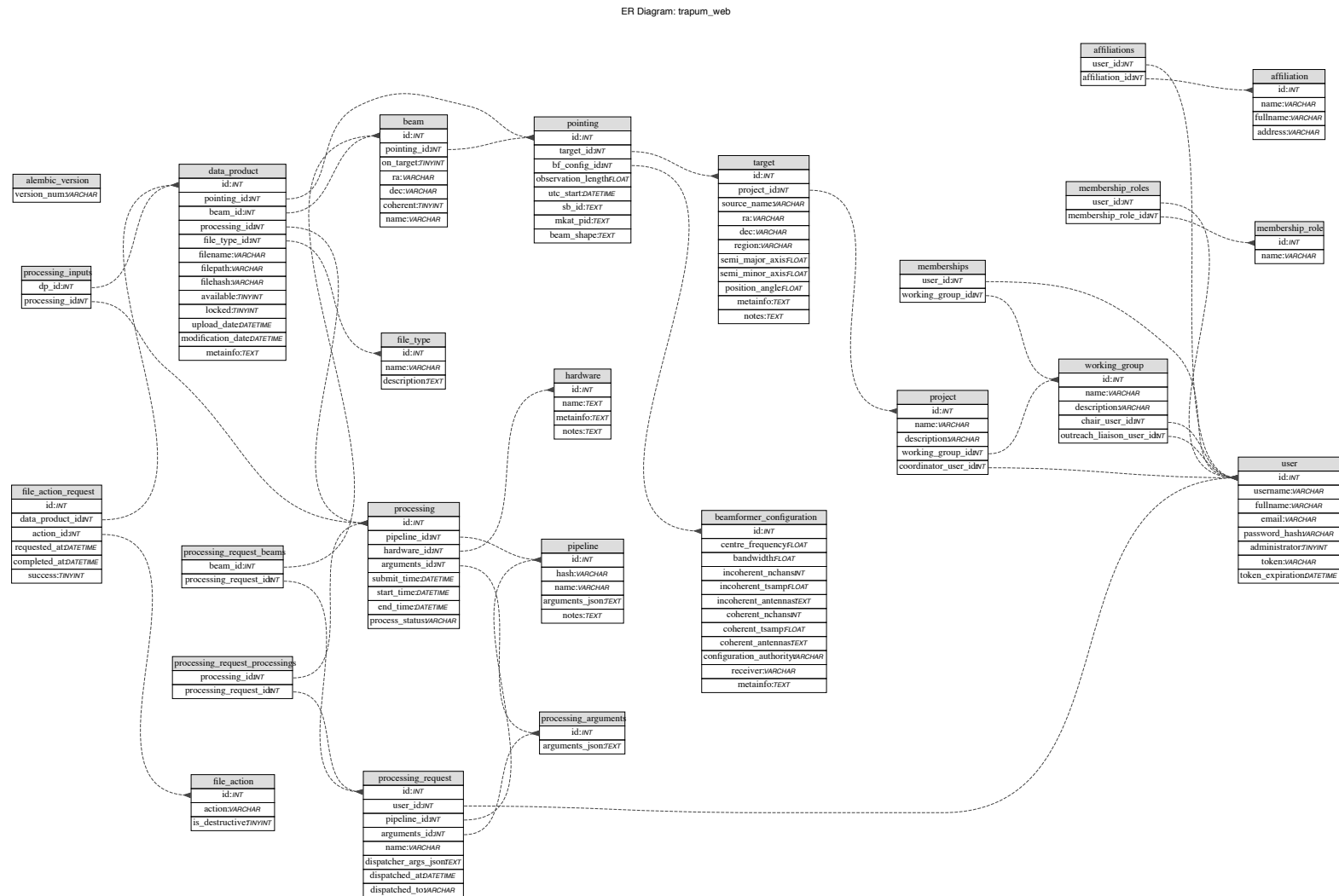


Figure 3.4: An entity relationship diagram of the TRAPUM database. The connections between tables indicate key constraints that assist in making complex queries. All the necessary metadata beginning from scheduling to the data products produced from specific pipelines are tabulated here. Figure provided by Vivek Venkatraman Krishnan.

a) Each individual filterbank file per beam that is generated spans 5 min in observing length thus generating 48 such files (assuming a 4 hour observation) per synthesised beam in one epoch²². The files are combined into one large filterbank with the `digifil` routine without compromising on the time and frequency resolution.

b) The combined filterbank file undergoes an RFI cleaning procedure using the Inter-Quartile Range mitigation (IQRM) procedure²³. In this routine, a narrow time window (60 seconds) is chosen to calculate the average bandpass of the observation file (i.e. average power across the time window per frequency channel). Using the bandpass, frequency channels are flagged when the difference between the original bandpass and multiple lagged/shifted versions of the bandpass exceed a certain threshold based on a factor of the Inter Quartile Range (IQR). The flagged channels are replaced by the respective medians of the channels given that the median is robust to outliers in the data. The IQRM procedure generates a new cleaned version of the original filterbank file.

c) The cleaned version of the combined filterbank file is subbanded (see Section 2.2.3.2) using the `ft_scrunch_threads` routine of the `filtools`²⁴ package. The number of frequency channels are scrunched by a factor of 16 (i.e. 4096 channels are reduced to 256 channels). This number was arrived as a compromise between reduction in the final output data size and minimal loss in sensitivity due to dispersive smearing. GCs with a high DM differential between known pulsars like Terzan 5 (see Section 3.5) can show dominant dispersive smearing effects. Thus the frequency scrunch factor does exceed a given threshold.

The final output of subbanding pipeline procedure is a cleaned and subbanded version of the original data with one filterbank file per beam. The intermediate data products created are deleted on-the-fly thus preserving storage space. A sanity check is run on the subbanded versions of the original files (e.g. checking known pulsar detections) before deleting the original full-frequency-resolution files. The subbanded files are then transferred to the storage server in Bonn for offline processing.

3.4.3.2 Acceleration search

The default search pipeline for TRAPUM is GPU based and utilises the `PEASOUP`²⁵ pulsar search library for time domain acceleration searching. The dedispersion prior to acceleration searching is performed using `DEDISP`²⁶. `PEASOUP` has been successful in finding new pulsars with previous and ongoing surveys (e.g. [Keane et al., 2018](#); [Morello et al., 2019](#)). However, `PEASOUP` has previously shown issues with poor detection sensitivity for 8-bit data ([Berezina, 2020](#)). This problem was further investigated and it was found that the dynamic range of the dedispersed timeseries was massively compressed

²²This is done since having too many short sized data products would cause a surge in the number of TRAPUM DB entries. On the other hand, large sized files (hundreds of gigabytes) are inconvenient I/O wise for data reduction.

²³https://gitlab.com/kmrajwade/iqrm_apollo

²⁴<https://bitbucket.org/mkeith/filtools>

²⁵<https://github.com/ewanbarr/peasoup>

²⁶<https://github.com/ewanbarr/dedisp>

TRAPUM Internal Profile Discoveries ▾ Processing ▾ Observations ▾ Organisation ▾ Logout

Observation: J1748-2446N

Pointing

Key	Value
Target	J1748-2446N
RA (hh:mm:ss)	17:48:04.92
Dec (dd:mm:ss)	-24:46:53.8
Duration (s)	14008.8
UTC start	Sep 5, 2020, 4:06:39 PM
Schedule Block ID	20200619-0057
MeerKAT Project ID	SCI-20180923-MK-04
Beam Shape	{ "y": 0.003310070170234607, "x": 0.004498623014720612, "angle": 133.76031334310878 }

Beamformer Configuration

Key	Value
Frequency (Hz)	1284000000.0
Bandwidth (Hz)	856000000.0
IB Nchannels	4096
IB Time Resolution (s)	7.65607e-05
IB Antennas	
CB Nchannels	4096
CB Time Resolution (s)	7.65607e-05
CB Antennas	
Configuration Authority	

Recorded Beams

Beam count: 253
[View beams](#)

Data Products

Primary product count: 12143
 Secondary product count: 253
[View data products](#)

Processings

Number of processings with inputs from this observation: 253
[View processings](#)

Figure 3.5: A sample user interface to the TRAPUM internal webpage of a particular target observed. Observing parameters like the integration time, source coordinates and the MeerKAT Project ID are specified. The MeerKAT Project ID is a unique identifier that helps distinguish different sub-projects within TRAPUM classified based on types of targets (e.g. globular clusters, Fermi sources). A visual representation of the beam pattern along with nearby pulsars is also shown. Finally, links to specific details on the original recorded files (primary products) as well as the data products generated from various processings (secondary products) are also provided.

due to inappropriate type casting and application of flawed scaling factors in DEDISP. The issue was resolved by modifying the library to read a floating point based (32-bit) timeseries generate before running the acceleration search, thus allowing for a wider dynamic range than before²⁷. The pipeline uses a default channel mask and birdies list. This been generated based on sources of RFI that are known to corrupt certain bands of the total observing bandwidth. Additionally, pipeline testing has also revealed previously unknown RFI sources that have accordingly been accounted for in the mask and birdie list. The range and step size for the DM trials are upto the user's discretion to provide as an input. The norm with the current processing is to use the `DDplan.py` package that generates an optimised dedispersion scheme based on ideas discussed in Chapter 2. The acceleration range is user-defined with a default tolerance of 1.1 to determine the step size.

Similar to the subbanding pipeline, contiguous files are combined and cleaned using `IQRM` before launching `PEASOUP`. The default search length is set to 10 min to prevent overheads due to I/O limitations as well as keeping a cap on the total RAM usage. `PEASOUP` automatically loads the entire filterbank into RAM before processing thus limiting the size of the input file that can be processed. The output from the `PEASOUP` pipeline is an XML format file that contains a list of search parameters, header parameters of the input file and a list of candidate metadata including the topocentric spin period, acceleration, DM and the detected S/N. These parameters are updated to a separate Mongo DB server.

3.4.3.3 Multibeam candidate filtering

Given that each synthesised beam generates a set of XML based candidates from `PEASOUP`, a pre-folding sifting mechanism is applied for reduction of unnecessary candidates for further processing. A two stage filtering process is applied using the `candidate_filter`²⁸ tool. Firstly, the tool scouts and removes candidates with periodicities that match with the known interference birdie list²⁹. Following this, a spaital domain multibeam coincidenting algorithm is applied. The algorithm clusters together candidates into groups based on similar spin periods. Within each cluster, the candidate with the maximum S/N is chosen and used as a reference to calculate the pairwise angular separation for the rest of the candidates in the cluster. An exponential decay law is fit to this curve and a threshold is applied on the decay exponent value. A high rate of decay is characteristic of a potential pulsar detection in one beam and weaker detections in the immediate neighbouring beams. In contrast, a low rate of decay increases the probability of the cluster belonging to an RFI source. Based on these features, only those clusters of candidates are retained that fall below the chosen decay exponent value threshold factor.

²⁷For validation, extensive testing with the modified version was done using data from the TRAPUM commissioning phase. The results showed S/N improvement factors of close to an order of magnitude for some known pulsar detections.

²⁸https://github.com/larskuenkel/candidate_filter

²⁹This routine has been incorporated in the fork of the original repository and can be found in https://github.com/prajwalvp/candidate_filter

This pipeline reads in a set of XML files corresponding to different synthesised beams of the same pointing and outputs a CSV formatted file giving all the metadata for the reduced set of candidates. Additionally, CSV files with parameters of the removed candidates are generated as a way to test the effectiveness of the filter. These CSV files are finally archived into one file³⁰.

3.4.3.4 Folding and ML based ranking

The folding pipeline for TRAPUM uses the `PulsarX` package. The `psrfold_fil` routine of `PulsarX` offers multi-threaded multi-candidate folding and has been engineered for enhanced processing speed. The pipeline reads in the CSV file generated from the `candidate_filter` routine and generates a predictor file for a particular beam of the corresponding pointing in consideration. The predictor file generation is repeated for every other beam in the pointing. The pipeline outputs a diagnostic PNG plot as well as a corresponding PSRFITS based archive file for each folded candidate. Thus the files are compatible with packages like PSRCHIVE (van Straten et al., 2012) for further inspection. Additionally, a CSV file is generated containing the input candidate parameters before folding as well as the final optimised parameters (spin period, DM and acceleration). Furthermore, the archive files are sifted through the PICS classifier (Zhu et al., 2014) which is a Convolutional Neural Network (CNN) based machine learning model that was originally trained on data from the PALFA survey (Cordes et al., 2006). The model generates a score between 0 and 1 with 1 indicating a most likely pulsar detection and 0 indicating noise or RFI. The scores generated per candidate are appended to the CSV file generated from `PulsarX`. The files produced for all the beams processed are combined into a single tarball that is transferred offline for candidate viewing.

As mentioned previously, each of the pipelines mentioned above are synchronised to be in connection with the TRAPUM DB through the execution process. These pipelines are also regularly maintained on GitHub³¹. The internal webpage has been configured to choose input parameters for the above mentioned pipelines and generate corresponding processing requests.

3.5 A case study with Terzan 5

In order to demonstrate the effectiveness of the entire computing infrastructure, an in depth analysis of observations with globular cluster Terzan 5 is presented below.

3.5.1 Introduction to Terzan 5

Terzan 5 (hereafter Ter5) is a globular cluster (GC) located in the inner bulge of the Galaxy. Although this cluster suffers from excess obscuration owing to its location, it

³⁰The BeeGFS file system performs poorly in terms of read/write speeds if the system contains several short sized files (order of several kilobytes). Hence small files are combined into tarballs wherever possible.

³¹<https://github.com/MPIfR-BDG/trapum-pipeline-wrapper>

Table 3.3: List of structural properties of the Terzan 5 globular cluster

Parameter	Value	Reference
Mass	$5.66 \pm 0.71 \times 10^5 M_{\odot}$	Baumgardt & Hilker (2018)
Distance	5.9 ± 0.5 kpc	Valenti et al. (2007)
Luminosity	$8 \times 10^5 L_{\odot}$	Lanzoni et al. (2010)
Core-density	$8.3 \times 10^6 M_{\odot} \text{pc}^{-3}$	Baumgardt & Hilker (2018)
Core-radius	$9''$	Lanzoni et al. (2010)
Half-mass radius	$31''$	Lanzoni et al. (2010)
Tidal radius	$277''$	Lanzoni et al. (2010)

has provided seminal results since its discovery in 1968 (Terzan, 1968). Some of the main properties of Ter5 are summarised in Table 3.3.

Ter5 has been extensively studied at multiple wavelengths. Near infrared studies by Ferraro et al. (2009) have shown evidence for two distinct stellar populations in terms of abundance of Iron. They also showed that Ter5 is possibly a remnant of primordial building blocks of bulges. Lanzoni et al. (2010) gave improved measurements of structural parameters like the central mass density and luminosity of the cluster using data from Very Large Telescope and the Hubble Space Telescope. They also revealed a higher mass than previous studies ($\sim 2 \times 10^6 M_{\odot}$). Given that optical and infrared studies can introduce significant reddening effects, Ter5 has also been observed in X-rays. There are at least 50 known X-ray sources associated with the cluster (Heinke et al., 2006) with recent studies also showing a possibility of some such sources being transitional millisecond pulsars (Bahramian et al., 2018). Recently, Bogdanov et al. (2020) analysed X-ray archival data from the Chandra observatory and found X-ray counterparts to eight radio MSPs known within the cluster. Ter5 has also shown promising evidence of gamma ray emission in the GeV (Abdo et al., 2010) as well as TeV range (H.E.S.S. Collaboration et al., 2011). On the radio end of the spectrum, Ter5 currently holds the record for housing the highest number of confirmed radio millisecond pulsars (MSPs) in a globular cluster (39^{32}) of which 20 sources are in binaries. The high number of binaries is not surprising given that Ter5 possesses the highest stellar interaction rate among all globular clusters in the Galaxy (Verbunt & Hut, 1987). Population studies (e.g. Chennamangalam et al., 2013) suggest that close to 150 MSPs are expected to lie in this region. Among the eclipsing binaries in Ter5, one such system is the fastest spinning radio pulsar currently known (Hessels et al., 2006). Discoveries of new radio pulsars in this cluster have also provided an independent method to probe the the intrinsic characteristics of the cluster via pulsar timing (Prager et al., 2017). Overall, the above mentioned factors motivate the need for continued observations of this globular cluster.

³²See <https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html> for the full list

3.5.2 Overview of radio pulsar searches with Ter5

The first radio pulsar discovered in Ter5 was an eclipsing binary system spinning with a period of 11.56 ms with a 1.7 hour orbital period (Lyne et al., 1990). Although this MSP is in a compact binary, a standard periodicity search alone was adequate to detect the source with the Parkes telescope at L-Band owing to its high flux density (2.5 mJy at L-Band as reported by Lyne et al. (1990)). Radio continuum observations of Ter5 with the Very Large Array (VLA) revealed numerous steep-spectrum radio sources indicating a possibility of more pulsars to be found (Fruchter & Goss, 1990; Fruchter & Goss, 2000). Lyne et al. (2000) discovered a second isolated millisecond pulsar with a 8.43 ms rotational period (Ter5C). This pulsar was found with Parkes at L-Band but with a search setup where the intra-channel smearing was a factor of 4 lower than the search done by Lyne et al. (1990). Ransom (2001) discovered a third pulsar in this cluster (Ter5D) based on acceleration searches run on 8 hour integration lengths of data obtained from Parkes. This was the first Ter5 pulsar discovered using the Fourier Domain acceleration search technique (Ransom et al., 2002) implemented via the PRESTO package (Ransom, 2001, 2011).

The gamechanger with the Ter5 searches was a result of an improved pulsar search backend, a different observing frequency and a swarm of new search techniques. Ransom et al. (2005) discovered 21 new MSPs using the Robert C. Byrd Green Bank Telescope (GBT). Observations were made with an S-Band (1.65-2.25 GHz) receiver which is less prone to dispersive and scattering effects in comparison to L-Band. Additionally, the S-band was mostly interference-free unlike the L-Band. This was complemented by the Spigot backend (Kaplan et al., 2005) of the GBT that provided a fine channel resolution of 0.78125 MHz thus reducing intra-channel smearing. Furthermore, the DM search trials ranged from 230-250 pc cm⁻³ with a small step size of 0.5 pc cm⁻³ and were used to search the full integration length (6 hours) as well as segmented lengths (10min, 20 min and 60 min). The partially coherent segmented search allowed for a boost in sensitivity for compact binaries. This was evidenced with 13 of the 21 discovered pulsars being in binaries. The success of these techniques prompted high cadence observations of Ter5 with the GBT. The re-application of different techniques on subsequent data sets led to the discovery of ten other pulsars including the fastest spinning (P=1.39 ms) pulsar (Hessels et al., 2006). A dynamic power spectrum search (Chandler, 2003) led to the discovery of Ter5ae (S. Ransom, private communication). Recently, two unique techniques applied to archival data sets have led to four new discoveries. Cadelano et al. (2018) used a stacking technique where Fourier power spectra were combined for 206 hours of observation revealing three new MSPs (Ter5aj, Ter5ak, Ter5al). Andersen & Ransom (2018) incorporated a Fourier domain jerk search algorithm into PRESTO that helped discover a new MSP (Ter5am) in GBT data recorded in 2008. Recently, the multimode PTUSE backend of MeerTIME (Bailes et al., 2020) allowed search mode recording of Ter5 data which is coherently dedispersed at the nominal cluster DM. A standard PRESTO search on this data confirmed a new pulsar namely Ter5an which was a candidate from previous GBT data (Ridolfi et al., 2021).

Although extensive searches have been conducted with Terzan 5, there can be mul-

tiple reasons for missing previously unknown pulsars. The existence of systems like Ter5 A, O and ad suggest that previous searches could have missed pulsars due to the eclipsing phenomenon. The radio eclipses are due to obscuration of the radio pulsations by ionized gas extending beyond the Roche lobe of the companion star (Fruchter et al., 1988). Besides this, some binaries could have certain orbital phases where binary effects causing acceleration, jerk and higher order derivatives are not that prominent. There also exist a set of tight binary systems where integration lengths of tens of minutes still suffer from significant binary modulation. Apart from these binary effects, high DM clusters like Ter5 can show long-term variations in the flux density of pulsars (timescale of months to years), and can change by a factor of 2 in brightness owing to refractive scintillation (Wang et al., 2018).

3.5.3 Benefits of TRAPUM for Ter5 searches

The TRAPUM infrastructure offers a unique set of capabilities that provide an edge compared to previous surveys of globular clusters. The advantages for searching Ter5 with TRAPUM are highlighted below.

- **Localisation:**

The size of a TRAPUM tied-array beam is of the order 15'' at L-Band. This is an order of magnitude smaller compared to beam sizes of single dish telescopes like Parkes (14') and the GBT (8') at the same frequency band. Thus any new source found with the TRAPUM system would already have a position that is convenient for follow-up timing analysis. The covariance between the true position of the source and other parameters while fitting for the timing residuals is reduced. Besides this, already known pulsars without a phase connected solution (e.g. Ter5al) can also be localised and timed making the process of phase connection easier.

- **Highly sensitive detections of known pulsars:**

TRAPUM uses a larger number of antennas than MeerTIME which uses the inner core antennas alone. The extended baselines thus reduce the beam size of TRAPUM in comparison with MeerTIME. Apart from the large number of beams and narrow beam size, the TRAPUM multibeam system is also capable of placing specific beams at known pulsar positions during observations. Thus TRAPUM can target more pulsars and offer a higher sensitivity than MeerTIME in most cases. The data recorded on the beams placed on known pulsars would also add to the timing baselines of the pulsars and in turn improve their timing models. Furthermore, the brightness profile from multiple detections of the same pulsar in the neighbouring beams of the specific tiling would help better characterise the point spread function of the synthesised beams.

- **Follow-up from imaging based surveys:**

Recently, Urquhart et al. (2020) demonstrated a strong case for radio imaging of Ter5 where they detected several known X-ray sources and radio MSPs in the

2-8 GHz band with the Very Large Array (VLA). They presented three steep spectrum sources that showed strong evidence for being eclipsing systems and do not match any known pulsars. Thus, information from imaging can be used to bootstrap and refine the time-domain searches for improved probability of new pulsar detections. The TRAPUM setup allows for a follow-up scheme where particular tied-array beams can be overlaid on source coordinate information retrieved from the imaging analysis. These beams could then be searched for radio pulsations.

- **Probing different regions of a globular cluster:**

The TRAPUM system uses 288 synthesised beams for globular cluster observations with a user defined overlap between tied array beams. Thus, the multibeam configuration allows flexibility in terms of balancing between maximal sky coverage and uniform sensitivity between beams. Given that the half mass radius of Ter5 is approximately 2-3 times the size of an individual coherent beam, hundreds of beams allow for a coverage beyond the half mass radius with the outermost beams lying around the tidal radius of the cluster (See example in next section). Discoveries of isolated and eccentric binary MSPs would thus help lay improved constraints on the rate of exchange interactions within the core as well as the outer realms of the cluster (Verbunt & Freire, 2014).

3.5.4 Observations and setup

Ter5 was observed on 16th April 2020 as the first TRAPUM science run. The aim of this observation was primarily focussed on quality assurance of the entire infrastructure including the observing setup, beamforming procedure and the search pipelines. The observations were made at L-Band with a 856 MHz bandwidth centered at 1284 MHz and divided into 4096 frequency channels. The data was recorded with a $76 \mu\text{s}$ time resolution for a 4 hour duration. The beamformer was configured for recording 288 synthesised beams (see Figure 3.6) with a 70 percent overlap between beams i.e. the neighbouring tied-array beam centres were separated by a distance where the maximum power in the beam drops off by 30 %. Two beams were placed on the brightest pulsars known in the cluster namely Ter5A and Ter5C and the rest of the beams were tiled around a central position chosen within the core of the cluster. The tied-array beam set form a roughly circular pattern with a radius close to the tidal radius of the cluster.

The first observation of Ter5 as part of Phase 1 of the TRAPUM Deep Survey for Globular Clusters was conducted on 5th September 2020. The aim of this survey is to find millisecond pulsars in a set of globular clusters (GCs) which are either core collapsed or already contain known pulsars. Given that Ter5 fits in with the latter condition, the cluster was chosen as one of 10 other GCs. Similar to the run in April, the observations were made at L-Band with the same observing frequency range (856-1712 MHz), time ($76 \mu\text{s}$) and frequency resolution (4096 channels). A total of 56 antennas were utilised for this 4 hour long observation. In order to thoroughly examine the sensitivity of the TRAPUM system, a beam was placed on every known pulsar

in Ter5³³ (see Figure 3.7). The rest of the beams were tiled around the position of Ter5N. Furthermore, search mode data was recorded from the PTUSE backend of the MeerTIME project (Bailes et al., 2020) simultaneously during the observation. The data was recorded with Full Stokes polarisation information at a 9.57 μ s resolution and 4096 frequency channels across the same bandwidth as the TRAPUM configuration. The data recorded via PTUSE utilised 38 antennas within the inner core and the corresponding beam was centered on Ter5N. Thus, the overall setup provides scope for benchmarking the TRAPUM sensitivity through a direct comparison with MeerTIME. The two Ter5 observations would hereafter be referred to as R1 (first run on 16th April 2020) and R2 (second run on 5th September 2020).

3.5.5 Sensitivity

The radiometer equation (Dewey et al., 1985) can be utilised to estimate the theoretical sensitivity of TRAPUM. From the perspective of searching, the expected mean flux density can be written as (Morello et al., 2020):

$$S_{\text{mean}} = \frac{S/N \beta (T_{\text{sys}})}{G \varepsilon \sqrt{n_{\text{pol}} \text{BW}_{\text{eff}} t_{\text{obs}}}} \sqrt{\frac{W}{P - W}}. \quad (3.1)$$

G refers to the gain of the telescope and is chosen to be 2.45 K/Jy which corresponds to the gain for 56 dishes. This number is chosen based on a constraint that a minimum of 56 dishes need to be operational for a TRAPUM observation. The system temperature (T_{sys}) is 26 K after taking into account the receiver temperature (18 K), the sky contribution (T_{sky} is ~ 3.5 K) and the ground spillover temperature (4.5 K at 45 degree elevation). The two orthogonal polarisations are summed and thus $n_{\text{pol}} = 2$. $\beta = 1.01$ accounting for digitisation losses. The integration time t_{obs} is chosen to be 4 hours. We chose a duty cycle of 14 percent (median of duty cycles of all known pulsars in the cluster) and minimum signal-to-noise ratio of 9 to derive the theoretical sensitivity. The search efficiency factor ε was chosen as 0.7 for Fast Fourier Transform (FFT) based searching (see Morello et al. (2020) for a detailed discussion). Figure 3.8 shows a comparison of the sensitivity between the latest VEGAS backend of GBT (Prestage et al., 2015), PTUSE backend of MeerTIME and TRAPUM at L-Band. The sensitivity of MeerTIME and GBT are approximately the same. Although the sensitivity of GBT is better for the fastest millisecond pulsars (approx. 1-3 ms), the advantage of TRAPUM is the uniform sensitivity across a wide area. This area is equivalent to the beam size for the S-Band receiver at GBT (Kaplan et al., 2005) which has been the most successful setup for discovering Ter5 pulsars.

³³The tied-array beam placement on Ter5A was missed due to an error while configuring the setup. Ter5al does not have a fully phase connected phase solution yet (Cadelano et al., 2018) and was hence also neglected.

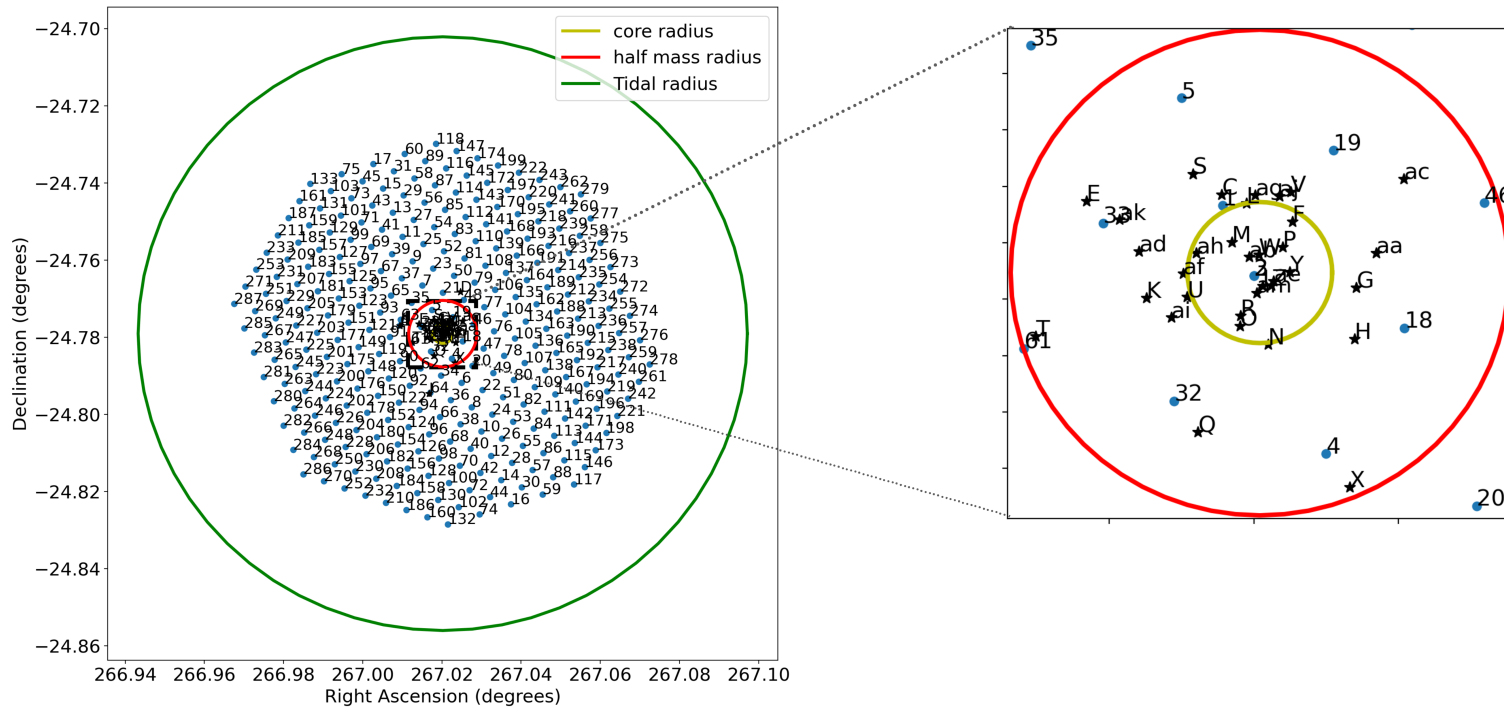


Figure 3.6: The tiling of the first Terzan 5 observation (R1 in text). Totally 288 beams were synthesised with an overlap fraction of 70% (see text), around the core position of the cluster. The extent of the tied array beam pattern is demonstrated by highlighting the core radius, half mass radius and the tidal radius of the cluster. The inset plot zooms into the region covered within the half mass radius and also shows the corresponding known pulsar positions. Majority of the pulsars are found near the core of the cluster.

3.5.5.1 Known pulsar folds

An ephemeris based folding was also done on R1 as well as R2 for all the known pulsars in Ter5. In R1, 34 known pulsars were redetected after folding on beams with the closest angular separation to the respective pulsars. The pulsars missed were Ter5ad, Te5raj and Ter5ak. Extensive quality checks on the data revealed a technical error which led to the overlaid synthesis beam pattern to change every 10 min thus leading to different positions every 10 min. As a compensation, the beams were restitched by taking sections of data every 10 min from beams that were closest in separation to a fixed set of coordinates. However, this did not prove useful in detecting the missed pulsars.

With R2, every pulsar had a corresponding beam tiled on its position. Thus, the respective beams were folded with an ephemeris revealing all pulsars except Ter5ad. This is not surprising since this system eclipses and is extremely faint (Hessels et al., 2006). Besides the TRAPUM data, R2 involved a simultaneous recording on PTUSE. In order to compare the known pulsar detections from each backend, the data were brought to a uniform sampling time (76 μ s) and frequency resolution (4096 channels). Additionally the same interference based mask was applied on both the data for folding. We compared the S/N detection of Ter5N with TRAPUM and MeerTIME (See Figure 3.9). Ter5N was the boresight source and thus any beam shape deviations were ruled out from comparative analysis. MeerTIME shows a higher S/N per antenna as compared to TRAPUM (190.19/38=5.0 for MeerTIME and 247/56=4.47 for TRAPUM). This is expected from the excess smearing that TRAPUM suffers due to incoherent dedispersion as seen from the widths of the profiles for each fold.

3.5.6 Search strategy and processing

In order to enhance processing speed without a significant loss in sensitivity, we performed acceleration searches with PEASOUP on subbanded versions of the beams. The TRAPUM subbanding pipeline was deployed with a reference DM of 237.75 pc cm⁻³ and a frequency scrunch factor of 16 (implying final output has 256 channels) for each of the 288 synthesised beams. We focus here on searches conducted on R1 alone. The searches conducted on R2 and future observing runs of Terzan 5 with TRAPUM will be discussed elsewhere. We also point here that the searches on R1 were conducted after restitching the beams due to reasons as explained in Section 3.5.5.1.

We took a two pronged approach to search the entire set of beams. The beams were categorised into two sets namely inner beams and outer beams based on those that lie within and outside the half mass radius of Terzan 5 (see Figure 3.6). This split the data into 10 inner beams and 278 outer beams to be searched. Given the high density at the Ter5 core, a partially coherent segmented search (Ng et al., 2015) was applied to the inner beams in order to improve sensitivity to wide as well as tight binary systems. Owing to RAM limitations imposed by PEASOUP, the longest integration length to be searched was restricted to 2.8 hours. The 4 hour integration was also searched in segments of 2 hours, 1 hour and 30 min. Shorter segment lengths were not considered

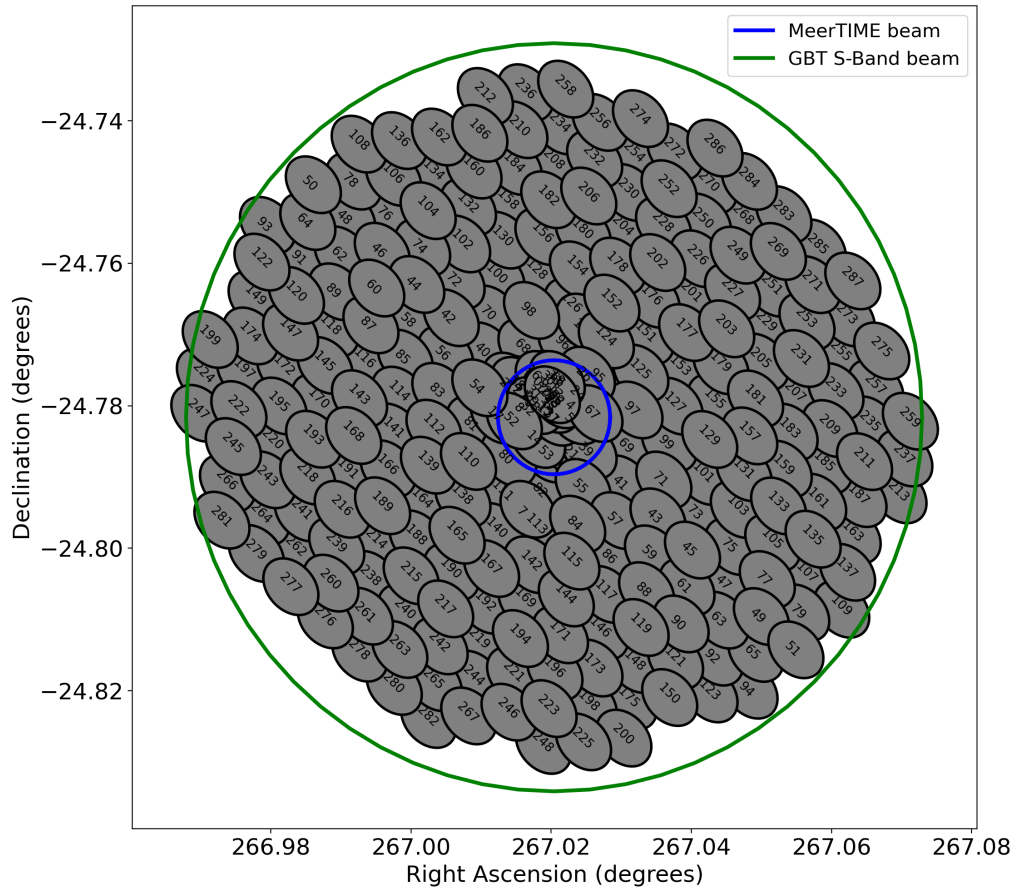


Figure 3.7: The tiling pattern for the second Terzan 5 run on 5th September 2020 with TRAPUM with 56 dishes. The MeerTIME beam is highlighted in blue which was formed with 38 dishes in the inner core. Out of the 288 beams, 36 are placed on known pulsar positions. Since most pulsars are found in the core of the cluster, multiple beams overlap nearly 100 percent at the centre of the configuration. The plot also shows that the tied-array beams approximately tile the beam size at S-Band with GBT (highlighted in green).

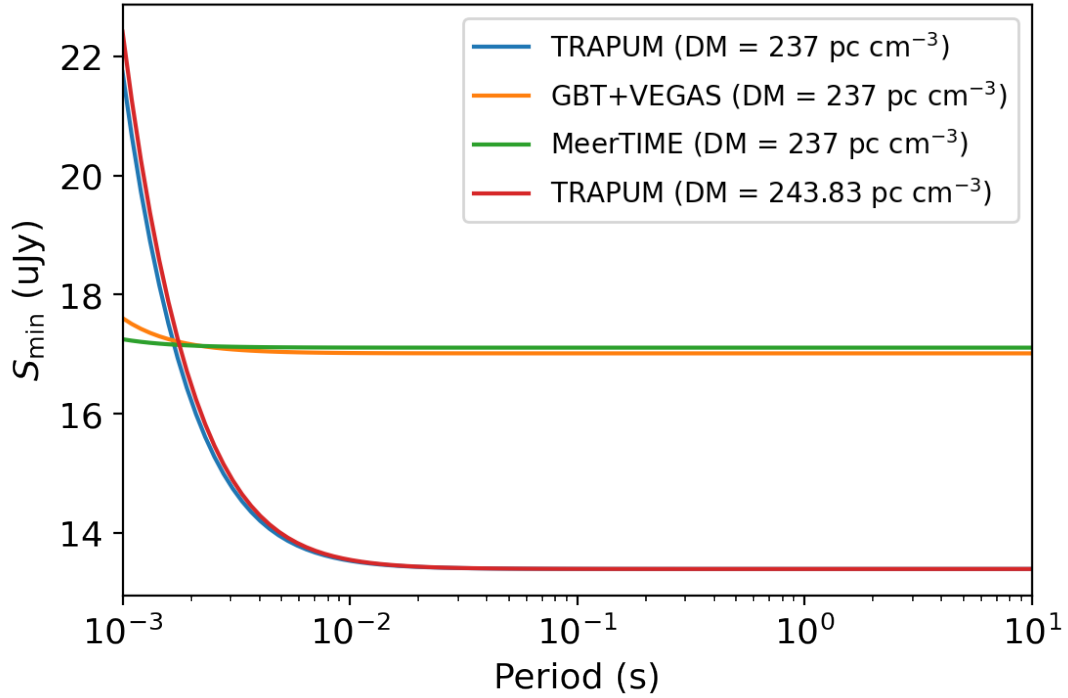


Figure 3.8: The theoretical sensitivity at the nominal DM of Ter5 (237 pc cm^{-3}) for the latest VEGAS pulsar search mode back-end at GBT and back-ends of MeerTIME as well as TRAPUM. While TRAPUM does marginally better (approx. 4 uJy) for high spin periods ($P > 10 \text{ ms}$), the GBT does relatively better for the fastest millisecond pulsars ($P < 2 \text{ ms}$). This is mainly owing to the fine time resolution achievable ($10.24 \mu\text{s}$) with VEGAS. The TRAPUM sensitivity curve is also plotted for $\text{DM}=243.83 \text{ pc cm}^{-3}$ corresponding to Ter5D. This is the maximum DM differential value with respect to the reference DM of 237 pc cm^{-3} used for subbanding. The excess smearing is $\sim 89 \mu\text{s}$ at 1284 MHz . The plot demonstrates minimal loss of sensitivity due to the subbanding process.

Table 3.4: A summary of the search parameters applied on the beams recorded with R1. The DM step size used was 0.1 pc cm^{-3} throughout. The equivalent acceleration ranges per segment of the inner beams are the maximum acceleration values calculated for circular orbits with pulsar mass of $1.4 M_{\odot}$ and a maximum companion mass of $4 M_{\odot}$.

Type of beam	Number of beams	Segment length (hr)	DM range (pc cm^{-3})	Acceleration range (ms^{-2})
Inner beams	10	2.8	225-250	-10 to 10
		2	225-250	-25 to 25
		1	225-250	-64 to 64
		0.5	225-250	-162 to 162
Outer beams	278	2.8	100-300	0 (Periodicity)

for R1 owing to the issue of restitching together beams which was known to drop the sensitivity of the already confirmed pulsars in the cluster. The DM and acceleration trial parameters chosen per segment are summarised in Table 3.4. The DM range was chosen to be $225\text{-}250 \text{ pc cm}^{-3}$ about the cluster DM. An optimal step size of 0.1 was chosen based on the smearing constraints calculated via the `DDplan.py` routine. The acceleration range was chosen assuming a maximum companion mass of $4 M_{\odot}$ in a circular binary (see Section 2.2.11) and an acc_{tol} value of 1.1 for determining the acceleration step size. Candidates generated from PEASOUP were folded via the `prepfold` routine of PRESTO (Ransom, 2011).

The strategy for the outer beams was mainly capped by the processing time constraint. We chose a DM range of $100\text{-}300 \text{ pc cm}^{-3}$ to expand the search for pulsars that are not bound by the cluster. Additionally, the search was restricted to periodicity alone and the candidates generated by PEASOUP were sifted through the `candidate_filter` pipeline. The pipeline removes known interference sources as well as clusters of candidates showing interference-like properties in the spatial domain. The remaining candidates were folded with `prepfold`. After folding, the candidates were scored through the PICS Machine Learning based classifier (Zhu et al., 2014). We took a conservative approach and inspected candidates scored above 0.1 with the classifier. This was done to avoid possible biases and discrepancies arising due to the training set being from a different survey.

Apart from the above mentioned searches, a refined search was conducted for Ter5al. This isolated MSP was first found by Cadelano et al. (2018) and was reported to be a weak source (8 uJy at L-Band). This was reflected in their analysis where they could not detect the pulsar a sufficient number of times in archival data to obtain a fully phase connected timing solution. Utilising the multibeam configuration of TRAPUM, we applied a search scheme fine-tuned to the already known parameters of the pulsar. Each of the 288 beams were dedispersed to a DM of 236.50 (as reported by Cadelano et al. (2018)). Each of the dedispersed timeseries generated were put through a periodicity search routine using `accelsearch` from PRESTO with `zmax=0`.

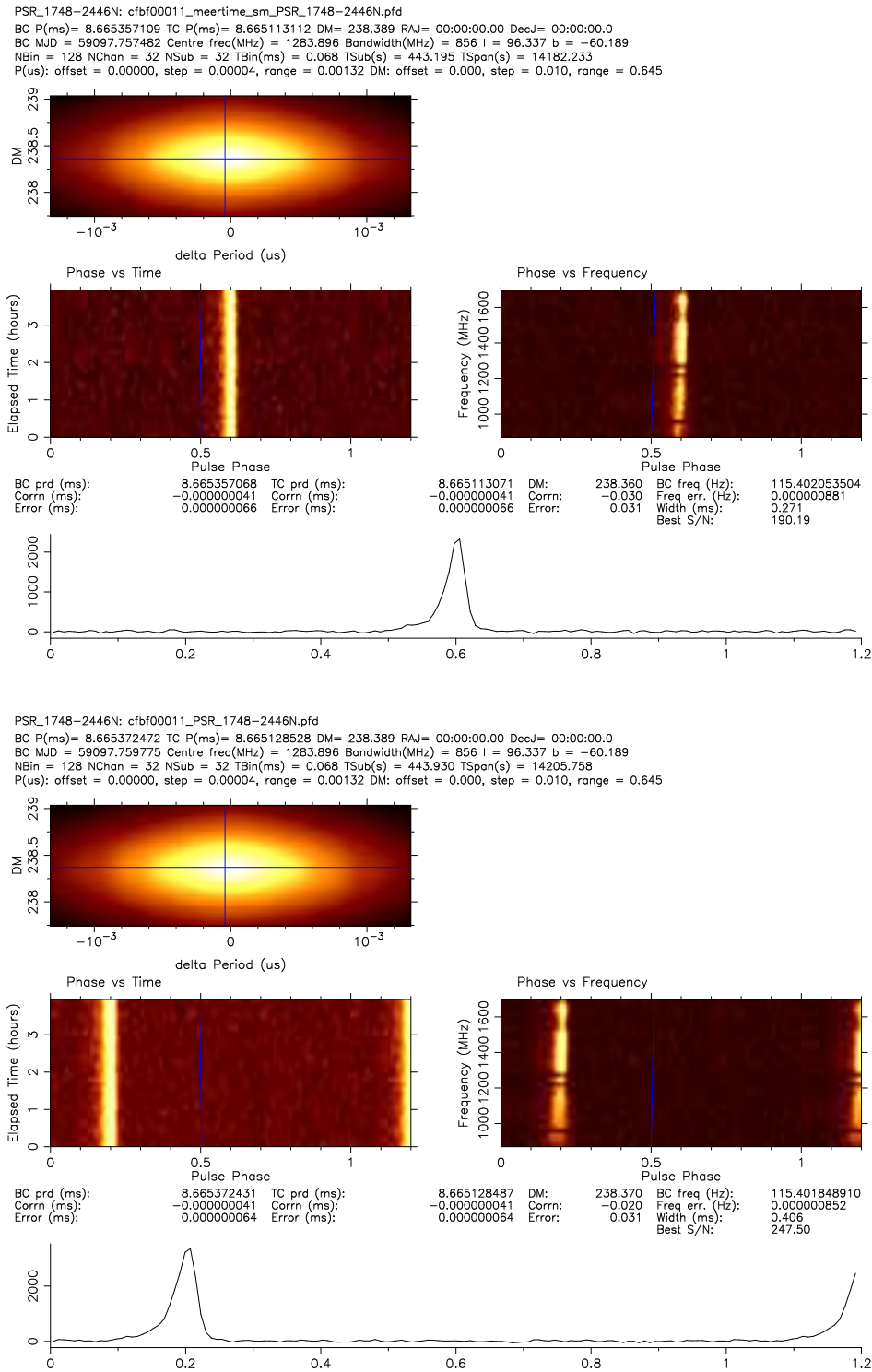


Figure 3.9: A comparison of sensitivity of MeerTIME and TRAPUM systems via the detection of Terzan 5N. Top panel: Detection plot with MeerTIME search mode data. Bottom panel: Detection plot with TRAPUM coherent beam data. The position of this pulsar was chosen as the boresight for recording on the respective backends. The sampling time, frequency resolution and the interference mask were made equal before folding with the ephemeris of Terzan 5N. The smearing at the bottom of the band is apparent for the TRAPUM system but not MeerTIME. This is due to coherent dedispersion applied per channel through the MeerTIME system with the cluster DM value as the reference.

3.5.7 Results

The searches conducted on data from R1 did not reveal a new significant pulsar detection. A few weak Tier-2 candidates were followed up by refolding the candidate parameters with neighbouring beams and extended segments but could not be confirmed. These candidates would be used as a database to cross-match candidates from future searches as well as used as a training set for exploring different machine learning models (see Chapter 5). The known pulsar detections were used as a test bed in order to analyse the efficacy of the searches conducted as described below.

3.5.7.1 Known pulsar redetections

Based on the searches conducted with R1, several known pulsars in the cluster were redetected in the inner beam and outer beam searches. The inner beam searches in particular revealed multiple detections of the same pulsar in different segment lengths. Out of the 38 known pulsars in the cluster, 19 pulsars were re-detected. Table 3.5 summarises the properties of the pulsars and the signal-to-noise ratio (S/N) obtained from the searches. Given that all beams required a restitching mechanism, the expected S/N per pulsar was not estimated. Alternatively, in order to test the efficiency of the pipeline, the S/N obtained from the searches was compared against the S/N obtained from an ephemeris based folding for each re-detected pulsar. This is represented as the Pipeline Efficiency Ratio (PER). Some pulsars with a poor PER score (below 0.8) are in compact binary orbits (Ter5O, Ter5M, Ter5W). For example Ter5M is in a 10 hr binary orbit and despite using the partially coherent segmented search, the acceleration effect alone is not enough to account for the binary modulation (see Figure 3.10). This leads to the smearing of the pulse across the observation. Some pulsars show a poor PER score as a consequence of the dynamic retiling error (e.g. Ter5Q, Ter5D, Ter5X). The error was such that beams further away from boresight experience the maximum change in sky position with every new tiling pattern regenerated. This leads to the pulsar signal coming in and out of the beam. Ter5D in particular also has a high value of ΔDM (6.83 pc cm^{-3}) i.e. the difference between the DM of the pulsar and the reference DM used when subbanding. A highlight of the searches was the redetection of Ter5an which was discovered recently with MeerTIME search mode data (Ridolfi et al., 2021). This pulsar is in a wide binary orbit with an orbital period of 9.6 days. Hence the detection was at nearly zero acceleration (see Table 3.5) Overall, the search pipeline shows a reasonable sensitivity for the pulsar redetections despite the limitations in the observing length and the usage of restitched beams.

3.5.7.2 Non-detections

Based on the remaining set of 19 pulsars that were undetected from the searches, 3 were undetected despite folding with the original full resolution recorded data as well as the restitched subbanded data (Ter5ad, Ter5aj, Ter5ak). The ephemeris folds on Ter5G, Ter5J, Ter5X, Ter5V, Ter5Z, Ter5aa, Ter5ac, Ter5af, Ter5ag, Ter5ah and Ter5ai revealed a weak (lower than 9-sigma) detection on the respective closest spaced

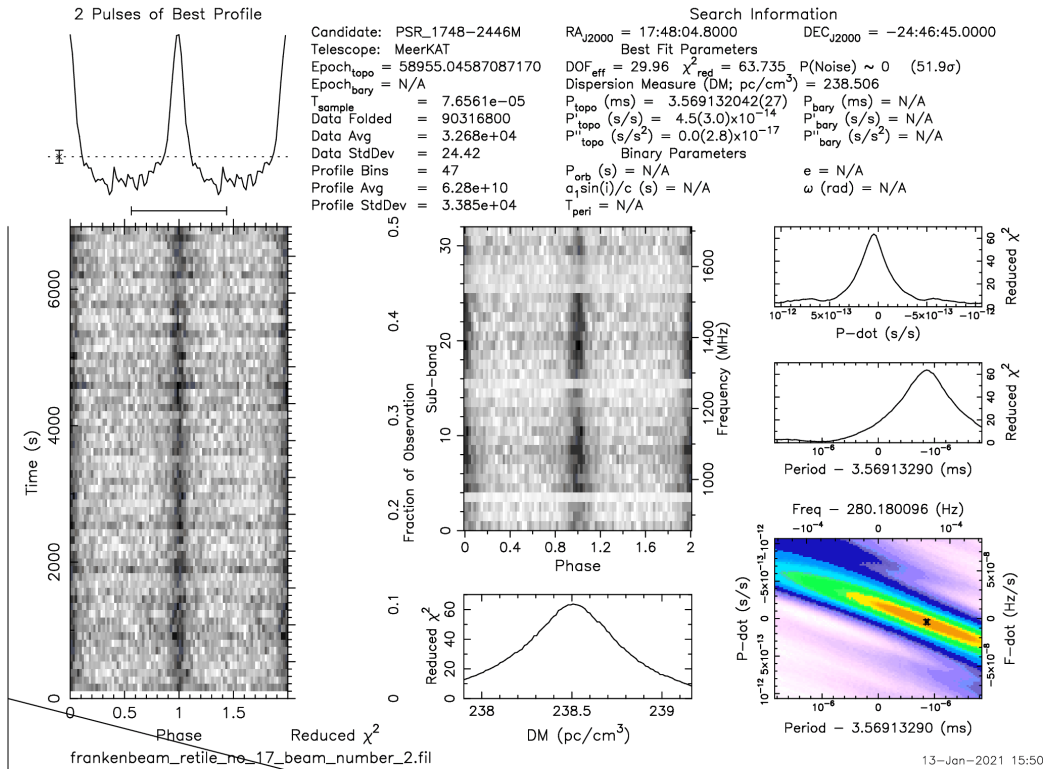
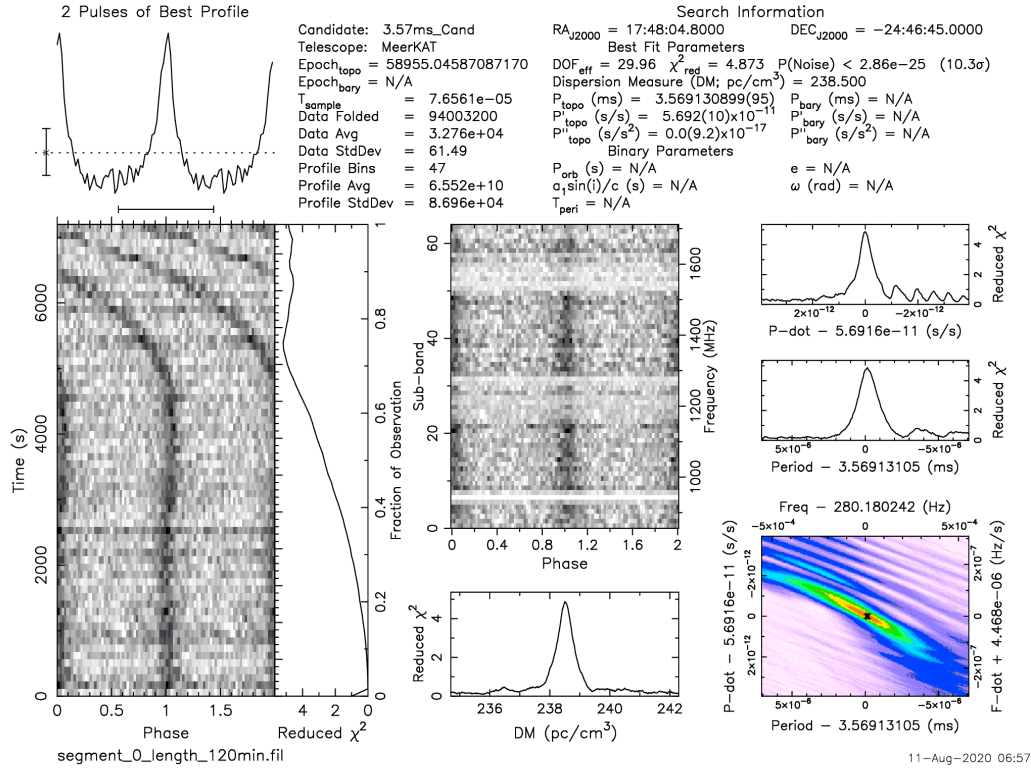


Figure 3.10: A comparison of Ter5M detections with an ephemeris fold and the fold generated by the search pipeline. Top panel: Detection plot from pipeline. Bottom panel: Detection plot from folding with ephemeris of Ter5M. This binary pulsar has an orbital period of 10 hours. The folds shown here are across the first half of the observation lasting about 2 hours. Given that the integration time is more than 10 percent of the orbital period, the acceleration alone does not suffice in removing binary motion smearing as shown in the pipeline detection plot.

beams. Ter5S and Ter5Y and were missed due to the limited integration length (2.8 hr instead of 4 hr) which drove the S/N below the fixed threshold for the pipeline. This was confirmed upon conducting a specific search on the closest beams to Ter5S (Beam number 1) and Ter5Y (Beam number 2). The search was done with the entire 4-hr integration length after downsampling in time by a factor of 2 (153 μ s sampling time) in order to adjust the file size within the RAM capabilities of the pipeline. The same search pipeline when run through the downsampled data was able to detect both the pulsars. Ter5ae is a compact binary pulsar with an orbital period roughly same as the integration length (approx. 4 hours). Although the pulsar was detected when folding with an ephemeris, the pulsar was most likely missed in the search (including segmented searches) due to smearing caused by effects of higher order derivatives of acceleration. Although a 30 min length is expected to be dominated by acceleration alone (since the integration length is less than 10 per cent of the binary period; see [Ransom et al. \(2002\)](#)), the folded S/N at these lengths were less than the detection threshold (approx. 7.5 sigma). Ter5am is in a binary system with an orbital period of 19 hours. Thus, the pulsar was most likely missed in the full length search due to smearing from higher order binary effects. This is not surprising given that the pulsar was initially discovered with a jerk search pipeline ([Andersen & Ransom, 2018](#)). The ephemeris fold revealed an S/N of 14 for the full integration length thus indicating that the segmented searches do not have the sensitivity to detect the pulsar. Ter5P is an eclipsing binary and the folding detection revealed that the pulsar was in an eclipsing phase for $\sim 85\%$ of the observation. The 30 min segmented searches also could not reveal the pulsar. On folding the final 30 min segment, the S/N obtained for Ter5P was 9.06. Given that this number is very close to the detection limit, it is not surprising that the detection was missed. Finally, the refined searches on Ter5al did not reveal a detection. This is not surprising given that [Cadelano et al. \(2018\)](#) stated that the pulsar was below the detection limit for most archival GBT observations. They also quoted a flux value of 8 μ Jy which is much lower than the theoretical limit of the TRAPUM survey and is thus consistent with a non-detection (see Figure 3.8).

3.5.8 Discussion

Although the segmented acceleration searches did not reveal a new pulsar detection, the Terzan 5 data provided a robust test case for the pulsar search infrastructure. Firstly, the infrastructure allowed for processing beams at a scale that is an order of magnitude higher than the number of beam elements used in previous and ongoing surveys (e.g. 13-beam receiver with Parkes, 7-beam receiver at Effelsberg and the 19-beam receiver with FAST). This was feasible owing to using an interferometer like MeerKAT and complementing instrumentation (FBFUSE and APSUSE systems). Benchmarking results revealed that the GPU search pipeline takes 35 milliseconds per search trial for the full observation length. In comparison, a standard CPU-based PRESTO pipeline takes ~ 6 seconds per search trial, thus demonstrating the advantage of using GPUs. The flexibility of the beamformer was demonstrated by the two Terzan 5 runs (R1 and R2) where specific tied-array beams could be configured to point at known pulsar

positions. The potential of the Kubernetes infrastructure deployed on APSUSE was demonstrated where upto 60 beams can be searched simultaneously.

The Terzan 5 acceleration searches provided an opportunity to carry out a large scale search including regions beyond the half mass radius of the cluster. The segmented searches on the inner beams were fine tuned to search for circular binaries with orbital periods of 5 hours and above. Additionally, the acceleration ranges were fixed based on a maximum pulsar companion mass of $4 M_{\odot}$. The non-detection of a new confirmed pulsar places a theoretical flux density limit of ~ 18 uJy on any such binary MSP detection (assuming a spin period of 1 ms) within the half mass radius of the cluster. The impact of dispersive smearing due to subbanding is minimal as shown in Figure 3.8. However, the practical limit is influenced by several external factors including refractive scintillation and radio frequency interference. A thorough examination of these factors with future observations could provide vital points to keep in mind while conducting future searches. On the other hand, the outer beam periodicity searches were able to detect Ter5D and also included multiple redetections of Ter5A. However, no new confirmed discovery implies a flux density limit of ~ 18 uJy for an isolated millisecond pulsar lying within a $3.6'$ radius of the core of the cluster.

The known pulsars missed in the searches demonstrate a need for higher order searches. Some pulsars were missed owing to smearing of the signal in tight binary orbits (e.g. Ter5ae). This calls for higher derivative searches incorporating the jerk component. Furthermore, given the massive processing capability of APSUSE, a coherent search for circular binaries using template bank routines (e.g. Allen et al., 2013) is feasible. If not a dedicated large scale search, such a coherent search routine could be utilised to confirm potential binary discoveries from the lower dimension acceleration or jerk searches. Another major hindrance to the searches was the cap in the observing length. This requires a revisiting of the memory management of the GPU pipeline and is already being looked into. By incorporating the full observation length into the pipeline, the searches would benefit from a 20 percent improvement in sensitivity. Additionally, the S-Band is the preferred band to observe high DM clusters like Ter5 owing to minimal dispersive and scattering effects. For this reason, future TRAPUM observations can also benefit from prioritising Ter5 observations at S-Band once the commissioning phase is complete.

3.5.9 Conclusions

We have conducted a large scale multibeam acceleration search for pulsars in Terzan5 beyond the realms of the half mass radius. No new detection was made for these searches but limits were placed on the flux density of the potential binaries and isolated MSPs that were missed. The search pipeline was validated using known pulsar redetections and it demonstrates a reasonable efficiency (above 80 percent) for majority of the pulsars. Although compact binaries like Ter5A ($P_{\text{orb}} = 1.8$ hr) and Ter5O ($P_{\text{orb}} = 6.2$ hr) were detected in segmented searches, binaries like Ter5ae ($P_{\text{orb}} = 4.1$ hr) and Ter5am ($P_{\text{orb}} = 19.2$ hr) were missed owing to higher order binary effects. This demonstrates a need for higher dimensional searches in this globular cluster going forward.

Table 3.5: List of redetected Terzan 5 pulsars from the acceleration searches conducted on R1 (see text). Some pulsars under the orbital period column are marked as ‘i’ indicating that these are isolated. The Δ DM indicates the difference between the DM of the pulsar and the reference DM at which the subbanding was implemented. ‘Closest/Best beam’ indicates the closest beam number by angular separation to the pulsar and the beam number with the best S/N detection. These two numbers need not always be same since the beam shape is irregular. PER indicates the Pipeline Efficiency Ratio which is the ratio of the S/N from the search pipeline (S/N_{pipeline}) and the S/N obtained from a direct ephemeris fold (S/N_{fold}). The ‘Acceleration’ column indicates the corresponding acceleration value for the best detections made by the pipeline. Compact binaries like Ter5A and Ter5N show high acceleration values as expected.

Pulsar	Orbital period (hr)	Δ DM (pc cm^{-3})	Acceleration (ms^{-2})	Closest/Best beam	Segment detections	Best S/N segment	S/N_{pipeline}	S/N_{fold}	PER
Ter5A	1.815504	5.1	31.55	0/0	All	30 min	549.49	628.96	0.87
Ter5C	i	0	0.03	1/1	All	Full	202.14	203.48	0.99
Ter5D	i	6.83	0.0	48/48	Outer beam full	Outer beam full	29	48.06	0.60
Ter5E	1441.44	-0.16	0.04	33/33	All	Full	83.8	89	0.94
Ter5F	i	2.18	0.02	1/19	Full	Full	16.89	20.38	0.83
Ter5H	i	1.13	0.0	18/18	Full, 120 min, 60 min	120 min	16.5	17.18	0.96
Ter5I	31.872	1.73	-0.82	2/2	All	Full	50.91	68.78	0.74
Ter5K	i	-2.19	0.03	33/32	Full, 120 min, 60 min	Full	27.11	29	0.93
Ter5L	i	0.74	0.03	1/1	All	Full	43.09	44.56	0.97
Ter5M	10.6344	1.65	4.78	1/2	All	120 min	41.41	64.58	0.64
Ter5N	9.252	1.47	-14.75	2/4	All	30 min	24.16	25.61	0.94
Ter5O	6.228	-0.62	-2.54	2/2	All	120 min	33.21	67.87	0.49
Ter5Q	727.08	-2.5	0.34	32/32	Full	Full	18.18	25.93	0.70
Ter5R	i	0.6	0.03	2/2	Full, 120 min, 60 min	Full	36.77	37.42	0.98
Ter5T	i	0.7	0.02	61/61	Full	Full	22.0	22.75	0.97
Ter5U	85.68624	-1.54	0.0	2/2	Full, 120 min	Full	27.69	33.42	0.82
Ter5W	117.048	2.14	-0.14	2/2	Full, 120 min	Full	27.82	38.08	0.73
Ter5X	119.964	3.03	0.18	4/4	Full, 120 min	120 min	17.13	26.18	0.65
Ter5ab	i	1.34	0.03	2/2	Full, 120 min	Full	21.95	22.37	0.98
Ter5an	230.64	0.73	-0.01	19/19	Full	Full	12.90	16.75	0.77

The MPIfR Galactic Plane Survey: System setup and early results

This chapter focuses on the pulsar searching aspect of the MPIfR Galactic Plane Survey. As part of this project, I conducted the simulations, implemented the processing pipeline and co-wrote the white paper for the survey strategy. I am also responsible for conducting the searches and the corresponding data analysis presented here.

Contents

4.1	Introduction	125
4.2	Survey strategy	126
4.3	Instrumentation	130
4.3.1	Front-end	130
4.3.2	Back-end	131
4.4	Sensitivity	131
4.5	Simulations	135
4.6	Processing	136
4.6.1	Candidate viewing	138
4.7	Pilot survey	139
4.7.1	Search strategy and processing	139
4.7.2	Results	142
4.8	Conclusions and Future Prospects	148

4.1 Introduction

The most sensitive telescope in the Southern Hemisphere is currently MeerKAT with a maximum gain of 2.8 K/Jy. This is approximately a factor of 4 higher than the Parkes telescope. Given the high success rate of surveys with the Parkes telescope (as discussed in Section 2.4), conducting a galactic plane survey with a sensitive telescope like MeerKAT is an ideal prospect for expanding the known pulsar population among different sub-classes and in turn provide scope for key insights into different applications (see e.g. [Lorimer et al. \(2019\)](#) and references therein).

The MPIfR Galactic plane survey (MGPS) is a multi-purpose survey, covering science cases not exclusive to pulsars but also fast transients, Galactic magnetic fields,

continuum emission and spectral lines. The observations carried out would be common for all the science cases, thus stepping up the scheduling time efficiency of this survey. The survey strategy also ensures minimal loss of quality in the individual science goals. The unique design proves useful in preparation for future telescopes like the Square Kilometre Array (Dewdney et al., 2009) where telescope time management is key. The synergy between the different science cases allows for a feedback mechanism where the results from the imaging domain has repercussions on the time-domain analyses and vice-versa. In context with pulsar searching, continuum images would help identify steep spectrum sources that could be specific pointing targets. Similarly, polarisation maps revealing polarised sources with pulsar-like rotation measure features could also be followed up for potential radio pulsations. This commensality between different fields also helps adapting to better strategies as the survey progresses. Complementary to the scientific drivers, is the MeerKAT instrumentation that makes such a large-scale high-time-resolution survey feasible. The instrumentation provides the capability of beamforming, fine channelisation (up to 32k channels spanning the entire bandwidth) and generation of visibilities for imaging. From the perspective of pulsar searching, the instrumentation includes: a) a receiver operating at frequencies (S-Band) where deleterious propagation effects via the interstellar medium are minimum b) a multi-beam beamformer capable of tiling a significant fraction of the primary beam and c) a dedicated high performance cluster equipped with the necessary infrastructure for real-time processing.

This chapter focuses on the blind pulsar search aspect of the survey. The setup and results for the other science cases including targeted pulsar searches of continuum sources will be discussed elsewhere. Section 4.2 discusses the strategy employed for deciding the targeted regions on sky. In Section 4.3 we describe the instrumentation used for conducting the survey. Section 4.4 discusses the sensitivity achievable with this survey. In Section 4.5 we discuss the simulations done to determine the number of new canonical as well as millisecond pulsars expected from the survey. In Section 4.6, we describe the processing pipeline implemented for the searches. In Section 4.7, we describe a 30-hour pilot survey serving as a platform to test and demonstrate the smooth functioning of the entire setup. We state our conclusions in Section 4.8.

4.2 Survey strategy

MGPS has 3000 hours of allocated time which has been partitioned into three surveys. In this section, we give a brief introduction to each of these surveys. The integration times and centre frequency chosen for each of the surveys are primarily based on maximising the yield of the number of discoveries within the allotted time. These discovery numbers come from simulations that are discussed in further detail in Section 4.5.

- **Deep S-band Galactic plane survey:**

This survey is aimed at discovering compact binary pulsars along the Galactic plane that were previously missed due to limitations from interstellar medium

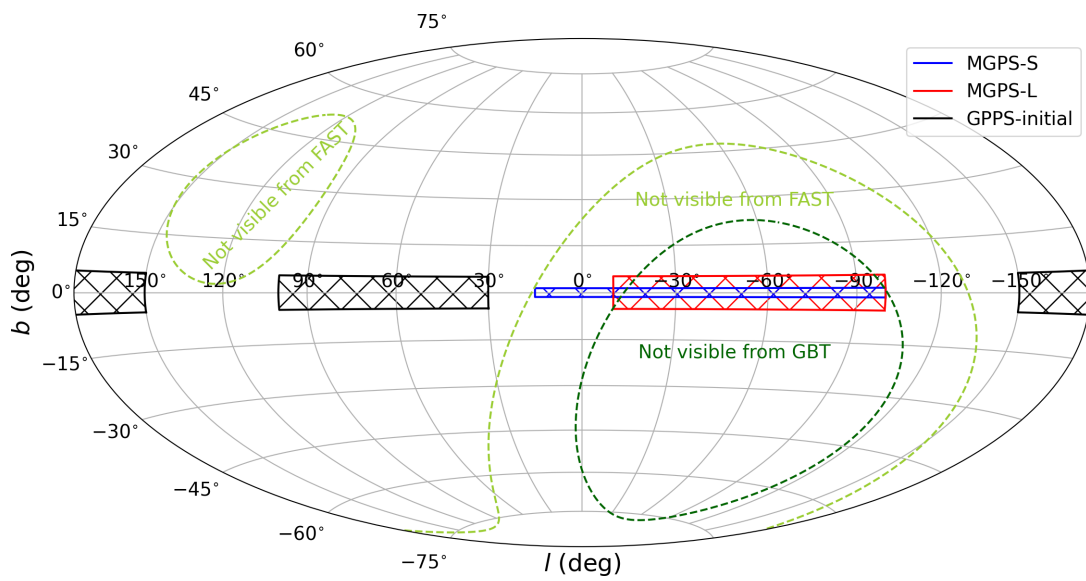


Figure 4.1: The survey region layout for the latest and most sensitive Galactic plane surveys. The red region indicates the L-band portion of MGPS while the blue region shows the deep S-band portion of the survey. It should be noted that the L-band region is continuous in latitude with dual frequency coverage achieved between $-100^\circ < l < -10^\circ$ and $|b| < 1.4^\circ$. The coverage of the Galactic Plane Pulsar Snapshot survey with FAST is shown as a comparison. The corresponding regions that are not visible from the Green Bank telescope and FAST are also overlaid. The regions are calculated based on the declination limits imposed due to their locations (see text).

Table 4.1: The observation parameters of the three MGPS sub-survey regions. The parameter t_{dwell} corresponds to the planned integration time for each sub-survey based on the time constraint and survey coverage limitations (See text). N_{chan} corresponds to the number of frequency channels, t_{samp} is the sampling time and ν_{centre} is the observation centre frequency.

Survey	Duration (hrs)	Lat. range (deg.)	Long. range (deg.)	t_{dwell} (s)	N_{chan}	t_{samp} (μs)	ν_{centre} (MHz)
MGPS-L	800	$ b < 5.2$	$-100 < l < -10$	637	4096	153	1284
MGPS-S	2000	$ b < 1.4$	$-100 < l < 15$	1274	512	153	2406.25
MGPS-Sgr A*	200	$b = 0$	$l = 0$	1274	4096	76	3062.5

(ISM) propagation effects . The survey coverage is 300 sq. deg. maximising the Galactic longitude coverage at the expense of latitude coverage (see Table 4.1) and uses about 2000 hours of the total allocated time. The integration length is fixed at 20 minutes ensuring an ideal balance between maximum yield and maximum detectable acceleration of the binary system (see Section 4.5 for details on deriving the best integration time). We refer to this survey as MGPS-S hereafter.

- **Shallow L-band Galactic plane survey:**

This survey is done over shorter observations of 10 minutes but covering a wider area (approx. 900 sq. deg.) of the Galactic plane. This survey covers 800 hours of observing time and is aimed at discovering pulsars that were previously missed owing to limited telescope sensitivity. Given that the survey is conducted at L-Band and at short integration times (10 min), MSPs discovered from this survey would most likely be integrated into the current PTAs. Henceforth, we refer to this survey as the MGPS-L.

- **Ultra-Deep S-band Galactic centre survey:**

The remaining 200 hours of the MGPS is used for observations centered on Sagittarius A* (see Figure 4.2). This survey is conducted at the highest available S-Band central frequency (i.e. 3062.5 MHz) in order to avoid maximum possible propagation effects due to the ISM. The S-Band primary beam width is ~ 0.5 deg. on the high frequency end and thus would be able to cover a ~ 0.2 sq. deg. field. Assuming that the Galactic centre is ~ 8.1 kpc away (Gravity Collaboration et al., 2019), the primary beam would span 70.65 pc. The corresponding tied-array beam at the high end of S-Band is 13 in size (assuming boresight) which would imply that the best achievable localisation is 0.5 pc around Sgr A*. To further exploit the richness of these data, multiple observations would be stitched post-fact to perform extremely deep searches for pulsars in orbit around the central black hole. We refer to this survey as MGPS-Sgr A* hereafter.

Table 4.1 gives a summary of the the regions for each of the survey’s observing frequencies, dwell times and extents for each of the survey regions. Each region was selected based on the following criteria:

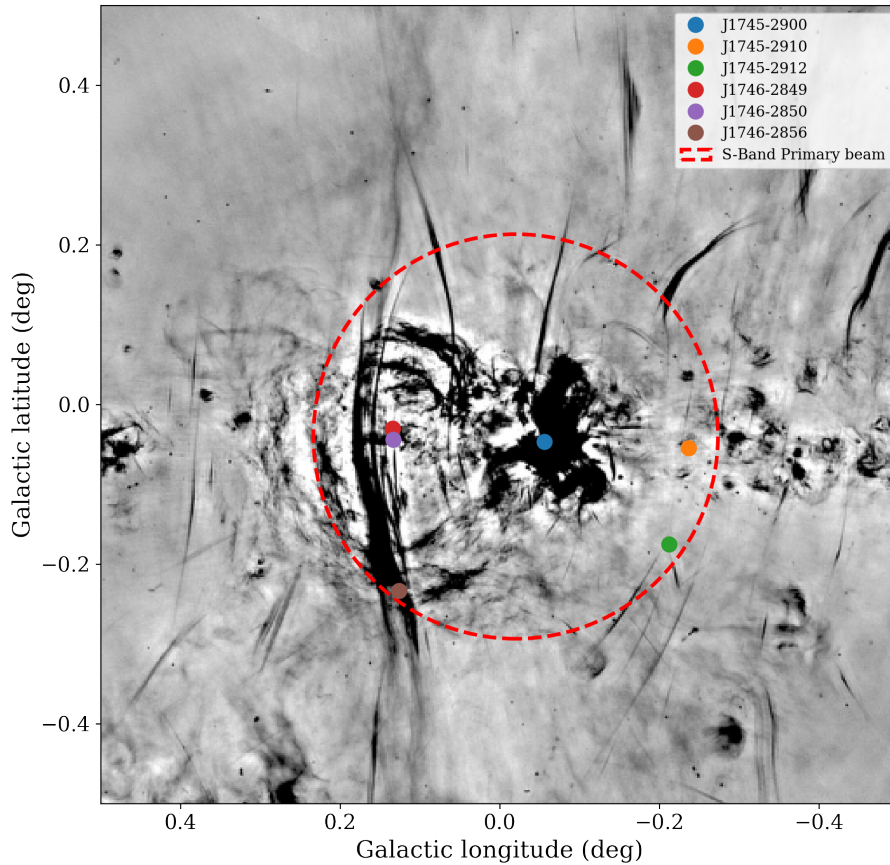


Figure 4.2: The expected coverage of the MGPS-Sgr A* survey is shown in context with the image published by [Heywood et al. \(2019\)](#) at L-Band using MeerKAT. The image of the inner sq. deg around the Galactic Centre has been overlaid. The primary beam for the high frequency end of S-Band (3062.5 MHz) is also shown with Sgr A* chosen as the boresight position. All 6 known Galactic centre pulsars can be accommodated within the primary beam.

- The total possible survey coverage at each frequency is constrained by the allocated time, integration length and the total number of synthesised beams that can tile a pointing. We have also assumed a 90% efficiency for the allocated time to accommodate the time required for flux and phase calibration.
- For each pointing, we assume that up to 960 tied-array beams are available and that we observe with an effective bandwidth of 684 MHz (accounting for known RFI bands and band edges).
- The upper longitude limits for the surveys are restricted by the declination limits of the most sensitive telescopes in the Northern Hemisphere which have previously conducted or are conducting Galactic plane surveys in these frequency bands. For the L-Band survey, the declination limit of the Green Bank Telescope (GBT) is imposed ($\delta < -40^\circ$). Similarly, for the S-Band survey, the upper longitude is limited to ensure no overlap with the FAST sky ($-15^\circ < \delta < 65^\circ$).
- The lower limits on the longitude are survey dependent. For the shallow L-Band survey, we have chosen a region to maximise the predicted yield of pulsar discoveries (see Section 4.4) with the caveat of ensuring that certain targeted regions (e.g. the Carina arm) fall in the specified region. On the other hand, the deep S-Band survey has the same lower longitude limit to ensure maximum overlap between L-Band and S-Band regions. The priority of the S-Band survey is to probe regions close to the Galactic plane. Hence we have traded latitude coverage for a wider longitude coverage.
- The survey grid scheme is a hexagonal tessellation pattern with the separation between pointing centres in a row set to $\text{FWHM}/\sqrt{5}$ where FWHM is the Full Width Half Maximum of the primary beam. This was chosen to ensure uniform sensitivity for imaging. Additionally, the corresponding filling factor i.e. the fraction of the incoherent beam area covered by the tied-array beams is also not compromised ensuring maximum efficiency and highest pulsar discovery potential.

4.3 Instrumentation

The setup and equipment used for data acquisition and processing of the MGPS survey is described below:

4.3.1 Front-end

4.3.1.1 L-Band receiver

The L-Band receiver is one of the two receivers provided by the MeerKAT facility, the other being the Ultra High Frequency receiver (as explained in Chapter 3). The feed horn is dual polarisation consisting of two orthogonal linear receptors. Specific parts of the bandpass are corrupted by known RFI signals leading to an effective bandwidth of 684 MHz. The system temperature of the receiver alone is 18 K and an extra

contribution ranging from 4-7 K is added from the atmosphere and ground spillover. The system equivalent flux density (SEFD) is as low as 400 Jy for an individual dish at the centre of the band (see Figure 4.3).

4.3.1.2 S-Band Receiver

The S-band receiver system is designed by the Max Planck Institute for Radio Astronomy, Bonn (MPIfR) as a complementary receiver to the existing L-Band and UHF receivers at MeerKAT. The frequency coverage of the receiver ranges from 1.75-3.5 GHz with a maximum digitised bandwidth of 875 MHz. The selected observing band can be centered at five different centre frequencies (2187.50, 2406.25, 2625.00, 2843.75 and 3062.5 MHz) within the usable bandwidth. The development of this receiver is motivated by different science cases including Galactic centre pulsar searches, rotation measure studies and probing star forming regions within the Galaxy. With regards to pulsar searching, this observing frequency range is particularly suited for deep searches along the Galactic plane where dispersion ($\tau_d \propto \nu^{-2}$) and scattering ($\tau_s \propto \nu^{-4.4}$ assuming Kolmogorov turbulence) effects begin to dominate. Similar to the other receivers, the S-Band receiver is equipped with a cross dipole dual polarisation receptor with direct digitisation at the focus. Latest testing results reveal a system temperature of 22 K with the SEFD of ~ 450 Jy per individual receiver (see Figure 4.3). The S-Band receivers are currently in the commissioning phase with receivers installed on all 64 antennas at the time of writing.

4.3.2 Back-end

MGPS utilises the same back-end infrastructure as the TRAPUM survey (as explained in Chapter 3). The digitised streams from each individual antenna enter into the Correlator/Beamformer (CBF) network and into the F-engines. The channelised data output from the F-engines are captured by the multibeam beamformer (termed as FBFUSE) via the CBF network. FBFUSE is used to beamform and generate tied-array beams. Thus, the total number of files generated are $N+1$ where N corresponds to the number of tied-array beams requested and the extra product is the incoherent beam that is formed by directly summing the power from each antenna. These files are captured and recorded on the APSUSE cluster where the pulsar search pipeline is deployed. A detailed description of each of the above mentioned components in the processing chain is given in Chapter 3.

4.4 Sensitivity

To estimate the sensitivity of MGPS-L and MGPS-S surveys, we use the modified radiometer equation given by Equation 3.1 (see Chapter 3). G refers to the gain of the telescope and is chosen to be 1.92 K/Jy which corresponds to the total gain for

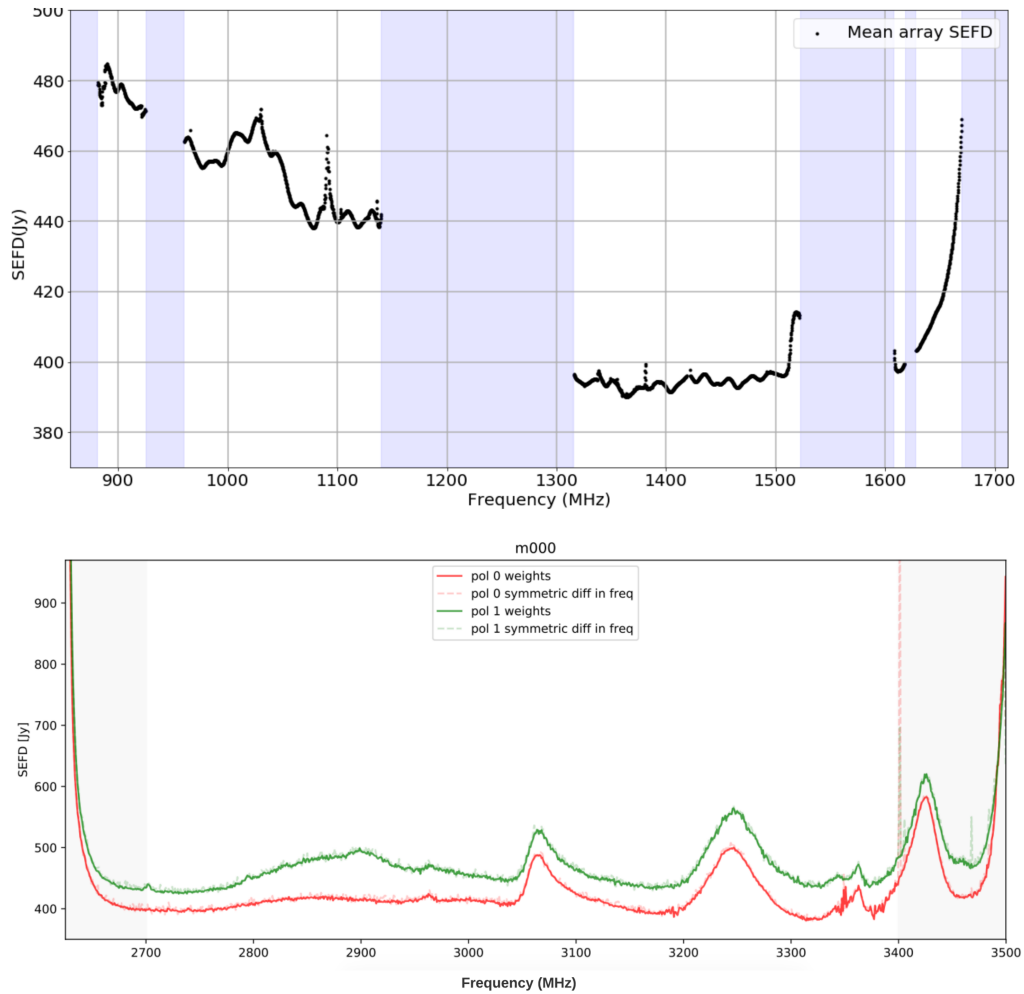


Figure 4.3: Top panel shows the mean measured SEFD for a single MeerKAT dish at L-band. The blue regions are masked out for RFI and band edges. This figure were provided by the South African Radio Astronomy Observatory (SARAO). Bottom panel shows the measured SEFD for both the polarisations (pol0 and pol1) for the MeerKAT dish m00 at the top half of the frequency coverage of S-Band. This is from a commissioning run on 14th December 2020. This figure was provided by Olaf Wucknitz.

44 dishes making up the inner core of the array¹. The system temperature at L-Band (T_{sys}) is 27.8 K. This is after taking into account the receiver temperature (18 K) as well as the sky contribution (T_{sky}) of 5.3 K. T_{sky} was derived from scaling the map at 408 MHz made by [Haslam et al. \(1982\)](#) with a spectral index of -2.6. The system temperature also accounts for the ground spillover temperature which is 4.5 K at 45 degree elevation. Similarly at S-Band, the system temperature is chosen to be 24 K. This includes the receiver temperature of 22 K and a sky temperature of 2 K following the method applied at L-Band. We have not considered the ground spillover at S-Band due to lack of experimental data. $n_{\text{pol}} = 2$ since the two orthogonal polarisations are summed. $\beta = 1.01$ accounting for digitisation losses ([Kouwenhoven & Voûte, 2001](#)). The integration time t_{obs} is chosen to be 10 minutes for the MGPS-L and 20 min for the S-Band survey as specified in Table 4.1. We chose a duty cycle of 5.5 percent which is the median duty cycle of the set of known pulsars with reported pulse widths². The signal-to-noise ratio (S/N) was chosen as 10 based on the False-alarm probability (see e.g. Equation 6.15 from [Lorimer & Kramer \(2012\)](#)).

The search efficiency factor ε was chosen as 0.7 for Fast Fourier Transform (FFT) based searching as determined by [Morello et al. \(2020\)](#). We compared the sensitivity estimates between the surveys of MGPS with legacy Galactic plane surveys like HTRU South low-lat (e.g. [Cameron et al., 2020](#)) and PALFA ([Cordes et al., 2006](#)) as well as with the ongoing Galactic Plane Snapshot Survey conducted with FAST. Figure 4.4 shows a comparison between each of the surveys. The Figure shows that MGPS-S is a factor of 2 more sensitive than the HTRU South low-lat survey. Similarly, MGPS-L, is about 1.2 times more sensitive than HTRU South low-lat (HTRU parameters are given by [Keith et al. \(2010\)](#)). Although PALFA and GPPS surpass the sensitivity of MGPS, the survey regions do not overlap. The improved sensitivity over HTRU South low-lat with a lower integration time (637 s vs 4300 s) boosts the probability of several more binary pulsar discoveries in particular. This is because a shorter integration time would be less prone to binary motion effects above the the linear acceleration regime. Thus the linear acceleration regime is a valid approximation for a wider range of binary pulsar orbits (assuming $T < P_{\text{orb}}/10$; see e.g. [Ransom et al. \(2003\)](#)). An important point to note is that the numbers are an overestimate of the true sensitivity of the survey. Apart from RFI, red-noise arising primarily due to the mains power supply would further reduce the sensitivity especially for long period pulsars.

¹The inner core alone is used for MGPS observations to ensure that the filling factor does not drop below a certain level due to narrow beams. Although the inner core consists of 42 antennas, the FBFUSE beamformer kernel requires the number of antennas to be a multiple of 4 for efficient data packing. Hence the observations use 44 dishes

²The W10 parameter was used for the calculation of the duty cycle as reported in the ATNF catalogue [Manchester et al. \(2005\)](#)

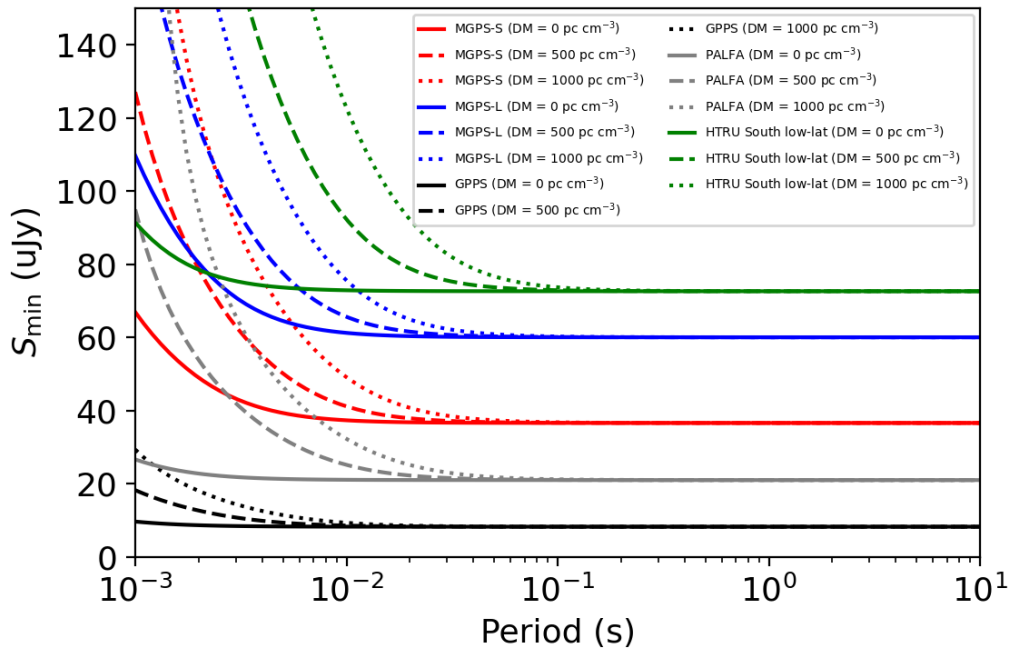


Figure 4.4: Theoretical flux densities (S_{\min}) as a function of spin period for current (GPPS, MGPS-L, MGPS-S) as well as previous (PALFA, HTRU-South low-lat) Galactic plane surveys. The sensitivity curves are calculated at three different DM values (0, 500 and 1000 pc cm^{-3}). The intra-channel dispersive smearing is responsible for different sensitivity limits at different DMs. A minimum detectable S/N of 10 is chosen and a duty cycle of 5.5 percent. Effects due to scattering are neglected here.

Table 4.2: List of input parameter distributions used for the PsrPopPy simulations of canonical pulsars and MSPs.

Parameter distribution	Canonical pulsar	MSP
Spin period	Log-normal as used in Lorimer et al. (2006) (mean=2.7, variance=-0.34)	Log-normal as used in Lorimer et al. (2015) (mean=1.5, variance = 0.58)
Luminosity	Log-normal as given in FaucherGiguère & Kaspi (2006) (mean=-1.1, variance = 0.9)	Log-normal as given in FaucherGiguère & Kaspi (2006) (mean=-1.1, variance = 0.9)
Radial	As described in Lorimer et al. (2006)	As described in Lorimer et al. (2006)
Height above plane	Exponential with scale height = 0.33 kpc (Lorimer et al., 2006)	Exponential with scale height = 0.5 kpc (Lorimer et al., 2006)
Spectral Index	Gaussian with mean =-1.6 and variance =0.5	Gaussian with mean =-1.4 and variance =0.5
Electron density	NE2001 (Cordes & Lazio, 2002)	NE2001 (Cordes & Lazio, 2002)
Duty Cycle	6 percent	20 percent

4.5 Simulations

In order to decide the optimal dwell times for each survey within the MGPS, simulations were carried out using PsrPopPy³ (Bates et al., 2014), which is a software package for pulsar population synthesis. The input parameters and the distributions used for the simulations are summarised in table 4.2. In order to make the simulations more robust we also incorporated the following constraints:

- Every realisation was generated with the snapshot method. This method is a Monte Carlo-based approach that uses statistical distributions of a set of input parameters (as given in Table 4.2) to generate a Galactic pulsar population realisation. The simulation generates pulsars using the input parameter models and stops populating the realisation when the number of pulsars detected reaches the expected number from a reference survey. The constraint here was 1038 pulsar detections from the PMPS.
- Since we are primarily interested in new discoveries, we added previous survey parameter files (including all surveys conducted with the multibeam receiver at Parkes) as predecessors to the MeerKAT survey. In the case of the simulations for MGPS-S, we also added the MGPS-L inputs as an extra prior for calculating the discovery numbers. This is a reasonable assumption given that discoveries at S-Band could be detected with L-Band.
- The snapshot simulation parameters are set from models for canonical pulsars

³<https://github.com/samb8s/PsrPopPy>

by default. In order to generate a robust population model for MSPs, we used a sample of 56 MSPs as mentioned in Lorimer et al. (2015) as a reference to generate the model (see Table 4.2 for specific parameters used).

- To make the estimation of expected discoveries more robust, we averaged the expected discoveries over 100 different realisations generated for both canonical pulsars and millisecond pulsars.
- To optimise the integration length, we ran the simulations for three different set of integration lengths for the L-Band as well as the S-Band survey. The results indicate that to maximise potential discoveries, an integration length of 637 s at L-Band and 2548 s at S-Band are the ideal combination⁴. Although 2548 s yields the most discoveries, the processing cost of performing acceleration searches on such data is unrealistically high. As such we limit the dwell times to a more manageable 1274 s.
- At S-Band, we ran the simulations for the 5 different centre frequencies available. We observed that using the lowest part of the band results in the highest pulsar yield (as might be expected from radio pulsars being steep spectrum sources).

The simulation results are summarised in Figures 4.5 and 4.6. The expected yield from MGPS-L is 476 ± 23 canonical and 22 ± 6 MSPs, whereas the numbers from MGPS-S indicate 220 ± 15 canonical and 10 ± 2 new MSP potential discoveries. The errors for the expected number of discoveries come from the 1-sigma estimates after averaging over the hundred realisations.

Due to the lack of empirical models for pulsars around Sgr A* (owing to a small number of pulsars found in this region), we did not conduct a separate simulation for MGPS-Sgr A*. The dwell time for MGPS-Sgr A* was chosen to be 20 min (see Table 4.1) based on the amount of data that can be recorded at a time and the corresponding limitation on the amount of RAM available for first-order acceleration searches.

4.6 Processing

Similar to the TRAPUM setup, the MGPS uses processing pipelines implemented via the APSUSE cloud-computing infrastructure that is managed by `Kubernetes` (K8s) and executed via `Docker` deployments as discussed in Chapter 3 (see Figure 3.3). Of the various processing pipelines available with the infrastructure (see Section 3.4.3), MGPS uses the acceleration searching, multi-beam candidate filtering and the folding with ML ranking pipelines in a sequential manner. The folding pipeline generates a diagnostic PNG plot, a folded archive file and a CSV file giving details of the input and final optimised parameters (i.e. spin period, DM and acceleration). The file also contains scores from the PICS machine learning classifier (Zhu et al., 2014) and the S/N

⁴The reason for choosing multiples of 637 is to ensure that the total number of samples is a power of 2. For example an integration time of 637 s at $76 \mu\text{s}$ resolution is 2^{23} samples. This speeds up computation for subsequent processing (e.g. FFT operations).

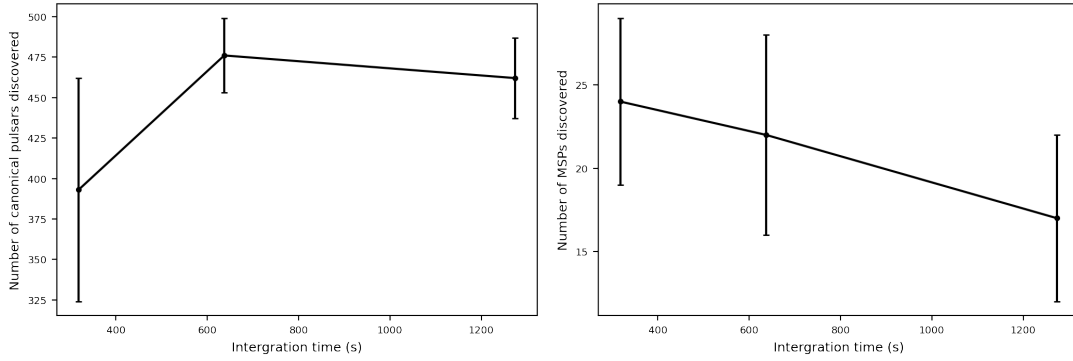


Figure 4.5: Number of expected canonical pulsar (left panel) and MSP (right panel) discoveries for different integration times at L-Band for the effective bandwidth.

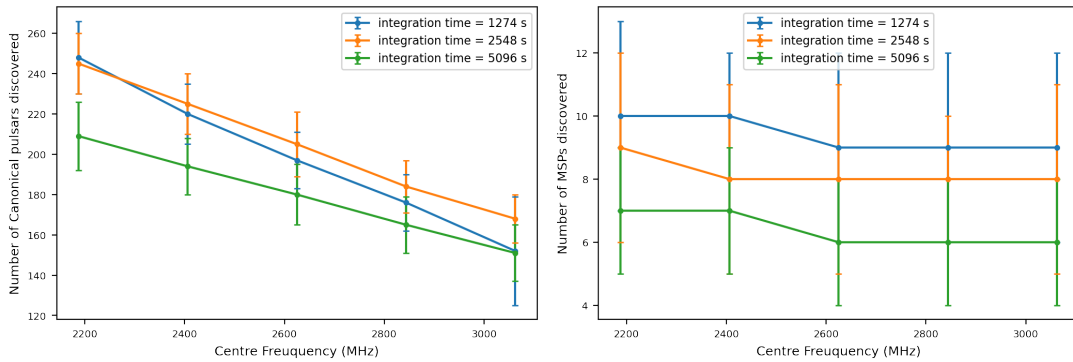


Figure 4.6: Number of expected canonical pulsar (left panel) and MSP (right panel) discoveries for different centre frequencies at S-Band for the full bandwidth.

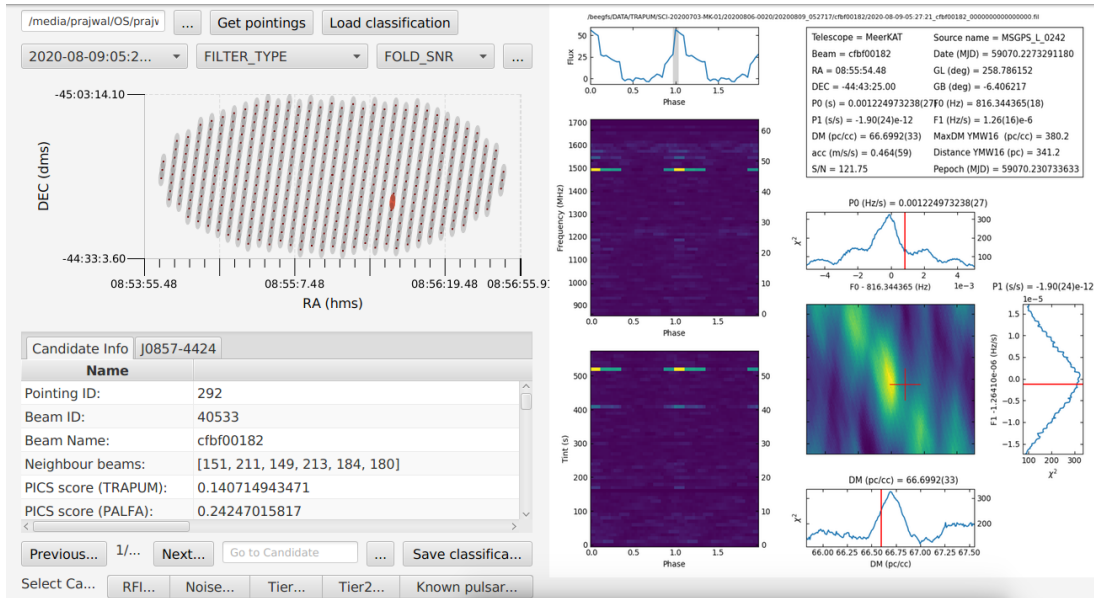


Figure 4.7: A snapshot of the **CandyJar** viewer interface. Apart from displaying the candidate plot generated from **PulsarX**, the viewer provides information on nearby known pulsars and candidate metadata that ease the process of human decision making. The labels decided by the user are stored into a file post-viewing. These labels prove useful in building training sets for candidate classifiers.

obtained from the FFT search as well as post folding. Other information parameters include the maximum DM and distance along the line of sight based on the YMW16 model (Yao et al., 2017) which helps in discerning the nature of the candidate. Apart from the CSV file, a meta file containing the tied-array beam coordinates and shape parameters of the corresponding observation is also retrieved. Candidates above a certain classifier score and S/N threshold are retained for visual inspection.

4.6.1 Candidate viewing

A specialised candidate viewing tool termed **CandyJar**⁵ has been developed to ease the human classification process (See Figure 4.7). The tool displays the diagnostic plot generated from **PulsarX** as well as corresponding candidate metadata parameters. The tool provides options to mark the candidate either as a known pulsar, RFI, noise or a potential pulsar. The user responses are recorded and cross-matched to select promising candidates to follow up. The classification would also be used in the future for training supervised machine learning classifiers.

⁵<https://github.com/vivekvenkris/CandyJar>

4.7 Pilot survey

We conducted a 24-hour pilot survey at L-Band in order to test and benchmark the entire observing and computing infrastructure. This survey was conducted across two 4×3 sq. deg. fields which is within the survey region of MGPS-L (see Figures 4.8 and 4.9). The field was chosen to overlap with a previously imaged field of MeerKAT. The motive of selecting this field was thus to not only validate the functioning of the pulsar search pipeline using known pulsars in the field but also test the imaging pipelines using known continuum and polarised sources. The survey grid scheme is shown in Figures 4.8 and 4.9. The known pulsars expected to be visible in the fields are also marked in the grid. The grid consists of 114 pointings in total with a dwell time of 600 s for each pointing. Each pointing generated 480 tied-array beams and an incoherent beam. The sampling time was set to $153 \mu\text{s}$ with 2048 frequency channels⁶ across 856 MHz bandwidth centered at 1284 MHz.

4.7.1 Search strategy and processing

The PEASOUP acceleration search pipeline (as explained in Section 3.4.3.2) was used for the searching. The numbers chosen for the DM and acceleration ranges for the search trials were mainly capped by processing time constraints. Similar constraints have been applied at different steps of the processing, the details of which are explained below. The main assumption was that the processing speed on APSUSE is eight times slower than real-time. On one hand, this ensures that the robustness of the pipelines is not compromised while focusing on enhanced processing speed alone. On the other hand, it also ensures that new observations can be scheduled on a weekly cadence given the caveat of the storage space on APSUSE.

We chose a DM range of $0\text{-}500 \text{ pc cm}^{-3}$ with a step size of 0.1 pc cm^{-3} as generated by `DDplan.py`. Apart from reducing processing time, the upper DM limit was also ensured to exceed the DM of all known pulsars in the survey field. The acceleration range was set to $-150 \text{ to } 150 \text{ ms}^{-2}$ with a default acceleration tolerance of 1.11. The number of candidates per beam produced from PEASOUP were limited to 1000 and the S/N threshold was set to 8.0 based on false-alarm statistics. The candidates generated from each of the 481 beams were put through the `candidate_filter` pipeline (Section 3.4.3.3) before folding. Post candidate filtering, a cap of 50 candidates per beam was set for the folding pipeline. This was a reasonable cap limit given that the candidates per beam produced from the candidate filtering pipeline for did not exceed 50 for 99 per cent of all the beams (including all pointings). Additionally, an S/N threshold of 9.5 was chosen as the cutoff to select candidates before filtering. The motive behind the higher S/N threshold was to reduce the time required for filtering and subsequent folding by a factor of at least two. We selected candidates which scored above 0.1 on the

⁶The first 4 hours of the survey was configured with 4096 channels. Although the intended frequency resolution for MGPS-L is 4096 channels, we chose to go a factor of 2 lower for the subsequent observations. This was done after several setups were trialled to assess differences in processing performance.

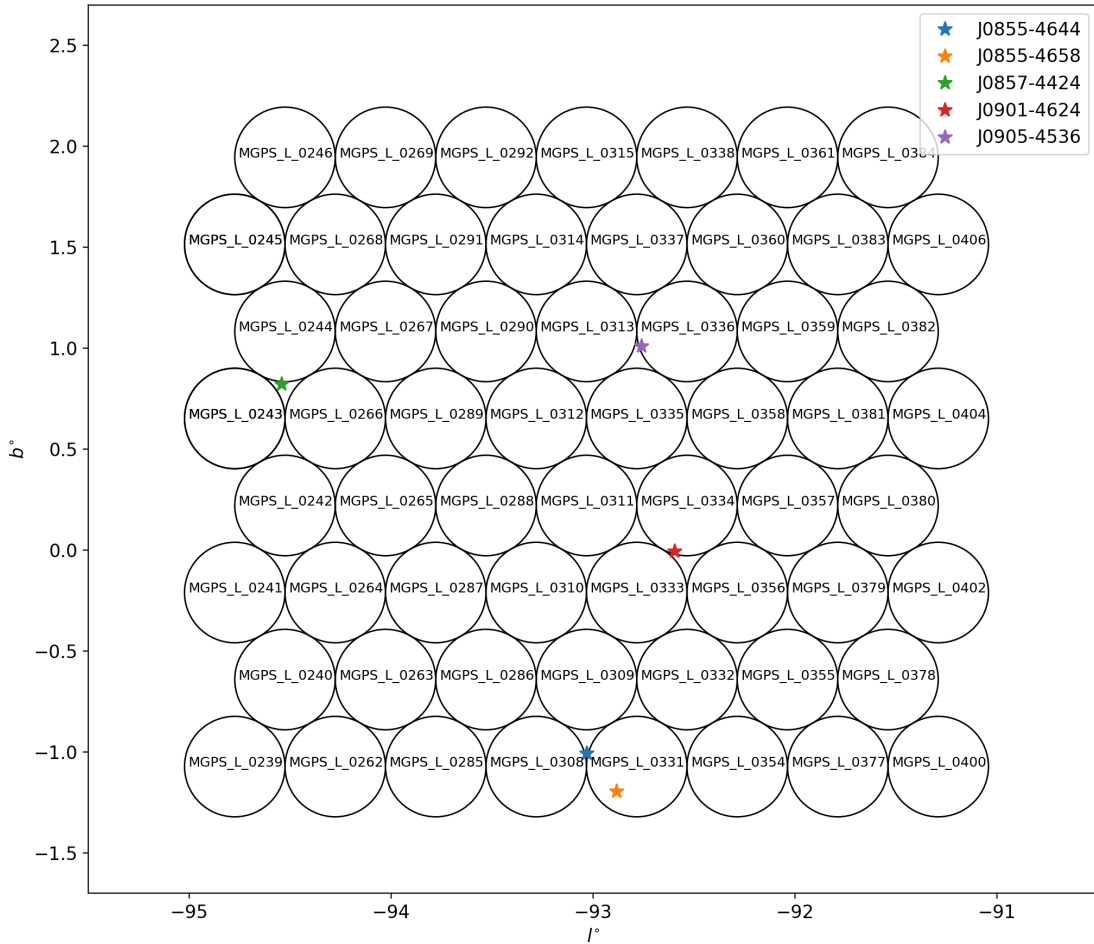


Figure 4.8: The gridding pattern for the pilot survey conducted within the time allotted for MGPS-L. The survey consists of two fields with one of the fields represented here. Field 1 (Top panel) and Field 2 (bottom panel). The radius of each pointing is scaled to $\text{FWHM}/\sqrt{5}$ where FWHM is the full width half maximum of the primary beam. The known pulsars within the fields are also marked.

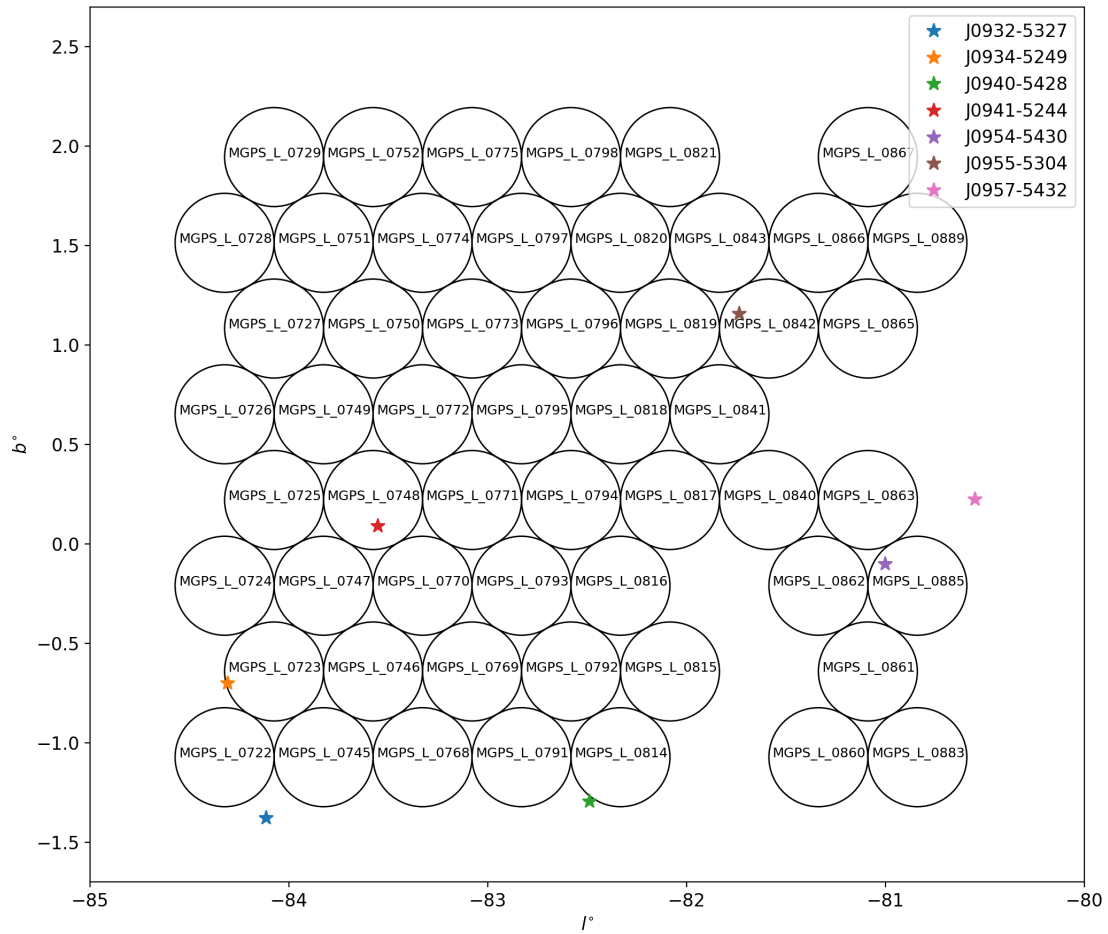


Figure 4.9: The gridding pattern for field 2 of the pilot survey. The pointing specifications are same as Field 1 which is shown in Figure 4.8 The gaps in pointings covered within Field 2 is due to a technical issue during one of the runs. The corresponding pointings were discarded.

PICS ML classifier for candidate viewing. This conservative approach on the classifier was chosen to avoid missing weak potential candidates that have scored poorly due to different statistics in comparison with the original training set used for generating the PICS model. Furthermore, we removed candidates below a DM of 3 pc cm^{-3} given that such candidates are most likely RFI.

4.7.2 Results

The ongoing searches have revealed candidates that await confirmation. In order to understand the efficiency of the pilot survey, we have conducted multiple analyses that are described below:

4.7.2.1 Survey beam coverage

We first estimated the coverage factor achieved per pointing based on the configuration parameters. In order to calculate this, we represent the survey beam coverage factor (SBCF) as a function of the tied-array beams (TABs) as:

$$\text{SBCF} = \frac{(\text{No. of TABs}) * (\text{TAB area})}{\text{Survey beam area} * 0.91} \quad (4.1)$$

The factor of 0.91 represents the best possible hexagonal packing efficiency of ellipses over a given area. Thus the SBCF value needs to be compensated for the additional sky coverage owing to the imperfect tessellation of the TABs. We achieve an SBCF value of at least 60 per cent in all pointings except eleven pointings where 289 tied-array beams were used instead of 480. The reduced number of beams was a consequence of a configuration error during scheduling. Figure 4.10 demonstrates the SBCF as a function of telescope elevation for each of the 114 pointings in the pilot survey. The synthesised beams are expected to elongate at lower elevations owing to shorter projected baselines. This tends to increase the SBCF value when observing sources closer to the horizon. Besides this, different sets of antennas leads to different combinations of their respective spacing and would thus change the shape and size of the beam. Figure 4.10 also demonstrates that a longer baseline can decrease the achievable SBCF for the same elevation. Although, the size of the TAB depends on the distribution of antenna baselines in the inverse Fourier image plane (better known as the u-v space), a longer baseline can significantly reduce the TAB size and thus the SBCF.

In order to understand whether the current tiling configuration is optimal for the survey, we simulated a tiling of 480 beams for a pointing chosen at random. We modelled each TAB with a two-dimensional Gaussian profile. We generated multiple iterations of the tiling as a function of the overlap factor between beams (0.05-0.95) as well as telescope elevation (10-90 deg.). Figure 4.11 shows a heat map of the coverage at half power fractional gain across a cross a hexagonal patch centered at the boresight of the survey pointing and whose edge is equal to the survey beam radius ($\text{FWHM}/\sqrt{5}$). We chose a hexagonal patch given that the survey beams are close packed circles, thus providing an improved measure of the mean fractional gain across a given survey area. Figure 4.11 suggests that the best achievable mean fractional gain is a combination of

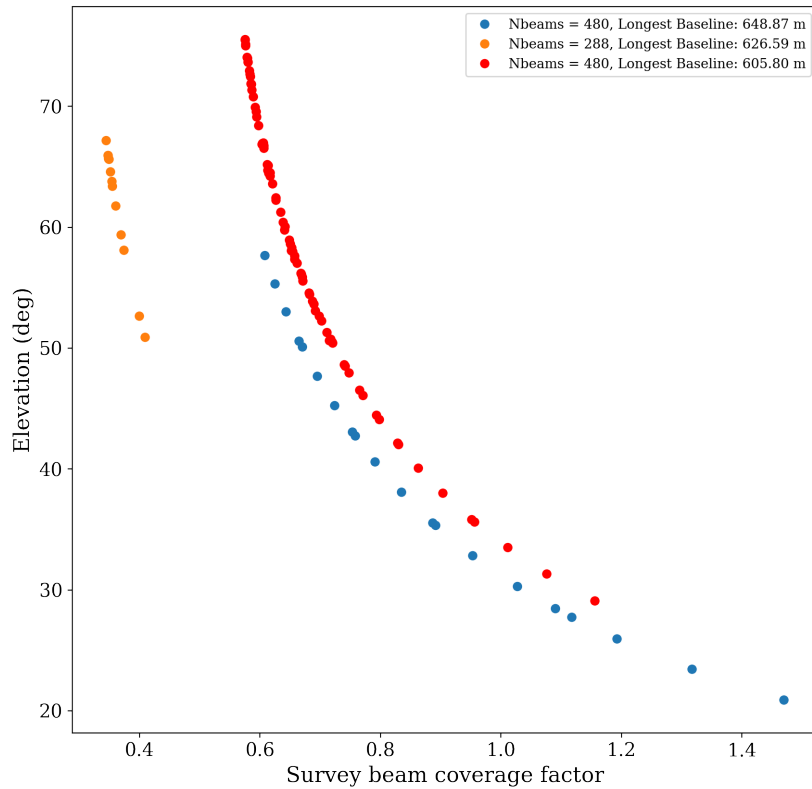


Figure 4.10: The dependence of the survey beam coverage factor on telescope elevation for the pilot survey pointings. Apart from the dependence on elevation, the antenna configuration used in an observing epoch also influences the achievable coverage factor for a given telescope elevation. The longest baseline between antennas pairwise is used as a measure to demonstrate this (see text). A factor above 1.0 indicates that the tied-array beams exceed the boundary of the survey beam i.e. $\text{FWHM}/\sqrt{5}$ of the primary beam.

low elevation (below 30 degrees) and a minimum overlap of 0.55. In this configuration, the TABs overlap beyond their FWHM points but are significantly elongated (due to the low elevation), thus achieving a right balance between coverage as well as overall sensitivity. Overlap factors below 0.3 show a more uniform gain value across elevation than for factors of 0.4 and above. This is due to the decrease in the filling factor of the TABs with a higher overlap, thus leading to significant voids within the hexagonal patch. Given that future observing runs are planned to span 8 hours per epoch, a wide range of elevations would be covered. This suggests that an overlap factor of 0.5 is reasonable for such runs. This ensures a near uniform sensitivity throughout the observing run (at all elevations), without compromising on the achievable coverage. Another important note is that low telescope elevations suffer from reduction in sensitivity due to increase in the the ground spillage and the atmospheric contribution to the system temperature. Accounting for this caveat in Figure 4.11 also suggests that the overlap factor of 0.5 is suitable for a reasonable coverage as well as sensitivity. Translating this information in terms of coverage, we achieve a minimum fractional gain of 0.5 for ~ 78 per cent of the survey field. In order to achieve a median coverage of 91 per cent (emulating a hexagonal closed packing strategy of survey beams) across all elevations at an overlap factor of 0.5, the minimum fractional gain is 0.16.

4.7.2.2 Known pulsar redetections

Four known pulsars were detected from the pilot survey searches. This included a primary beam side-lobe detection of the Vela pulsar which is known to be a very bright source ($S_{1400} = 1050$ mJy). Of the remaining three pulsars, one pulsar was detected in the incoherent beam (PSR J0955-5304). Despite the incoherent beam being ~ 6.5 times less sensitive than the coherent beams, this detection is not surprising. This is primarily due to the bottom half of the band (856 MHz to 1284 MHz) where the pulsar brightness increases significantly (spectral index estimated from published flux densities at 800 and 1400 MHz is -3.5).

The details of the other two pulsars are summarised in Table 4.3. We refrained from using the detections made from the incoherent beam for the survey sensitivity (PSR J0955-5304 and Vela). This is due to the poor sensitivity and major impact of RFI on the incoherent beam. The expected S/N for the other two pulsar detections were calculated using the radiometer equation along with the following conditions:

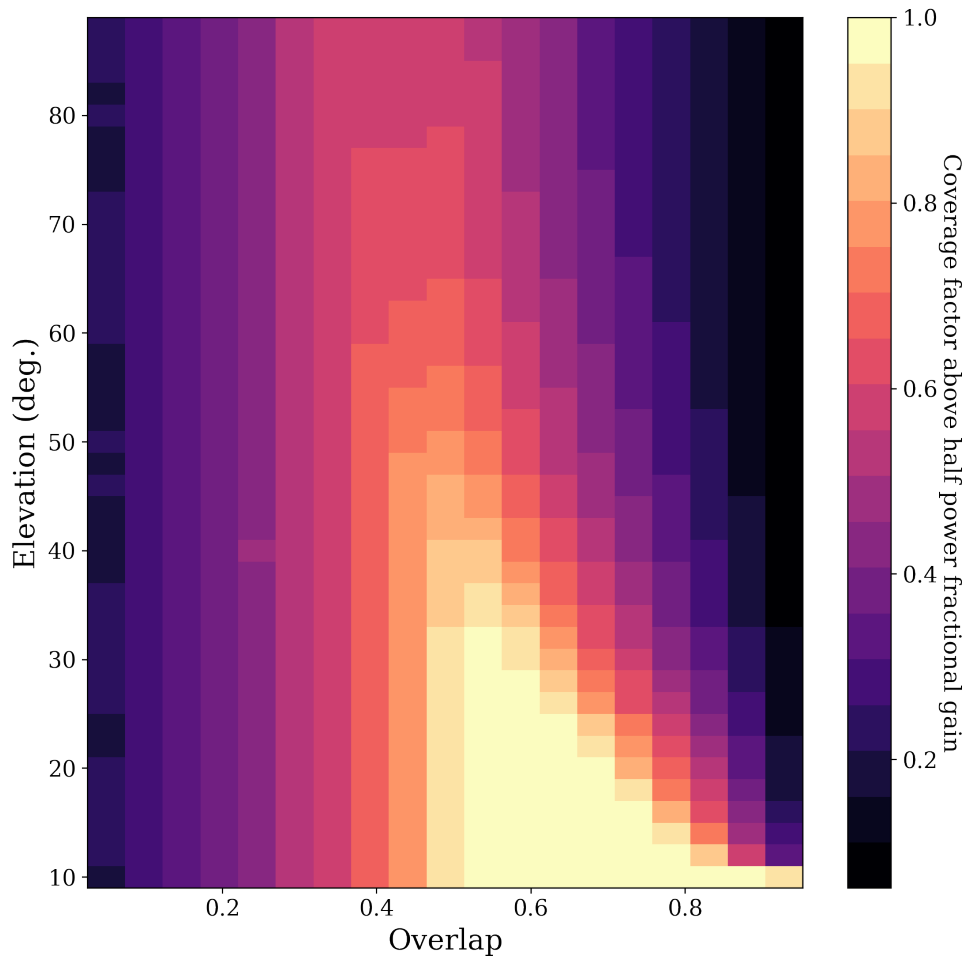


Figure 4.11: The simulated coverage for a minimum fractional gain of 0.5 (i.e. half power) as a function of the overlap factor between tied-array beams and the telescope elevation for a pilot survey pointing chosen at random. The coverage is dependent on the sensitivity profile within the beam tiling as well the area covered by the tied-array beams as a fraction of the survey beam area. Low overlap factors increase the spacing between the beams and in turn increase the overall coverage. However, they decrease the achievable sensitivity between beams, thus reducing the overall gain. High overlap factors decrease the coverage but provide a more uniform sensitivity between beams within the tiling. High and low elevation tends to shrink and elongate the size of the beam respectively.

- An S/N reduction factor was calculated for pulsars offset from the TAB positions. This was done assuming that the beamshape can be modelled as a 2D Gaussian profile. Thus, points lying along the perimeter of the model would have half the value of the expected S/N at the beam centre.
- The sky temperature was derived using the Haslam map (Haslam et al., 1982) deduction (as done previously for the survey sensitivity analysis; see Section 4.4)

Table 4.3: The MGPS-L pilot survey redetections of two known pulsars. The presented parameters are: J2000 name, spin period, dispersion measure, the name of the MGPS-L pointing where it was detected, The tied array beam width in degrees (w), The tied array beam height in degrees (h), The angular offset between the pulsar and beam positions, expected S/N (S/N exp.) calculated according to the radiometer equation, observed S/N (S/N obs.) and expected S/N error (S/N err.) which is estimated from the uncertainty on the expected flux density. The corresponding units are provided in the table.

PSR	P0 (s)	DM (pc cm ⁻³)	Pointing	w (deg.)	h (deg.)	offset (deg)	S/N exp.	S/N obs.
J0857-4424	0.326786	184.43	MGPS_L_0243	0.020	0.0074	0.0233	114.4 ± 5.6	91.9
J0855-4658	0.575072	472.70	MGPS_L_0331	0.0096	0.0079	0.00607	41.5 ± 4.8	40.4

at the beam position where the detection was made.

- The reported W50 parameter from the ATNF catalogue ([Manchester et al., 2005](#)) was used for the estimate of the intrinsic pulse width at FWHM.
- We derived the flux density by first fitting a power law to available flux density values from the ATNF catalogue. The estimate was then weighted by the observing bandwidth of interest (856-1712 MHz).

The detection of PSR J0857-4424 was not expected given that the pulsar was located in the void between three survey beam pointings (see Figure 4.8). However, one of the pointings (MGPS_L_0243) had a SBCF value exceeding 1 owing to a low telescope elevation (23.44 degrees). Hence, a TAB from the outermost part of the tiling was able to detect this pulsar (see Figure 4.12). As reported in Table 4.3, PSR J0857-4424 was detected with an observed S/N value which is 0.8 times the expected S/N value. The 20 per cent reduction is most likely a consequence of minor artefacts of the beam pattern outside the FWHM region. PSR J0855-4658 was detected in the third spin frequency harmonic and the observed S/N matches with the expected S/N within the error bar.

4.7.2.3 Known pulsars undetected/missed

Eight of the remaining nine pulsars in the two pilot survey fields (see Figures 4.8 and 4.9) were missed due to inadequate coverage. Some pulsars were missed as a consequence of not being enclosed by any survey beam pointing but are expected to be covered in future survey pointings (PSR J0932-5327, PSR J0940-5428 and PSR J0957-5432). Some pulsars were within the survey beam enclosure but the SBCF value was not high enough for the outermost TABs to overlap with the pulsar position (PSR J0934-5249, PSR J0901-4624, PSR J0905-4536 and PSR J0954-5430). One pulsar was missed as a consequence of being in the void between pointings (PSR J0855-4644).

One pulsar which was missed in the searches despite being covered by the TABs was PSR J0941-5244. The searches revealed no signal but folding the data with the

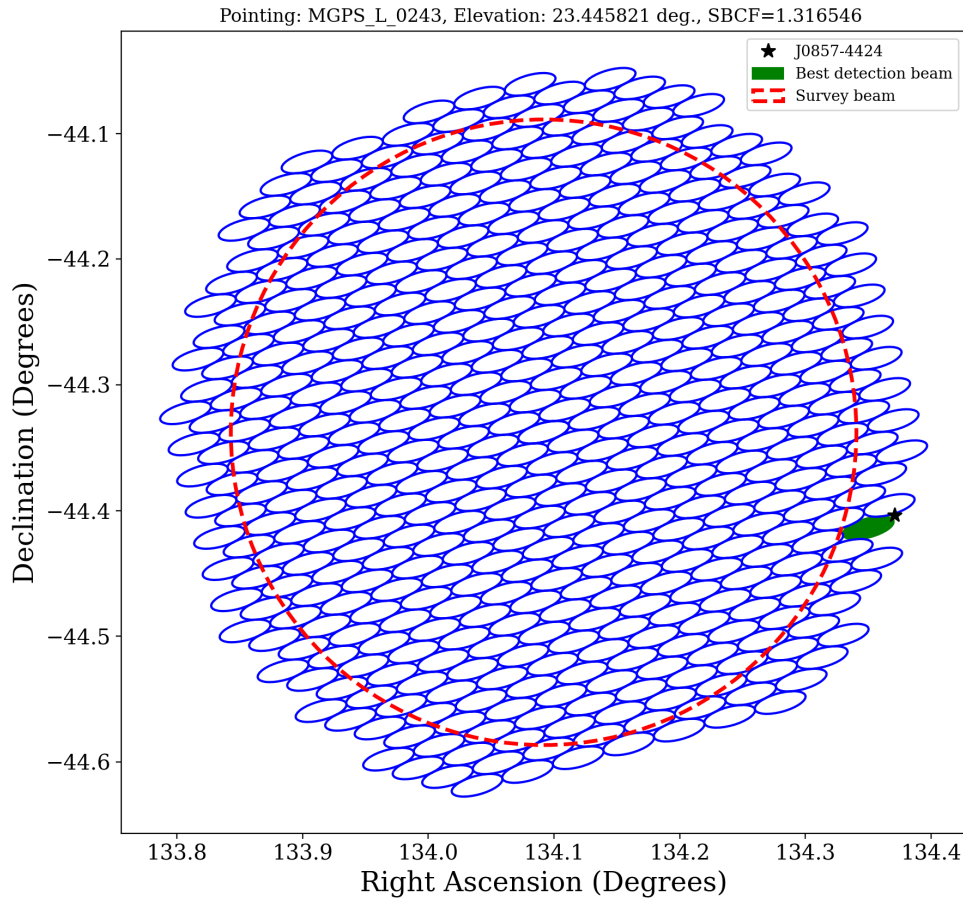


Figure 4.12: The tiling pattern of MGPS_L_0243 where PSR J0857-4424 was detected. The survey beam (corresponding to $\text{FWHM}/\sqrt{5}$ of the primary beam) is marked in red. The tied-array beams are overlaid (marked in blue) with the respect to the boresight position of the pointing. This pointing shows that the 480 beams exceed the boundary set by the survey beam thus leading to a survey beam coverage factor above 1.0 (see text). The pulsar position is marked as a black star and the beam with the best detection S/N is marked in green.

ephemeris parameters revealed a detection with significant portions of the observations showing RFI bursts. Thus, the most likely explanation is reduction in sensitivity due to RFI.

4.8 Conclusions and Future Prospects

We have conducted a pilot survey as part of the commensal science driven MPIfR Galactic Plane Survey. The survey covered 114 pointings of the expected 4140 pointings from the shallow L-Band sub-survey. A significant number of pulsars in the pilot survey were missed due to poor coverage of the TAB tiling. Future pointings can benefit from survey beam configurations where a few tied-array beams are placed on pulsar positions if within the survey beam radius. This would give more robust estimates of the sensitivity as the survey progresses. We demonstrated through simulations that the current survey setup is able to achieve a 78 per cent coverage factor for a minimum fractional gain of 0.5. The searches yielded yielded S/N values within the expected range derived from the radiometer equation for two known pulsars (PSR J0857-4424 and PSR J0855-4658). The searches also revealed that a long period pulsar PSR J0941-5244 was not detected by the search pipeline despite being located within the TAB coverage. This calls for improvements to RFI mitigation and red-noise removal schemes for future pointings.

Pulsar candidate classification using SAP Data Intelligence

The work presented in this chapter is part of a collaborative effort with the software company SAP. I am co-responsible for developing the pipeline along with colleagues from SAP. I am also responsible for the data reduction and results presented here. This work will be presented in an article submitted to *Astronomy and Computing*, in which I will be the lead author.

Contents

5.1 Abstract	149
5.2 Introduction	150
5.3 Machine Learning	151
5.3.1 Artificial Neural Networks	152
5.3.2 Common evaluation metrics	153
5.3.3 Feature Selection	155
5.3.4 Overview of ML classifiers for pulsar candidate classification . . .	157
5.4 Candidate processing schemes for current and next generation telescopes	158
5.5 SAP Data Intelligence	159
5.6 Data preparation	160
5.7 Processing	162
5.8 Preliminary results	162
5.9 Discussion and Future work	165
5.10 Conclusions	168

5.1 Abstract

Pulsar candidate classification schemes have predominantly involved human intervention and decision making to determine the likelihood of a new detection being a pulsar. However, the last decade has seen a surge in the use of machine learning (ML) classifiers for identifying radio pulsars among large survey datasets. With the advent of telescopes like the MeerKAT radio telescope, Five-Hundred-Metre Aperture Spherical telescope (FAST) and the upcoming Square Kilometre Array (SKA), a corresponding

enhancement in data processing is a must to keep pace with the expected data rates. While on-line data processing is crucial, it is equally important to have a platform that offers flexibility in applying models and metrics to the data and assists in offering solutions for the candidate classification problem. We describe a system for pulsar candidate classification using the SAP Data Intelligence platform. We use 3.2 million candidates generated from the High Time Resolution Universe (HTRU) South survey to demonstrate a use-case for the platform. We show that the SAP system offers a flexible and scalable system for storing pulsar candidate metadata and deploying ML classifiers that are already well tested or require testing on a large scale. We also compare the performance of existing classifier models on these candidates. Domain experts can also quickly investigate important candidate features through the data analytics tool-kits provided within the platform. Such a setup can provide vital feedback on the different classifier models. This would help better understand the circumstances and constraints under which certain models are better suited for improving the accuracy of pulsar candidate classification.

5.2 Introduction

The method of searching for new radio pulsar signals involves a series of techniques which account for intrinsic pulsar properties, propagation effects and instrumental aspects within the signal processing chain. Pulsars signals being broadband suffer from interstellar dispersion quantified by the DM (see Section 1.3.1). Candidate signals generated from search techniques (see Chapter 2) are used to phase coherently fold the data using the DM trial and the spin-period values to form candidate data-cubes (see Section 2.2.8). These data-cubes contain information along time, observing frequency, rotational phase as well as power of the signal. At this stage, a series of diagnostic plots are generated using different projections of the data-cube and the plots are visually inspected. Pulsar astronomers use these plots to make an informed decision whether the candidate is a true pulsar signal or not.

The number of pulsar candidates generated from pulsar surveys over the last few decades has increased by several orders of magnitude (see [Lyon et al. \(2016\)](#) for details on the numbers). The second Molonglo survey ([Manchester et al., 1978](#)) conducted 5 decades ago produced about 2500 candidates. Whereas, modern surveys like The HTRU North survey ([Barr et al., 2013](#)) have produced more than 80 million candidates. The reasons for these high numbers are several. A significant contributing factor has been the rise of sources of radio frequency interference (RFI) in the form of mobile phones, communication satellites and other electronic communication signals that produce narrowband as well as broadband signatures in the radio spectrum. These signals show up as false positive candidates in the pulsar search algorithms, thus increasing the candidate numbers. Another major contributor to the increased number of candidates is the increased number of trials. This is a consequence of high time and frequency resolution as well as the usage of binary pulsar search techniques that add extra dimensions like acceleration and jerk parameters apart from the DM and spin

period in the search parameter space. Despite these refinements in searching, the number of true pulsar candidates (including known pulsars and new pulsars) as a fraction of the total candidates can be as low as 0.01% (Lyon, 2016).

Given that the candidate numbers are only bound to increase, applying person hours for manual pulsar candidate classification is impractical. This has led to the development of automated techniques which enhance the processing speed without compromising the accuracy of the classification. Although certain automated ranking procedures with human scoring schemes have been used in the past (Faulkner et al., 2004; Keith et al., 2009; Lee et al., 2013), machine learning techniques have emerged as the prime solution for classifying pulsar candidates (e.g. Eatough et al., 2010) (detailed overview given in Section 5.3.4).

5.3 Machine Learning

Machine Learning (ML) is an active branch of computer science involving algorithms that learn and improve with experience. In the traditional scenario of problem solving, humans specify a series of steps required by a computer to model the input data and solve a given problem. This straight-forward approach is not always applicable while building a model and it is preferable that the computer learns to make decisions and infer the necessary steps required to achieve a given task. ML can be categorised into two broad categories based on the level of human inputs required in the process of learning by experience. These are :

- **Supervised Learning:** The computer is provided with a set of human generated classifications corresponding to the input data. These labels serve as a guideline in the training phase where the model learns and adapts to the expected outcomes from the input data. The model generated from training is validated in the testing phase on data different from the previously input data. This process ensures that the model is robust and not biased by the training data. Examples of classifiers built around supervised learning include Artificial Neural Networks (see Section 5.3.1), Support Vector Machines (Hearst, 1998) and Deep Learning algorithms (LeCun et al., 2015). In the context of pulsar candidate classification, the training data would consist of labels denoting if the input candidate data cube is a pulsar or not. The model built by the learner would then be aimed at accurately classifying any candidate as a pulsar or non-pulsar in an automated fashion.
- **Unsupervised Learning:** In this scenario, no labels are provided and the algorithm is expected to independently determine patterns and structures within the input data. Some examples in this category include k-means clustering (Macqueen, 1967) which aims to organise the data into smaller groups and anomaly detection where outlier points are identified from the statistical distribution that best describes the input data. For the pulsar candidate classification problem, the learning algorithm is not provided with any candidate metadata information and

is expected to find patterns and group data into categories segregating pulsars from non-pulsar data.

From the point of view of pulsar candidate classification, supervised learning is a convenient choice. This is mainly because the domain knowledge about pulsars can be used to help build better models. However, the task of generating a large number of human classified labels is laborious. This has recently encouraged the usage of semi-supervised learning which is a mixture of supervised and unsupervised techniques (Balakrishnan et al., 2020). In the case of unsupervised learning, algorithms like anomaly detection would be suitable in specialised circumstances like identifying spurious candidates from a large candidate dataset. Given that the spurious candidates are outliers in the candidate distribution, they could for example be produced due to a particular type of interference source or a specific type of astronomical signal. Since a wide variety of pulsars as well as interference sources exist in general, unsupervised learners do not add value to the problem at hand. These factors are reflected in the literature where majority of the classifiers built for pulsar candidate classification use supervised learning algorithms (see Section 5.3.4).

5.3.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) belong to the set of supervised machine-learning algorithms and are a common choice for classification problems. The term neural network is inspired from the functioning of biological neuron cells that are activated under certain conditions. The network consists of multiple nodes (termed as neurons) which are grouped in different layers. There are a minimum of two in this construct comprising of an input layer where the data interest enters the network and an output layer that produces a result for the given input. In context with pulsar candidate classification, the input nodes represent unique features of the pulsar candidate and the output has 2 nodes representing the classifications, in this case: pulsar or not a pulsar. The most simplistic network is the single layer perceptron (see Figure 5.1) with no hidden layers. The neural network can also be expanded to have multiple hidden layers which is termed as a deep neural network (see Figure 5.1). All nodes belonging to the same layer are connected to the same set of nodes from the previous layer (each connection is termed as a synapse). However each synapse is weighted differently. Mathematically speaking, the weighted sum of inputs arriving at a layer node can be expressed as

$$\hat{O} = f \left(\sum_{i=1}^m X_i w_i + Y \right) \quad (5.1)$$

where w_i is the weight on the synapse from each of the i nodes in the previous layer and Y is a bias factor that is constant for a particular layer. f denotes the activation function which eventually determines whether a particular node is activated for a given set of weights. Some common activation functional forms used are the sigmoid function and the Rectified Linear Unit (ReLU) function. Once the layers

and the activation function are configured, the learning process for a neural network begins by minimising a loss function. The loss function is the difference between the actual labels and the predicted labels for a given input data. Depending on the type of problem, the loss function can be minimised in different ways. The standard approach is using back-propagation. Here, the gradient of the loss function with respect to the weights is calculated over every input/output example in an iterative manner. The network adjusts its weights over every iteration and converges to one combination of weights. The training phase is complete here and in order to check how robust the model is, a series of metrics are used.

In order to ensure that the training process was successful, a different data set is chosen for cross-validation. The performance of the model on this set is then evaluated using metrics that are discussed in the next section.

5.3.2 Common evaluation metrics

In context with pulsar classification, there are 4 different possibilities that the learning model produces:

- i) True Positives (TP): A true pulsar candidate that is correctly classified.
- ii) True Negatives (TN): A true non-pulsar candidate that is correctly classified.
- iii) False Positives (FP): A true non-pulsar candidate is predicted to be a pulsar.
- iv) False Negatives (FN): A true pulsar candidate is predicted to be a non-pulsar.

These four outcomes are collectively defined as the confusion matrix (see Figure 5.2). Using each of these outcomes, some of the standard metrics and the corresponding formulae are defined below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5.2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5.3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.4)$$

$$\text{F1} = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.5)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (5.6)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.7)$$

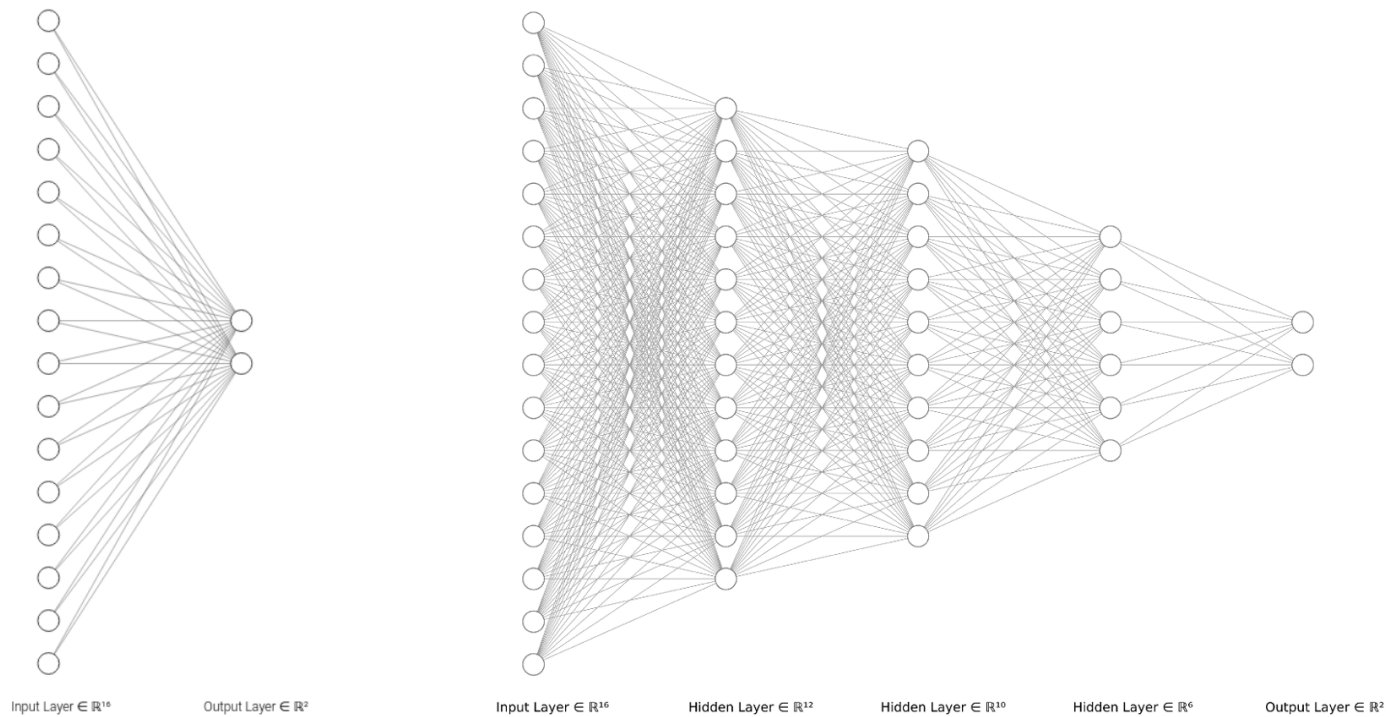


Figure 5.1: Left panel: Representation of a single layer perceptron with eight input features and 2 output features. Right panel: Representation of a multi-layer artificial neural network with three hidden layers. Each hidden layer contains nodes that are fired based on inputs from the previous layer. While, the complexity of the learning process increases with more layers, the network also tends to learn better. This figure was generated using the NN-SVG online tool (<https://alexlenail.me/NN-SVG/>).

While evaluating a pulsar candidate classifier, each metric plays a significant role in giving feedback about the classifier model. A high accuracy value is a plus point assuming that the training data is class balanced. This means that the training data should roughly have the same number of input labelled pulsar candidates as non-pulsar candidates¹. A high precision value implies a low value of False Positives. This indicates that the number of non-pulsar candidates are low which in-turn reduces the overall candidate inspection time. A high recall is a higher priority than precision since pulsar astronomers would prefer to not mistake any true pulsar candidate as a false negative. The F1 score is an amalgamation of Precision and Recall and is a robust indicator of the overall efficiency of the model. FPR stands for False Positive Rate and is on similar lines with Precision but gives the proportion of false positives from all the classified labels. A low score is desirable for the FPR metric while testing the robustness of a classifier. Similarly, TPR (True Positive Rate) measures the proportion of true positives and a high score on this front is ideal for classification.

5.3.3 Feature Selection

An important step prior to training is to select inputs that best represent the data. While a large number of such input variables could collectively represent the data better, the training process would take a long time to converge to an appropriate selection. For the pulsar search problem, the aim is to use a set of reduced parameters (termed as features) that provide the most useful information while distinguishing between pulsar, RFI and noise candidates.

Pulsar astronomers typically handcraft features derived from domain knowledge. As the neural network is exposed to these features for several candidates in the training phase (with the corresponding set of true labels), the predictive capability of the model is expected to improve. A few common attributes that have been used for feature selection are listed below:

- *Dispersion Measure*: In contrast to RFI, pulsar signals travel through the interstellar medium and are thus subject to dispersive effects. Hence, the DM value has been used as a feature (e.g. [Bates et al., 2012](#)). Other classifiers have used the DM value as a starting point to derive different metrics. For example, [Eatough et al. \(2010\)](#) used a fit made to the DM vs Signal-to-Noise Ratio (SNR) curve. The curve would peak at the true DM of the pulsar whereas RFI would peak at DM=0. Building on this, [Lyon et al. \(2016\)](#) used the mean, standard deviation, kurtosis and skewness of the DM-SNR curve.
- *Persistence in time*: Majority of pulsars show continued emission of pulses with time. An example of a feature based on this attribute is the persistence measure used by ([Morello et al., 2014](#)). They use a metric that is scored over each time

¹Class imbalance is a common problem with pulsar data given that majority of the candidates generated from a survey are interference and noise. Solutions to this problem are discussed in Section 5.3.4

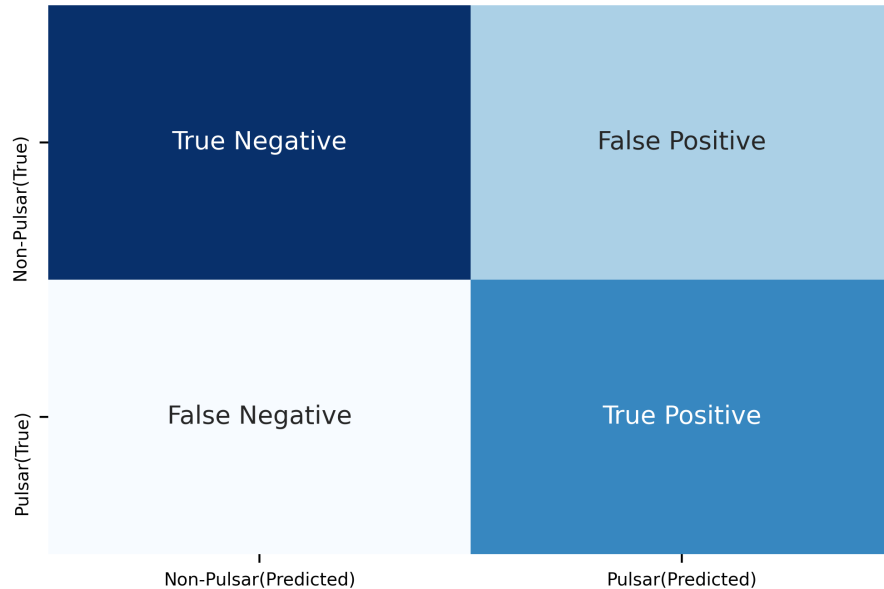


Figure 5.2: A visual representation of the confusion matrix in terms of true and predicted labels for pulsar candidate classification.

sub-integration and finally averaged. Thus a higher score indicates a higher degree of persistence.

- *Persistence in Frequency*: Pulsars emit broadband signals thus leading to significant emission throughout the observable bandwidth. This attribute also helps in filtering out narrowband RFI candidates whose power is mainly concentrated on a small range of observing frequencies. An example is the linear correlation score across frequency sub-bands used by (Eatough et al., 2010). Similarly, (Zhu et al., 2014) used the information from the two dimensional plot of frequency vs phase as an input feature.
- *Pulse profile*: The profiles of pulsars typically tend to show narrow peaks (average duty cycle ~ 5 per cent; (Manchester et al., 2005)). Classifiers have thus used quantities derived from the standard intensity pulse profile as a feature. For example, (Lyon et al., 2016) used the mean, standard deviation, kurtosis and skewness of the integrated pulse profile as features.

A major caveat is that the above mentioned guidelines have exceptions. For example, there exist RFI sources that are broadband in nature, are persistent with time and can also mimic the shape from average profiles of pulsars. From the perspective of pulsars, there exist pulsars which do not show persistent emission in time. This includes nulling pulsars (see Section 1.2.8), intermittent pulsars (see Section 1.2.8) and

eclipsing binaries (see Section 3.3). Millisecond pulsars tend to have broad peaks with duty cycles up to 50 per cent (Morello et al., 2014).

In reality, feature selection for pulsar candidate classification is a complex problem. A finite set of features used to train a model could overlook a true pulsar candidate that is an outlier in the feature hyper-parameter space. Apart from computational costs, increasing the number of input features introduces problems related to the Hughes phenomenon (Hughes, 1968) which states that the predictive power deteriorates beyond a certain feature dimensionality. Hence, the scope for choosing a robust set of pulsar candidate features is an open question.

5.3.4 Overview of ML classifiers for pulsar candidate classification

A vast set of supervised ML techniques using ANNs have been implemented for pulsar candidate classification in the last decade. The ANN is input with labelled hand-crafted candidate features (i.e. features made from domain knowledge) from folded data-cube archive files generated from a particular survey. The labels generally indicate whether a candidate is a pulsar, RFI or noise. (Eatough et al., 2010) developed an ANN that utilised 12 features with a multi-layer perceptron model and discovered a previously missed pulsar in the Parkes Multi-beam Pulsar Survey (Manchester et al., 2001). Bates et al. (2012) developed a similar network for the HTRU South survey (Keith et al., 2010) and correctly classified 85 per cent of the known pulsar discoveries. This drove the need for improved strategies and models which are able to retrieve majority of true pulsar detections among a large number of candidates. Zhu et al. (2014) developed an ensemble classifier termed PICS using Convolutional Neural Networks (CNN) along with Support Vector Machines with 2D features like the Frequency-Phase and Time-Phase information as inputs. These were combined with 1D features like the pulse profile and the DM-curve input into an ANN. The training was done on ~ 4000 candidates from the PALFA survey (Cordes et al., 2006) and validated with candidates from the GBNCC survey (Stovall et al., 2014). The classifier managed to rank all pulsars in the top 1 percent of 90,000 test candidates and was able to identify six previously undiscovered pulsars. Morello et al. (2014) developed an ANN classifier that managed to rank 99% of the pulsar candidates within the top 0.11% of 91,200 candidates (1200 pulsar candidates + 90,000 non pulsar candidates) generated from the HTRU South Survey (Keith et al., 2010). The classifier also provided a significant enhancement in processing speed by reducing the amount of candidates to be visually inspected by four orders of magnitude. (Bethapudi & Desai, 2018) used the same training set as used by Morello et al. (2014) to evaluate the ANNs as well as other ensemble classifiers like Adaboost and Gradient Boosting Classifiers. Their results showed a high degree of accuracy (99 %) for all types of classifiers and they also managed to reduce the false positive rate by an order of magnitude compared to results obtained by Morello et al. (2014). They also applied a Synthetic Minority Over-sampling Technique (SMOTE), where synthetic data were manufactured to compensate for the class imbalance between pulsar to non-pulsars candidates (which had a highly skewed ratio of 1:100).

The class imbalance problem has been brought up as a matter of concern before (e.g. Lyon et al., 2013) and multiple other solutions have been provided for the same. Lyon et al. (2014) proposed the Hellinger distance measure as a means to deal with imbalanced data streams. They tested the effectiveness of the algorithm with data from the HTRU South survey (Keith et al., 2010). They were able to demonstrate a significant improvement in recall rates at the cost of a reasonable increase in the false positive rate. Guo et al. (2017) tackled the class imbalance problem by using deep convolution general adversarial networks (DCGAN) to generate samples and enable feature learning. Xiao et al. (2019) proposed a pseudo-nearest centroid neighbour classifier which runs after evaluating the data for class imbalance using a scatter matrix based metric.

Besides some of the above mentioned works, many other techniques different from standard ANN architectures have been explored for classifying pulsar candidates. Lyon et al. (2016) adapted the Hellinger distance measure into an online learning scheme to address real-time classification as a feasible solution for large incoming data rates. They showed that this new algorithm termed as the Gaussian Hellinger Very Fast Decision Tree (GH-VFDT) along with a new set of features was able to process ~ 1 millions of candidates in 15 seconds with a reasonably high recall rate (90 %). This procedure has been incorporated into the LOTAAS survey (Sanidas et al., 2019) and has been responsible for multiple pulsar discoveries. Wang et al. (2019) modified the PICS classifier developed by Zhu et al. (2014) to incorporate a 15 layer residual networks model in place of CNN. This classifier has been incorporated into the CRAFTS survey being conducted with FAST. Recently, Balakrishnan et al. (2020) have explored a semi-supervised learning classifier that is also able to deal with a mix of few labelled and largely unlabelled data sets without compromising on the accuracy of classification. This scheme could prove useful in the early stages of a survey where the candidate numbers are still low to build a suitable training set.

5.4 Candidate processing schemes for current and next generation telescopes

Modern day telescopes equipped with capabilities including multi-beam beamforming and large bandwidth receivers produce high fidelity data, thus enhancing the data inflow and outflow rates. As discussed earlier in Section 5.2, these high data rates contribute to massive candidate data volumes generated from pulsar search pipelines. The performance of the latest pulsar search pipelines are ever increasing via the use of Graphics Processing Units (e.g. Dimoudi et al., 2018) and Field Programmable Gate Arrays (FPGAs) (e.g. Wang & Sinnen, 2015). In contrast, the subsequent candidate filtering routines have improved algorithm wise but not performance speed wise. However, the predicted candidate numbers from current and future surveys raise the importance of moving to strategies centered on online data streaming. For example, the Galactic plane survey with the MeerKAT telescope discussed in Chapter 4, is expected to generate up to 0.3 billion candidates (~ 0.2 PB) in total. The SKA1-Mid which is a subset

of the SKA is expected to produce 1000 candidates per beamformed data stream (Levin et al., 2018) and up to 2500 candidates per second amounting to 3.3 TB for a single observation (Lyon et al., 2019).

Lyon et al. (2019) conducted a detailed analysis describing the feasibility of online processing for pulsar candidate data streams. They divided the pulsar candidate problem into multiple components including candidate sifting, feature extraction, known source matching and candidate selection. They discussed the inner workings of each component at a computation level and provide solutions to help migrate to the online platform. They also provided a prototype design for an end-to-end candidate filtering pipeline using the Apache Storm framework. This design was bench-marked at a small scale on a single quad-core machine as well as at a large scale via the Amazon Web Services Elastic Compute Cloud infrastructure. The results at both scales showed a high computational efficiency with the pipeline handling more than 6000 candidates generated per second. These rates match the SKA requirement scales thus proving to be a promising approach for the future. We refer to Lyon et al. (2019) and references therein for specific details regarding the proof of concept of the design.

While the prototype design was able to achieve desirable processing rates, the pulsar recall rate as reported by Lyon et al. (2019) did not exceed 88 percent. They argued that the dip in recall rate was a consequence of prioritising speed over accuracy by using a stream classifier. They reported that a further improvement in recall rate would require sophisticated schemes and would come at a cost of computation time. However, they also suggested that there is scope for building complex algorithms that enables high recall rates without a heavy compromise on computation.

In the following section, we describe a system where pulsar astronomers can populate a large number of candidates (order of millions to billions). Furthermore, the system also provides resources and tools to test and benchmark different pulsar candidate classification algorithms on a grand scale. This way, the robustness of injected algorithms can be better understood. The algorithms can then be adapted and fine-tuned for improved performance based on evaluation metrics. Such a large scale platform also allows pulsar astronomers to develop new evaluation metrics as well as input features that can simplify the candidate filtering process.

5.5 SAP Data Intelligence

SAP Data Intelligence is an all-in-one orchestration solution to analyse, refine and visualise data across a distributed data landscape. The platform is built by SAP, a multinational software company primarily known for enterprise resource planning (ERP) software. Based on container technologies including Docker and Kubernetes, SAP Data Intelligence can be used to define data flows (aka pipelines) of arbitrary complexity connecting multiple data sources to multiple targets. These flows are directed acyclic graphs made of interconnected operators. The function of an operator is to ingest a certain data type and output a relevant data product based on a predefined set of instructions. Alongside the standard operators delivered with the product, users

can define and use custom operators by providing the corresponding `Docker` images to launch containers. The user is also provided with an interface to drag and drop customised operators built on a specific algorithm or library. The interface serves as an abstraction layer to run the required process chain. The workload can be scaled dynamically and uniformly across the cloud infrastructure offering an easy way to scale-out and accelerate the execution of the algorithms with negligible overhead. Owing to Kubernetes, hybrid clusters can be defined with a mix of CPU and GPU hardware, distributed in different data centers which can be geographically far apart. This offers optimization possibilities for data locality and associated hardware.

5.6 Data preparation

In order to test the SAP platform, we used candidates generated from a re-processing of observations from the HTRU South low-lat Survey which is a subset of the HTRU South survey (Keith et al., 2010). A partially coherent segmented search pipeline (as described in Ng et al. (2015)) was applied on 284 pointings from the survey chosen at random. This pipeline was chosen given that it has been successful in finding more than 100 new pulsars (Ng et al., 2015; Cameron et al., 2020) and redetected more than 800 previously known pulsars. The candidates generated from the pipeline are in AR2 format which is a variant of PSRFITS archive format. About 3.2 million candidates from this pipeline were chosen for further analysis.

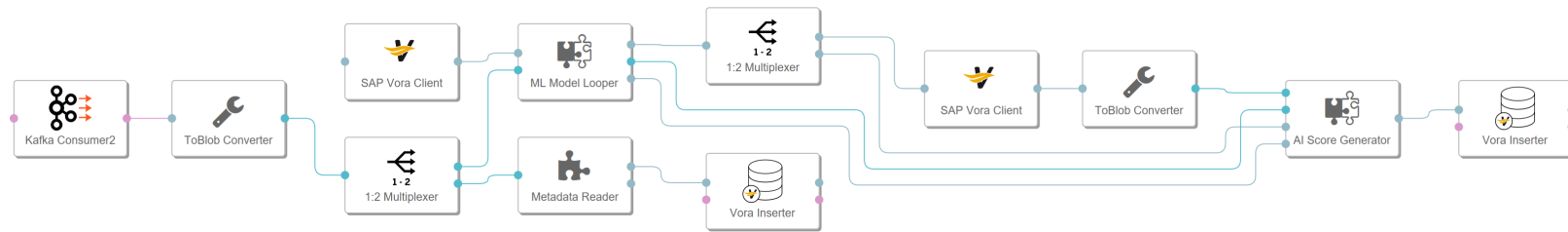


Figure 5.3: Schematic of the pipeline used on SAP Data Intelligence for testing ML models on pulsar candidate data. Each block in the schematic represents an operator with a specific functionality.

5.7 Processing

A schematic of the SAP Data Intelligence (SDI) pipeline employed on the candidates is shown in Figure 5.3. The candidates generated from the search pipeline are originally stored in the MPIfR’s HERCULES computing cluster located at the Max Planck Computing & Data Facility (MPCDF). These candidates are streamed to the SAP Data Intelligence instance via an Apache Kafka service. Apache Kafka is an open-source distributed event streaming platform that is commonly used by companies for high performance message passing applications. Each pulsar candidate is converted into a message which is published to a unique Kafka channel known as a topic. A machine connected to the SAP Data Intelligence cloud subscribes to this channel and receives the candidate blobs (this is done by the ‘Kafka Consumer2’ module in Figure 5.3). Each candidate blob is sent through two different channels enabled by a multiplexer. One channel reads in the candidate blob and extracts the necessary candidate metadata from the blob. These metadata are stored on a scalable distributed database system known as SAP Vora. The other channel is scored through a loop of ML models that are already stored in the SAP Vora database in the form of Python pickle files. In this case, we use 8 different models that were generated from the PICS classifier (Zhu et al., 2014) and were made publicly available². These 8 models were generated from training a combination of candidates from different surveys including PALFA (Cordes et al., 2006) and HTRU (Barr et al., 2013; Keith et al., 2010) surveys as well as different sets of candidates within the HTRU survey. A brief description of each model is given in Table 5.1. PICS reads in the candidate blobs and applies each of the models to generate a score between 0 and 1 where 1 indicates a most probable pulsar detection and 0 implies a strong likelihood of the candidate not being a pulsar. The scores generated per PICS model are combined with the metadata information retrieved via the other channel. SQL Queries can be run on the SAP Vora database to retrieve all necessary information required by the user pertaining to every candidate put through the system. The pipeline is built in such a way that any user can test their algorithms by just supplying a model file generated from training. This model file can be quickly integrated into the pipeline and used to generate scores that would provide a feedback on the robustness of the model.

5.8 Preliminary results

As a first step, we bench-marked the incoming rate of candidates into the SAP system. The AR2 candidates were 300 KB in size on average and the streaming rate was 0.3-0.5 seconds per candidate depending on the exact size of the candidate blob that was sent through. In order to test the scalability of the system, the pipeline was scaled to process more than one candidate blob in parallel. We were able to score all 8 models in parallel. The pipeline was also scaled to spawn 60 parallel scoring pipelines. The average rate of scoring candidates through the scaled pipeline setup was 6 candidates

²https://github.com/zhuww/ubc_AI

Table 5.1: List of different trained models based on the PICS classifier (Zhu et al., 2014) used for the analysis with SAP Data Intelligence. Each model uses four input features: time vs phase, frequency vs phase, pulse profile and the DM-curve as obtained from a folded archive file. The models are generated with different conditions that are explained in the second column. The model naming convention is explained as follows: ‘clf’ stands for classifier; ‘l2’ stands for L2 regularization and a one-layer neural net to combine the scores of sub-classifiers for different features; ‘svm’ stands for usage of support vector machines to combine the scores from sub-classifiers. The HTRU data were segregated into different training sets, thus leading to three unique models (denoted by 0,1,2). More details on the input features and the training process can be found in Zhu et al. (2014).

Model	Characteristic
clf2_PALFA	trained with data from PALFA survey
clf2_BD	trained with data containing known pulsars with a broad pulse (duty cycle > 40 per cent)
clf2_FL	trained via feature labelling.
clf2_HTRU_0	trained with data from the HTRU South low-lat survey
clf2_HTRU_1	trained with data from the HTRU South low-lat survey
clf2_HTRU_2	trained with data from the HTRU South low-lat survey
clf2_SP_0.0	trained for searching single pulses
clfsvm_HTRU_2	trained with data from HTRU South low-lat survey

per second.

We tested the efficiency of the models by selecting the known pulsar candidates from the entire candidate dataset. Based on the candidate metadata and scores available on SAP Vora, we first retrieved candidate information that best matched with known pulsar parameters. This was done by cross-matching between the SAP Vora database and a table of known pulsar detections available on the HTRU Survey website³. This table however contains information regarding known pulsars discovered prior to the survey. Thus, we also included a table of information about the discoveries resulting from the survey as reported by Ng et al. (2015) and Cameron et al. (2020). The first criterion used through the cross-matching process between the Vora database and each of the known pulsar tables were matching the pointings and beams. Given that only a fraction of pointings (284 out of 1200) were put through the SDI pipeline, we filtered out the known pulsar information that was irrelevant to the pointings of interest. The second criterion used was a match between the spin period as well as DM as reported in the candidate metadata with the spin period and DM parameters of the known pulsars with a 5 percent tolerance for each. The scores generated from the models for the cross-matched candidates were also retrieved⁴. A summary of the results obtained from the cross-matching process is given below:

- A total of 60 unique pulsar detections were retrieved from cross-matching the table from the HTRU website. Two pulsar candidates matched with the discoveries reported in Ng et al. (2015). Similarly, there were 14 pulsars that matched with discoveries reported in Cameron et al. (2020). These numbers matched with the expected number of unique pulsar detections from the selected pointings.
- The scores of all pulsar redetections were compared between multiple PICS models using a cumulative histogram as shown in Figure 5.4. The plot shows that the data trained on the PALFA data set seems to be the best performing model. The HTRU models show relatively consistent scores. On the other hand, the SP model which is primarily trained on single pulse data shows poor performance. This is not surprising given that the persistence in time for single pulse candidates is much smaller compared to standard pulsar candidates. This biases the Time vs Phase plot (used as a feature in training PICS) to score low and hence reduce the cumulative score.
- In order to further test the robustness of the classifiers, we plotted a receiver operating characteristic (ROC) curve for each of the classifiers as shown in Figure 5.5. This curve has been obtained by calculating the TPR (see equation 5.7) and FPR (see equation 5.6) at different PICS score thresholds ranging from 0 to

³<https://sites.google.com/site/htrusouthdeep/home/redetections?fbclid=IwAR0BCaKspWiGDufjwUVkSX3mQ1Yoj22Yvo1QflqrkP0v1UQmJdvGclH9nic>

⁴Due to a technical error, all candidates were not scored through each of the 8 models. All candidates were scored through one model but less than 33 per cent of candidates (approx. 1.2 million candidates) were scored through each of the other models. Thus the analysis regarding the scoring is limited in this work.

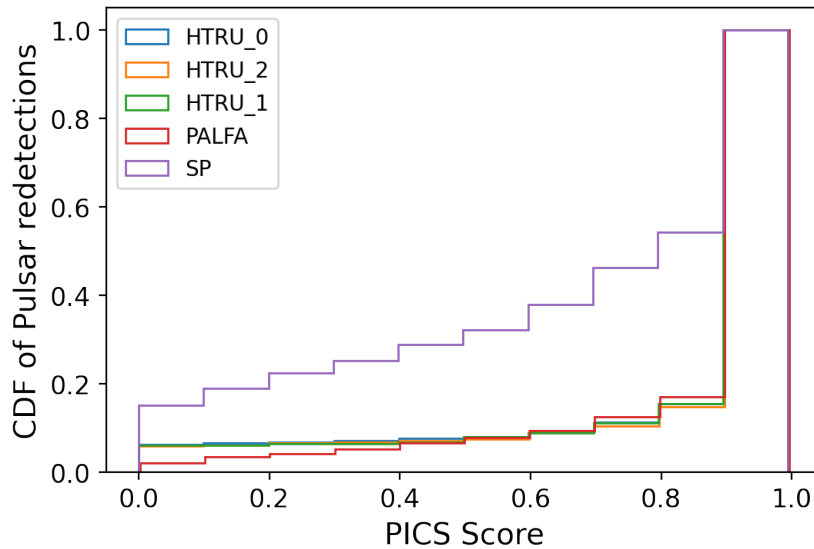


Figure 5.4: A cumulative histogram of the scores generated for 784 known pulsar redetections and five different PICS models each used by the SAP pipeline. Multiple pulsars are detected more than once in this set. The PICS models use different sets of training data from different surveys. The HTRU_0, HTRU_1 and HTRU_2 models have been generated with data sets from the same survey (i.e. HTRU) but without an overlap in candidates used for training. The PALFA model was trained on data from the PALFA survey (Cordes et al., 2006). The SP model was generated from training sets containing single pulse candidates from PALFA.

1. In order to calculate FPR, we assumed the number of pulsar detections to be insignificant in comparison to the total number of candidates (hundreds of pulsar detections compared to 3.2 million candidates in total). Thus the FPR was directly measured assuming that all candidates are non-pulsars. As plotted in Figure 5.5, an ideal classifier would yield a singular point in the ROC space where each candidate classification is accurate. The PALFA model once again proves to be the most robust classifier of all the models, given its proximity to the ideal classifier and also demonstrated by Table 5.2 which gives the area under the curve (AUC) for each classifier. The AUC value is a direct indicator of the measure of separability for a binary classifier and thus a high value is desired. We also observe that using L2 regularisation in comparison with the supporting vector model for the same HTRU training data (HTRU_2) does not significantly change the performance of the corresponding classifiers.

5.9 Discussion and Future work

The flexibility and scalability offered by the SAP system provides scope for extensive tests to be conducted at various levels. A large scale transfer is underway where about

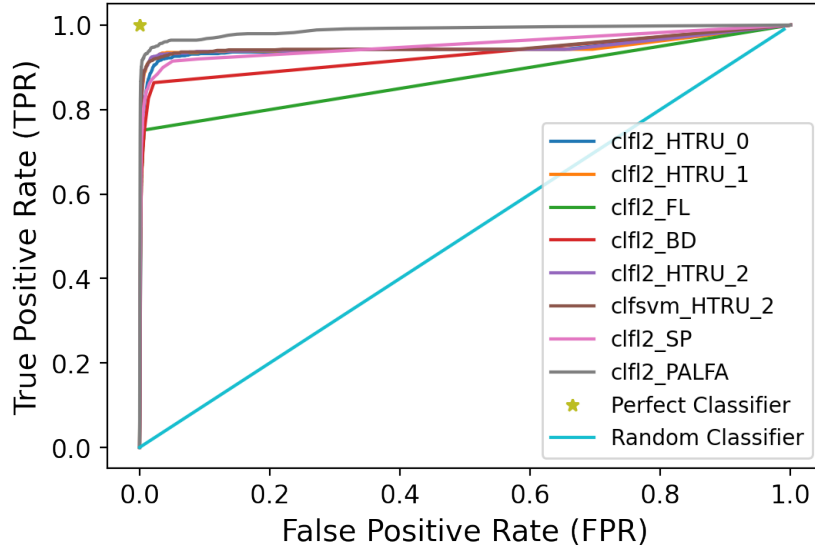


Figure 5.5: A receiver operating characteristic (ROC) curve for each of the PICS classifiers is plotted. The ideal classifier would always predict the true label and hence have a TPR of 1.0 and FPR of 0.0 as shown. A random classifier would predict each label (i.e. pulsar or non-pulsar) with equal probability of 50 percent. Thus the robust classifiers should feature in the ROC space above the diagonal given by the random classifier. While all classifiers are not random, the PALFA classifier performs better than the rest given that the curve is closest to showing an ideal performance.

Table 5.2: List of different trained models based on the PICS classifier (Zhu et al., 2014) used for the analysis with SAP Data Intelligence and their corresponding Area under the curve (AUC) values for the ROC curves plotted in Figure 5.5.

Model	AUC
clf2_PALFA	0.988
clf2_BD	0.927
clf2_FL	0.873
clf2_HTRU_0	0.947
clf2_HTRU_1	0.948
clf2_HTRU_2	0.949
clf2_SP_0.0	0.952
clfsvm_HTRU_2	0.951

16 million candidates generated from other reprocessings of the HTRU South low-lat survey would be transferred to the SAP system and scored through the various PICS models. Such a large and varied dataset would help better understand the advantages and drawbacks of the PICS models employed and tweak them to improve their robustness. Furthermore, a balanced training dataset can be prepared for retraining the existing models. Apart from PICS, we also plan to expand on integrating other open-source classifiers to the system. Such an expansion would help compare ML models against one another and identify which models perform better under certain circumstances. The pipeline bench-marking results showed that the processing rate is 0.1% the rate achieved by [Lyon et al. \(2019\)](#) (6 candidates per second scored vs 6000 candidates generated per second). Assuming that 3 candidates could be streamed per second at best with the current infrastructure, the entire set of candidates generated from SKA would take hundreds of years to transfer. This calls for developing a large scale parallelised uninterrupted streaming service along with improvements to increase candidate transfer rates. Another alternative to solving the streaming overhead is to integrate the SKA hardware with a cloud infrastructure through which candidates are automatically processed. Further investment in acquiring the right resources and technical know-how would be crucial to meet the high demanding data rates with SKA.

The analysis on the current models revealed that the PALFA model is the preferred classifier. This matches expectations given that the training data prepared for this model contained different types of pulsars including harmonics with large duty cycles (above 40 per cent) as well as narrow pulse slow pulsars ($P > 500$ ms) ([Weiwei Zhu, private communication](#)). Diversifying the training data is beneficial to ensure that the models generated are not biased to identify specific types of pulsars alone. Given that some models (e.g. `clf2_FL` and `clf2_BD`) were trained with candidates that focus on certain similar properties (see [Table 5.1](#)), they perform poorly in comparison to the PALFA classifier. Besides analyses on existing models, the SAP platform also serves as a test bed for developing new algorithms. SAP Data Intelligence comes with a set of user-friendly data visualisation and machine learning tool-kits. The aim in the future is to use these tools to help visualise different combinations of the candidate metadata. This can help identify unique features not only in the pulsar hyper-parameter space but also better characterise interference sources. With the upgrades in wireless technology (fifth generation i.e. 5G communication in particular), the impact on pulsar astronomy would be significant and non-negotiable. New classifiers that can adapt to these changes are thus a must. The aim in this process is to build a telescope agnostic super classifier which takes the best of existing ML algorithms and at the same time is adaptable (similar to GH-FVDT from [Lyon et al. \(2016\)](#)) without compromising on pulsar recall rates. An important part of this process is selecting a limited number of features that can improve candidate sifting schemes. Input features constructed so far have mainly been adapted from human diagnostic measures as explained in [Section 5.3.3](#). The SAP platform offers scope to employ different pattern recognition and feature selection metrics that have not found applications in pulsar candidate classification before.

Ultimately, the goal through this setup is to provide a platform for pulsar astronomers to compare different datasets across several telescopes and determine the

best combination of classifiers for different circumstances ensuring high performance in terms of speed as well as accuracy. A plausible outcome of such a setup is the potential to not only discover pulsars in archival datasets but also new astrophysical sources that were missed from previous analyses.

5.10 Conclusions

We demonstrated a use-case of the SAP Data Intelligence platform for the long standing problem of efficient pulsar candidate classification. We built a streamlined candidate metadata extraction and model scoring pipeline using operators within the SAP Data Intelligence platform. We integrated multiple models of an existing open source ML classifier (Zhu et al., 2014) to test its robustness. We showed that the system is scalable by processing upto 60 candidates in parallel for 8 different models. We demonstrated that the model trained on data from the PALFA survey (Cordes et al., 2006) is the best model suitable for pulsar candidate classification over a large dataset containing millions of candidates from the HTRU-South Low lat survey (Keith et al., 2010). We aim to further improve this platform by enhancing the speed of streaming and also provide a test bed for comparing other existing algorithms as well as develop new models that are able to overcome the complexities arising from current and future pulsar surveys.

Revisiting profile instability of PSR J1022+1001

In the previous chapters, I have presented new techniques in pulsar searching ranging from enhancements in computing to robust candidate classification. When a pulsar is newly discovered, the only parameters that are relatively well known are the spin period, dispersion measure (DM) and a rough estimate of the position. Additionally if the pulsar is seen to have a variable spin period over the course of the discovery observation we may also infer that it is part of a binary system. In order to refine the parameters of the pulsar further, it is monitored regularly via multiple followup observations. The technique behind the monitoring and retrieval of the pulsar parameters is known as pulsar timing.

Pulsar timing is an integral aspect of pulsar astronomy which deals with consistent tracking of pulses from a pulsar and determining the times of arrival (ToAs) of pulses. The interpretation of the ToAs are done based on a timing model. This model accounts for multiple effects including orbital motion (if the pulsar is in a binary), intrinsic spin evolution of the pulsar, ISM propagation factors and the correction of the ToAs to the presumed inertial reference frame of the solar system barycentre. The model is regularly updated based on a fit of the the ToAs over a long period of time (typically years). The key aim of pulsar timing is accurate modelling of the spin evolution of pulsars through precise ToA measurements. This precision is required for achieving the sensitivity needed by pulsar timing arrays to detect gravitational waves (as motivated in Section 1.4.3). The standard way of calculating a ToA is to measure the pulse arrival time with respect to a certain standard reference. This is typically done by cross correlating the observed pulse profile with a standard template profile in a process known as template matching. The time offset obtained by cross correlation between a standard template and the observed profile can be converted into a phase lag on dividing by the instantaneous pulse period. The observed profile $P(t)$ can be expressed as a function of the template $T(t)$ by (Lorimer & Kramer, 2012):

$$P(t) = a + bT(t - \tau) + N(t) \quad (6.1)$$

where a is an offset, b is a scaling factor, τ represents the phase offset used for the ToA calculation and $N(t)$ is a noise term. The cross-correlation technique typically used is known as the Fourier phase gradient algorithm. The technique transforms the profiles to the Fourier domain before applying a reduced chi squared minimisation to keep the residuals between the template and the observed ToAs to a minimum (Taylor, 1992).

Although this process has proven successful in determining the pulsar parameters with great precision, decades of pulsar observations have revealed phenomena that can advance or delay the expected ToA in different ways. These phenomena range from short term effects like pulse-to-pulse jitter (Osłowski et al., 2011) to temporal correlation of the timing residuals over timescales of several years (also known as timing noise; see e.g. Shannon & Cordes (2010)). Besides these phenomena, an important assumption of the pulsar timing methodology is that the template pulse profile is representative of the pulsar at all times. This can bias the timing model estimates made and potentially lead to discrepancies in long term timing if the pulse profile is unstable. It is thus important to understand all possibilities that influence the stability of average pulse profiles prior to generating templates. As an example, I present a case study of a millisecond pulsar PSR J1022+1001 that is regularly timed as part of pulsar timing arrays and whose profile stability has come under heavy scrutiny since its discovery.

This chapter is based on an article titled “Revisiting profile instability of PSR J1022+1001”, which is published in Monthly Notices of the Royal Astronomical Society, Volume 500, Issue 1, Pages 1178–1187, January 2021¹. The manuscript remains unchanged, but is formatted to fit the template of the thesis.

As the lead author of this publication, I have developed the methods discussed, reduced the entirety of the data, performed the analysis and written the manuscript.

The full author list is **P.V. Padmanabh**, E.D. Barr, D.J. Champion, R. Karuppusamy, M. Kramer, A. Jessner and P. Lazarus.

Contents

6.1	Abstract	171
6.2	Introduction	171
6.3	Previous profile stability studies of J1022+1001	172
6.4	Observations and Data Reduction	174
6.5	Methods	176
6.5.1	Previous methods	176
6.5.2	Peak optimiser	176
6.5.3	Modified Hotan method	178
6.6	Results	179
6.6.1	EPOS	179
6.6.2	CPSR2	181
6.6.3	PSRIX	183
6.6.4	Comparing telescope data set distributions through KS test	183
6.7	Scintillation based analysis	183
6.8	Discussion	187
6.9	Conclusions	188

¹Available at <https://ui.adsabs.harvard.edu/abs/2021MNRAS.500.1178P/abstract>

6.1 Abstract

Millisecond pulsars in timing arrays can act as probes for gravitational wave detection and improving the solar system ephemerides among several other applications. However, the stability of the integrated pulse profiles can limit the precision of the ephemeris parameters and in turn the applications derived from it. It is thus crucial for the pulsars in the array to have stable integrated pulse profiles. Here we present evidence for long-term profile instability in PSR J1022+1001 which is currently included in the European and Parkes pulsar timing arrays. We apply a new evaluation method to an expanded data set ranging from the Effelsberg Pulsar Observing System back-end used in the 1990s to that of data from the current PSRIX backend at the Effelsberg Radio Telescope. We show that this intrinsic variability in the pulse shape persists over time scales of years. We investigate if systematic instrumental effects like polarisation calibration or signal propagation effects in the interstellar medium causes the observed profile instability. We find that the total variation cannot be fully accounted for by instrumental and propagation effects. This suggests additional intrinsic effects as the origin for the variation. We finally discuss several factors that could lead to the observed behaviour and comment on the consequent implications.

6.2 Introduction

Millisecond pulsars (MSPs) show properties that serve as an ideal platform for testing general relativity in the strong field regime (e.g. [Kramer et al., 2006b](#)), probe the equation of state for neutron stars (e.g. [Lattimer & Prakash, 2001](#); [Antoniadis et al., 2013](#)) and may allow to detect gravitational waves in the nanohertz regime (e.g. [Foster & Backer, 1990](#)). Key to these investigations is the technique of pulsar timing where the arrival time of the pulsar signal is monitored by comparison against a reference template of the average pulse profile. Here, typically, tens of thousands of pulses are averaged to form an integrated profile. A template is generated from such integrated profiles across many observations. In order to avoid self imaging effects ([Hotan et al., 2005](#)), the grand average profile is often converted into a template by modelling the pulse as a superposition of multiple Gaussians or von Mises functions. This noise free template is then used as a reference and cross correlated with profiles from each epoch to calculate the times of arrival (ToAs). An accurate timing model of the spin and binary parameters of the pulsar should yield timing residuals which follow a Gaussian distribution with zero mean. However, the residuals most commonly show systematic variations due to several sources of noise that could be deterministic or stochastic in nature (e.g. [Cordes, 2013a](#)).

The method described above assumes that the integrated profile is invariant with time for a given observing frequency. However, several pulsars are known to show variable integrated pulse shapes. Some variations are stochastic. For example, when a few pulses (tens to hundred) are combined to form a profile, the phase jitter introduced causes noise in the timing residuals. This is mainly caused due to individual pulses fluctuating in shape, flux as well as phase. Longer integration times tend to average

out these changes and form much more stable profiles, though some variability may still persist at some level due to other effects. Changes may arise from the integrated profile switching between two or more profiles termed as “moding”. These transitions typically switch back within a few thousand pulses (e.g. [Backer, 1970](#); [Rankin et al., 2013](#)). Some pulsars have been observed to show variations in pulse profiles on longer time scales of weeks, months or years (e.g. [Lyne et al., 2010](#)) which is longer than typical moding timescales.

Several studies have tried to characterise jitter ([Osłowski et al., 2011](#); [Liu et al., 2012](#); [Shannon et al., 2014](#)), intermittency (e.g. [Mottez et al., 2013a](#)) as well as moding ([Cordes, 2013b](#)), but the underlying mechanism that induces the transition between stable configurations of the pulse shape is not understood. Other variations are secular, being especially prominent in some binary pulsars over timescales of years. These arise from relativistic spin precession (e.g. [Kramer, 1998](#); [Desvignes et al., 2019](#); [Krishnan et al., 2019](#)) due to misalignment of their spin with the orbital angular momentum. This effect can be modelled and helps constraining the pulsar beam’s structure.

MSPs are generally known to have a stable average pulse profile (e.g. [Shao et al., 2013](#)). Their short spin period allows many pulses to be averaged on time scales of minutes. Any change in the average profile of a pulsar can cause subtle unexpected variations in the timing residuals. This is due to the varying fiducial point while calculating ToAs. Devising techniques that can cope with profile changes are important to resolve this issue. On the other hand, studying these variations carefully can also give a deeper insight into factors that could impact the pulse shape. These can be broadly categorised into instrumental and non-instrumental effects. To understand these factors in more detail, we have conducted a detailed study of PSR J1022+1001, for which past analyses have reported contradictory results regarding the profile stability of this pulsar ([Kramer et al., 1999](#); [Ramachandran & Kramer, 2003](#); [Hotan et al., 2004a](#); [van Straten, 2013](#); [Liu et al., 2015](#); [Shao & You, 2016](#)).

In Section 6.3, we review previous profile stability studies of PSR J1022+1001. In Section 6.4 we explain the procedure for data reduction of the different data sets as well as the analysis tools used. In Section 6.5 we revisit the procedures and methods for characterising pulse profile stability. Section 6.6 discusses the results from applying new techniques to various data sets. We also cover the impact of polarisation miscalibration as well as interstellar scintillation effects. Section 6.8 discusses the scientific explanations for the results obtained. We state our conclusions in section 6.9.

6.3 Previous profile stability studies of J1022+1001

PSR J1022+1001 is a binary pulsar with a spin period of 16.45 ms in a 7.8-day orbit ([Camilo et al., 1996](#)) with a CO white dwarf companion ([Lundgren et al., 1996](#)). The average pulse profile exhibits a two peaked structure at 1.4 GHz (see Figure 6.1). The pulsar is currently a part of the European Pulsar Timing Array (EPTA) and the Parkes Pulsar Timing Array (PPTA) among other pulsars in the quest to detect nanohertz gravitational waves. Several studies across many years using different telescopes have

shown conflicting results regarding the pulse profile stability of this pulsar.

Camilo et al. (1996) reported unusual pulse profile variations on the time scale of a few minutes within the first few months of timing this pulsar with the Arecibo telescope. Kramer et al. (1999) observed this pulsar with the Effelsberg radio telescope and found profile changes across integration times of approximately 40 min. Given that the shape changes persist even after integrating for hundred thousand pulses, Kramer et al. (1999) suggested that moding cannot explain the observed instability.

Turbulence in the interstellar medium causes constructive or destructive interference leading to enhancing or weakening of the propagating pulsar signal in time as well as observing frequency. This is termed as scintillation. This propagation effect coupled with profile evolution across the observing bandwidth could lead to changes in the integrated pulse profile (Ramachandran & Kramer, 2003). However, Kramer et al. (1999) observed profile changes within a narrow frequency band (40 MHz) that could not be explained by scintillation effects. The calculated ToAs showed that the profile variation impacted the timing stability of this pulsar. The root-mean square (r.m.s) residuals post fitting were approximately 15-20 μ s.

Altitude-Azimuth (Alt-Az) telescopes have to account for the rotation of the feed with respect to the sky as given by the parallactic angle during observations. Any miscalibration in polarisation could lead to leakage in power between the polarisation channels. This would in turn affect the shape of the profile if the pulsar is highly polarised as PSR J1022+1001. To confirm that the seen changes were not telescope and mount dependent, Ramachandran & Kramer (2003) observed this pulsar with the equatorial mount Westerbork Synthesis Radio Telescope (WSRT), as an equatorially mounted telescope is not affected by the change in parallactic angle. Ramachandran & Kramer (2003) observed shape variation in time thus confirming the results obtained by Kramer et al. (1999). Ramachandran & Kramer (2003) also hypothesised that the shape changes are possibly due to magnetospheric return currents and profile variation.

Subsequent studies were carried out by Hotan et al. (2004a) using the Parkes observatory. They noticed that the integrated profile was remarkably stable and showed no significant signs of instability with time. In contrast to the results obtained by Kramer et al. (1999), they obtained a much smaller r.m.s in timing residuals of 2.27 μ s. They also observed that calibrated profiles showed higher degree of instability in comparison to uncalibrated profiles. While this may appear surprising, it may indicate that instrumental errors could possibly contribute to profile variations.

Purver (2010) conducted another analysis of profile variations in PSR J1022+1001. He used data from Effelsberg, Parkes as well as the WSRT to study possible profile changes on timescales of 10 min. He also analysed profiles with a bandwidth of 8 MHz to ensure minimal impact from profile evolution with frequency. The reduced chi-squared values of his shape change parameters indicated significant changes in the pulse profile. He argued that these changes are not biased by the metrics themselves. He also argued that any instrumental errors that could cause the shape change would be well in excess of the two *per cent* change in total intensity as reported by Hotan et al. (2004a)

van Straten (2013) suggested that instrumental polarisation artefacts could lead to systematic errors in the timing residuals. He observed that cross-coupling vari-

ations over 7 years of data with the Parkes telescope introduced systematic errors close to $1 \mu\text{s}$ in ToA estimation. He developed a new calibration technique and used PSR J1022+1001 as a test bed and demonstrated that the weighted r.m.s residuals could be reduced to 880 ns, a factor of at least two lower than previous analyses. The integration time used was roughly one hour. Through simulations, [van Straten \(2013\)](#) suggested that [Kramer et al. \(1999\)](#) observed the variations due to changes in the parallactic angle resulted in mixing of the total intensity with the linearly polarised power of the profile. He also commented that the same variation seen by [Ramachandran & Kramer \(2003\)](#) can be attributed to inaccurate calibration of the polarimetric response of the system.

[Liu et al. \(2012\)](#) analysed five MSPs including PSR J1022+1001 using parameterisation of the profile shape and phase jitter to analyse the profile stability. This analysis suggested that the profile does not vary significantly with time. However, they also pointed out that the shape parameter used in their analysis describes the entire on-pulse region, and may not be sensitive to profile variations occurring exclusively at the peaks.

More recently, [Liu et al. \(2015\)](#) carried out single pulse as well as integrated profile studies using simultaneous observations from Effelsberg and Westerbork. They investigated effects of potential polarisation miscalibration of the Effelsberg data by using the template matching technique ([van Straten, 2006](#)). After modelling the differential gain and cross coupling of the feeds, [Liu et al. \(2015\)](#) showed that applying different models to calibrate the data accounts for only a marginal fraction of the observed profile variability. They showed that subpulses from the leading and trailing components of the pulse are correlated. They also showed that subpulse variation is not correlated with the integrated profile variability observed.

[Shao & You \(2016\)](#) analysed Parkes data taken from the PDFB4 backend. The relatively large bandwidth of this system (256 MHz) allowed for studying pulse profile changes across frequency. The data were polarisation calibrated using the method described in [van Straten \(2004\)](#). They observed that the pulse profile changed across the bandwidth. They also noticed that the flux density for each frequency subband was different. They suggested that the profile variations are a consequence of profile evolution with frequency coupled with interstellar scintillation.

While all the studies have shown methodologically robust ways of characterising the pulse profile stability, the conflicting results provide little clarity on the origin or even the existence of variability in this pulsar. The profile variability observed could result from ISM effects in the form of scintillation or miscalibration of the instrument or due to phenomena intrinsic to the pulsar and its magnetosphere.

6.4 Observations and Data Reduction

To understand and seek to resolve the differences seen by the aforementioned analyses, we collected data that span close to two decades. We chose the data sets used by [Kramer et al. \(1999\)](#) and [Hotan et al. \(2004a\)](#) given that their results disagree with one

Table 6.1: Different data sets used as well as their year spans for this analysis. (N) indicates the number of epochs used for analysis

Telescope	Backend	Span	N	Dedispersion
Effelsberg	EPOS	1995-2005	50	Incoherent
Parkes	CPSR2	2004-2010	19	Coherent
Effelsberg	PSRIX	2011-2015	25	Coherent

another. [Kramer et al. \(1999\)](#) used data from the Effelsberg radio telescope recorded with the Effelsberg Pulsar Observing System (EPOS) ([Jessner, 1996](#)). [Hotan et al. \(2004a\)](#) used data from Parkes recorded with the CPSR2 backend ([Hotan, 2005](#)) which generates coherently dedispersed profiles. We extended the time baseline of the data set by using data from the Effelsberg telescope from the latest PSRIX backend ([Lazarus et al., 2016](#)). Table 6.1 provides exact details of the span of each data set used.

The data presented here correspond to the L-band radio wavelengths (1.3 - 1.4 GHz) with the receivers tuned to slightly different centre frequencies. The data were homogenised through clipping to a common overlapping band centred at 1390 MHz with a 40 MHz bandwidth. The EPOS backend was connected to an online incoherent dedisperser along with an adding polarimeter and did not provide frequency resolution ([von Hoensbroech & Xilouris, 1997](#)). Thus, we averaged across the entire clipped bandwidth for all epochs of the CPSR2 and PSRIX data to maintain consistency across data sets. We also optimised for any smearing caused by an inconsistent ephemeris by refolding the archives with the best known parameters after fitting for a linear drift in the pulse across the observed time.

Reduction of the archive files was done using the PSRCHIVE² package ([Hotan et al., 2004b](#)). The profiles were manually cleaned to mitigate Radio Frequency Interference (RFI) using `pazi`. A few epochs from CPSR2 showed inaccurate linear polarisation and circular polarisation both before and after calibration. These epochs turned out to have no signal from one of the polarisation channels for every alternate sub integration of an epoch. Data showing these artefacts were removed from the analysis to avoid inconsistencies.

The PSRCHIVE Python module was used to extract data from archive files before applying methods and techniques to characterise the pulse profile shape. The details of these methods are discussed in the following section. Archive files with signal-to-noise ratio (S/N) below 100 were excluded from the analysis to avoid uncertainties limited by poor S/N. The data sets from the CPSR2 and PSRIX backends were polarization calibrated using `pac`³. The Jones matrix solutions for the Parkes multibeam were derived using observations from PSR J0437-4715 based on the measurement equation template matching (METM) technique as described in ([van Straten, 2013](#)). We calibrated the PSRIX data using the single axis model ([Britton, 2000](#)).

²<http://psrchive.sourceforge.net/>

³<http://psrchive.sourceforge.net/manuals/pac/>

To use a uniform analysis environment, we modified the current PSRCHIVE plugin⁴ for the legacy EPOS data. Subsequently we used the PSRCHIVE Python plugin for the rest of the analysis. We used a Python based pipeline (EPOS_calibrator⁵) in order to calibrate the EPOS data. The pipeline implements calibration based on equations for an adding polarimeter as specified in von Hoensbroech & Xilouris (1997).

6.5 Methods

6.5.1 Previous methods

Kramer et al. (1999) used a simple but straight forward metric to describe the pulse profile stability of PSR J1022+1001. They used the amplitude ratio computed from the two peaks in the integrated profile and took a ratio between them with one of the peaks as a reference. If the profile is stable, one would expect the amplitude ratio to be stable across time. They observed that a profile integrated up to 10-16 min could show different amplitude ratios with a 30 *per cent* scatter.

Hotan et al. (2004a) used a different metric to characterise the profile stability. Every integrated profile spanning 5 min was first normalised based on the flux density in the on-pulse region. A high-S/N normalised profile was chosen as a standard reference template. The template was then subtracted from the normalised profile for every epoch. All the subtracted residuals were stacked and the standard deviation was calculated per phase bin across epochs. If the profile were stable with time, the residuals should behave like white noise and the standard deviation across phase bins should ideally remain constant across the profile. This result reported in Hotan et al. (2004a), indicates no sign of instability.

6.5.2 Peak optimiser

In order to first investigate the data with a method similar to the first method in the previous subsection, we apply a robust approach that uses specific features of the profile. We focus on the peaks of the two components of the pulse profile. The integrated profile is generated for every epoch such that the on-pulse is roughly centered and not impacted by the pulse profile wrapping in phase. To ensure consistency across all data sets, we generate integrated pulse profiles with 1024 bins across rotational phase.

Since PSR J1022+1001 has two well defined peaks in its profile at 1.4 GHz (see Figure 6.1), we choose the first peak (hereafter FP) and the second peak (SP) as features to analyse the stability of the pulse profile. We next fit for both phase and amplitude of FP and SP.

Firstly, the phase bin with the maximum amplitude is used as a starting reference for each of the peaks. A window centered on the reference point is chosen to represent the peak. We then solve for the turning point of a Least-Squares second order polynomial fit. The fit has the advantage of yielding a single maximum and does not over-interpret

⁴<https://sourceforge.net/p/psrchive/code/ci/master/tree/Base/Formats/EPOS/>

⁵https://github.com/prajwalvp/EPOS_calibrator

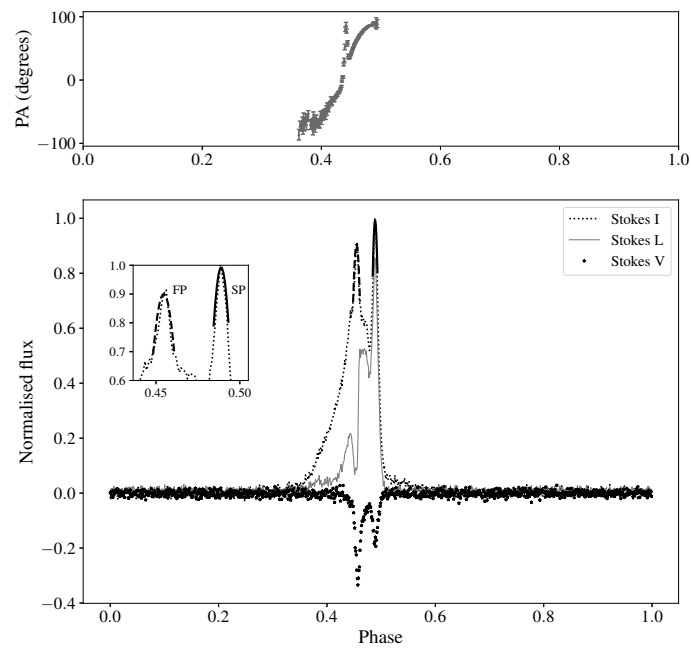


Figure 6.1: Bottom panel shows a sample Stokes profile for PSR J1022+1001 at 1.4 GHz observed with CPSR2. The plot inset shows an example second order polynomial fit on the first peak (FP) and second peak (SP). The top panel shows the position angle (PA) as a function of rotational phase.

the shape of the profile. It also requires fewer points compared to a higher degree polynomial fit. A different window size is chosen for FP and SP since they differ in shape and sharpness of the peak. The size of the selected window for each peak is chosen by using the reduced χ^2 . For this, we chose different window sizes and applied the above mentioned method. We then calculate the reduced χ^2 value for the obtained fit with respect to the original profile in the window of interest. We eventually choose the window size that yielded the reduced χ^2 value closest to 1. This yielded a window size of 20 bins for FP and 10 bins for SP.

To make a robust determination of the uncertainties in the fit, we take a Monte Carlo-based approach. We calculate the off-pulse r.m.s and generate multiple realisations of the selected window per peak with data in the window drawn from a Gaussian distribution with mean as the value of the point within the window and the off-pulse r.m.s as the standard deviation for that point. We then apply the parabolic fit for each realisation and solve for the phase bin value at which the turn over of the fit occurs. The final reported value comes from the median of the distribution of the peaks obtained from the different iterations. The error bars reported come from choosing the values at the 5th and 95th percentile of the distribution.

To make this procedure more robust, we use the new phase value as a starting point reference with the same previously selected window size per peak. The Monte Carlo approach is applied once again to solve for the maximum. This bootstrapping technique helps improve the characterisation of the peak. We also apply a similar procedure to characterise the amplitude. Once the windows are finalized and the fit is applied, we solve for the amplitude keeping the phase bin value fixed. This fixed phase bin value comes from solving for the maximum from the second order fit as mentioned previously. Once the phases and amplitudes for each of the peaks have been solved for, we use two different metrics to characterise the profile stability. Firstly, we take a difference of the solved phase bin values for FP and SP. Secondly, we take an amplitude ratio of the solved amplitude values between both the peaks similar to [Kramer et al. \(1999\)](#). We will refer to the phase difference and amplitude ratio calculation together as the Peak Optimiser (hereafter PO).

6.5.3 Modified Hotan method

In addition to these two metrics referring to phase and amplitude, we also apply a slight modification to the method used by [Hotan et al. \(2004a\)](#). The original method uses a high S/N reference profile template and stacks the difference amplitudes of profiles with respect to the template from every epoch. The standard deviation per phase bin of the stacked profiles is also calculated to make this more robust. We modify this method and calculate a median of all average profiles across all chosen epochs to use as a reference template for subtraction. This makes the analysis less susceptible to statistical outliers. We align each profile in phase before subtraction. This is done by rotating the profiles in the Fourier domain and additionally optimising for fractional phase bin variation if any. To avoid residual errors from normalisation, we also model for an offset in the on-pulse region of the standard deviation per phase (STDP) plot

and apply a correction using a cost function to minimise the r.m.s of the STDP. We will hereafter refer to this method as MHM.

In order to compare data between backends, we have applied the Kolmogorov-Smirnov test (Massey, 1951). The advantage of this test is that it is non parametric. We select data pair wise between the three backends in this analysis and test the null hypothesis that the two distributions in consideration come from the same parent distribution.

6.6 Results

Using the new and modified stability metrics as discussed in the previous section, we analyse data from three backends (see Figure 6.2). Within each data set, we distinguish between polarisation calibrated and uncalibrated data. Polarisation calibration is essential in eliminating systematic effects. For instance, incorrect calibration can affect the two detected polarisation channels differently and eventually result in a Stokes I variation. van Straten (2013) suggested that miscalibration had a major effect on precision pulsar timing for several pulsars including PSR J1022+1001 and so developed a technique that modelled possible artefacts in order to account for this. We thus expanded the analysis to study the effect of improper polarisation calibration on the pulse profile stability across all the data sets. We discuss the results obtained on each data set below.

6.6.1 EPOS

The EPOS data span approximately 10 years from 1995 to 2005 and the observation lengths for all epochs range between 20 and 30 min. The results from applying PO and MHM are shown below.

Figure 6.2 shows the results of applying PO to the EPOS data set. It is seen that there is significant scatter in both phases and amplitude ratios signifying profile instability. While calibration reduces the scatter marginally (1-2 *per cent*), there is still unaccounted variability observed for the profile across time. Kramer et al. (1999) had observed a 30 *per cent* scatter in the amplitude ratio across an observation and we observe a slightly higher (38 *per cent*) scatter across a long range of epochs.

The results from applying MHM can be seen in Figures 6.5(a) and (d). The standard deviation around the on-pulse region is high for uncalibrated (21-sigma significant) and calibrated data sets (16-sigma significant). The difference amplitude plots also show clear variations in the on-pulse region. Hence, the figures clearly indicate that the profile is unstable on timescales of 20-30 min which is commensurate with the results from applying the PO.

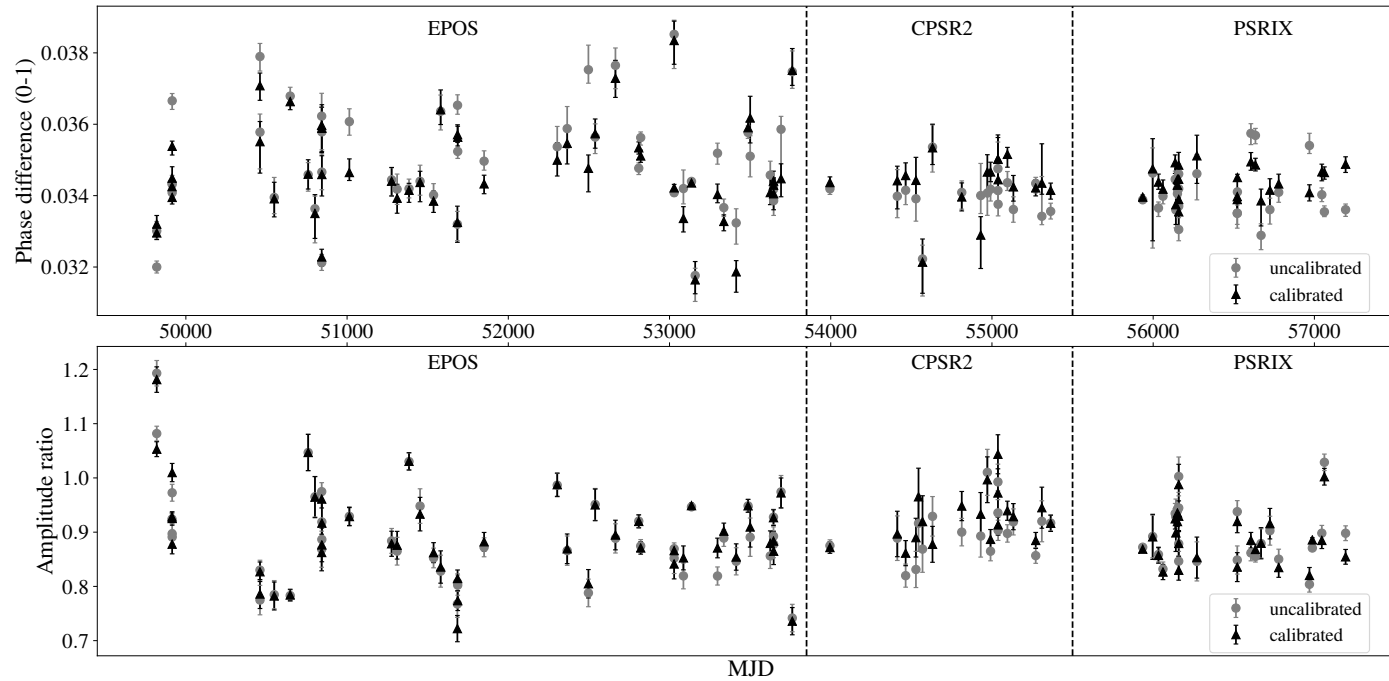


Figure 6.2: PO results for all three data sets across 20 years. Top panel shows the difference in peak phases plotted against time. The bottom panel shows amplitude ratios between peaks plotted against time.

To assess the effect of polarisation calibration, we varied the ratio of the gains of each polarisation and applied PO on each gain ratio (see Figure 6.3). The gain levels in each polarisation are typically within 10 dB of each other. Figure 6.3 shows that the range of phase differences seen by changing the gain ratios from 0.1 to 2 is close to 0.002 (in phase). This is at least a factor of 2 smaller than the range of phase differences observed in Figure 6.2. Similarly, the amplitude ratio changes by about 0.05 which is close to a factor of 10 smaller than the change observed in Figure 6.2. This indicates that although we change the gain ratio by more than an order of magnitude, it cannot fully account for the observed variability in the integrated pulse profile. This provides strong evidence that instrumental gain effects cannot fully explain the profile instability observed for the EPOS data.

6.6.2 CPSR2

The Parkes CPSR2 data span approximately 7 years from 2004 to 2010 and were obtained from the CSIRO Data Access Portal⁶. The observation lengths chosen were between 20 and 30 min.

Figure 6.2 shows that the scatter in phase difference increases by about 0.002 units when the data are calibrated. But we observe that calibration helps to reduce the scatter on the amplitude ratio. We further investigated the impact of miscalibration by inspecting the cumulative distribution function for the calibrated and uncalibrated data. We focused on the amplitude ratio given that this parameter is less susceptible to artefacts compared to the phase difference (as seen from the polarisation analysis in the previous subsection). We observe a scatter of roughly 30 *per cent* (see Figure 6.4) for the cumulative distributions. We also observe a slight offset between the distribution functions of the uncalibrated and calibrated Parkes data sets. However, this difference is not significant to ascertain that calibration makes an impact (as indicated by the p-value).

In order to investigate if the profile is unstable at smaller timescales than 20-30 min, we first applied the original method used by Hotan et al. (2004a) on their data with an integration length of 5 min (See Appendix 6.9 for details). The results for the uncalibrated data are consistent with the results from Hotan et al. (2004a). We also compared the results for applying the original method on the shorter integration (5 min) as well as longer integration (20-30 min). We observe that the calibrated as well as uncalibrated data for the long integrations show significant variability and the calibration is observed to improve the profile stability across epochs.

Apart from applying the original method, we also applied the MHM on the data set used by Hotan et al. (2004a). The difference amplitudes show no clear evidence for instability but the standard deviation of all epochs across the rotation phase of the pulse show a near 4-sigma significance in the on-pulse region.

The PO analysis showed some ambiguity about the stability indicators for the 20-30 min integration data. Thus, using a different metric would help in throwing some more light on the issue. Figures 6.5(b) and (e) show results from applying MHM to the

⁶<https://data.csiro.au/dap>

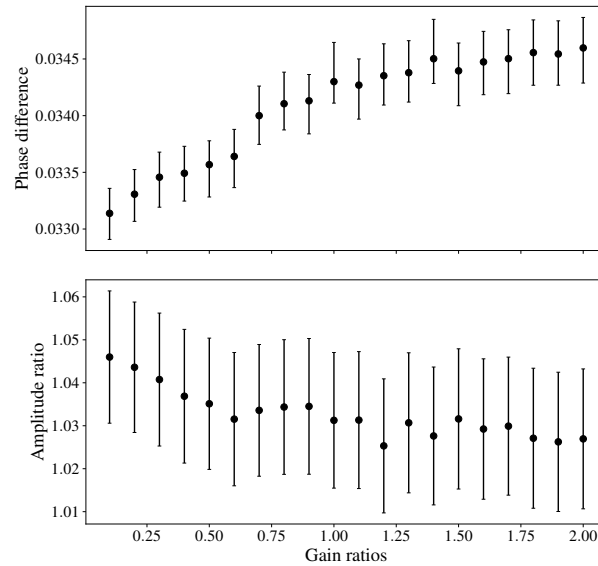


Figure 6.3: Top panel: Plotted is the difference in phase of the two peaks for a sample EPOS profile against the ratio in the gains for the two polarisations. Bottom panel shows the amplitude ratio as a function of the gain ratio.

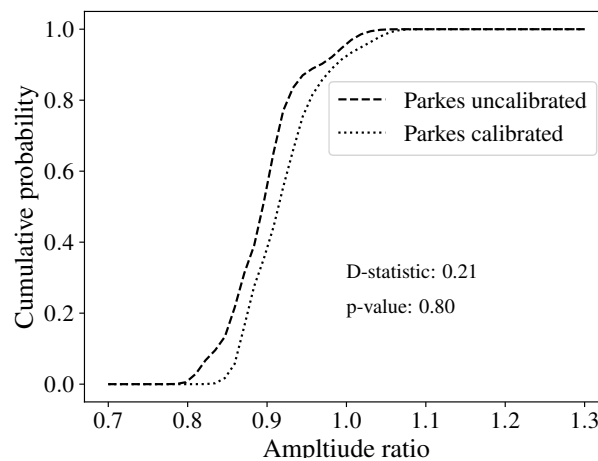


Figure 6.4: Comparison of cumulative distribution functions for CPSR2 data between uncalibrated and calibrated data.

long integration data. The maximum standard deviation around the on-pulse region is 8-sigma above the off-pulse deviation for the uncalibrated data. This value is close to 6-sigma for the calibrated data set. These values are much more significant compared to the value obtained for integration length of 5 min. This indicates that different integration timescales can play a crucial role in these studies. The dependence on integration timescales could plausibly explain the difference between this analysis and that of [Hotan et al. \(2004a\)](#).

6.6.3 PSRIX

The data taken with the PSRIX backend of Effelsberg span approximately 5 years from 2011 to 2016 with observation lengths ranging between 20 and 30 min per epoch, similar to previous data sets.

Figure 6.2 shows results from applying the PO on the PSRIX data set. We observe that the scatter in phase differences as well as the amplitude ratio lowers on calibration. Still, both plots support the hypothesis that the profile is unstable on these timescales as seen from the previous two data sets.

The results from applying MHM to the PSRIX data are consistent with results from applying PO (See Figures 6.5(c) and (f)). We observe a significant standard deviation difference between the maximum deviation and the off pulse deviation for both calibrated (7-sigma significant) and uncalibrated (7-sigma significant) data sets. The difference amplitude plots complement the observations of the standard deviation.

6.6.4 Comparing telescope data set distributions through KS test

In order to investigate different backends, we observe the trends in the cumulative distribution functions obtained for the amplitude ratios. This is a robust indicator of any instrumental effects that have impacted our pulse profile analysis.

Ideally, one would expect the distributions in pulse shape changes to be independent of the telescope backend, but as seen in Figure 6.2, we see a narrow range of amplitude ratios for PSRIX and a much wider range for EPOS. On injecting a white noise source with nearly 4 times the r.m.s of the amplitude ratios of the PSRIX data, we are completely able to explain the differences observed (see Figure 6.7). The p-value which was 0.86 became 1.0 after injection of noise. This suggests that some part of the EPOS backend chain significantly increased the noise floor of the incoming signal.

We also conducted a KS test between CPSR2 and PSRIX data post-calibration to check for similarities or differences. In Figure 6.8, we observe an offset between the distributions. However, the p-value indicates that we cannot reject the null hypothesis that the PSRIX and CPSR2 distributions arise from the same parent distribution.

6.7 Scintillation based analysis

Some studies ([Shao & You, 2016](#); [Ramachandran & Kramer, 2003](#)) had suggested that profile evolution with frequency along with interstellar scintillation could lead to varying

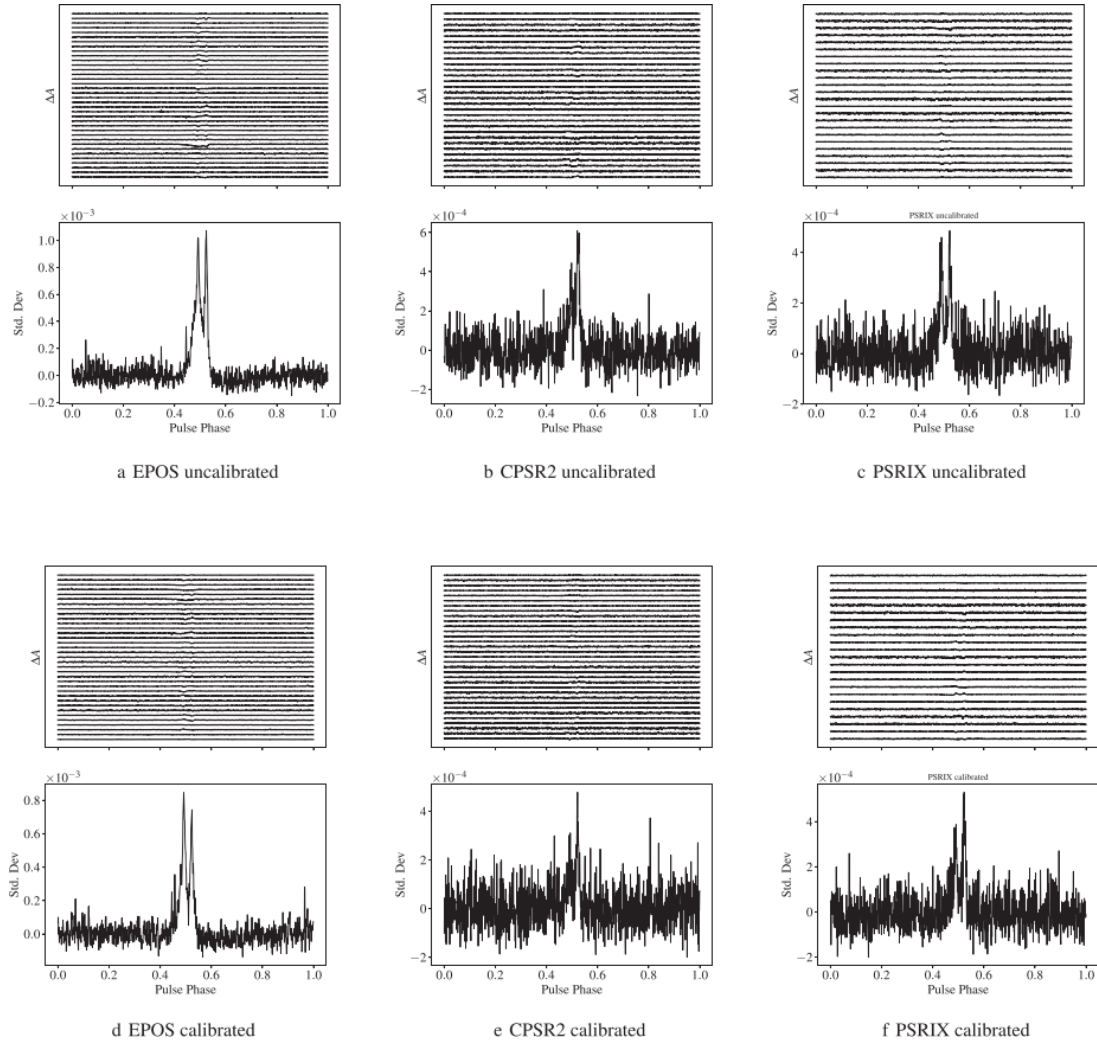


Figure 6.5: MHM results for all calibrated and uncalibrated data. The top panel in each individual plot is denoted by ΔA which refers to the difference amplitudes stacked for flux normalised profiles. The bottom panel is the standard deviation across all epochs as a function of the rotational phase.

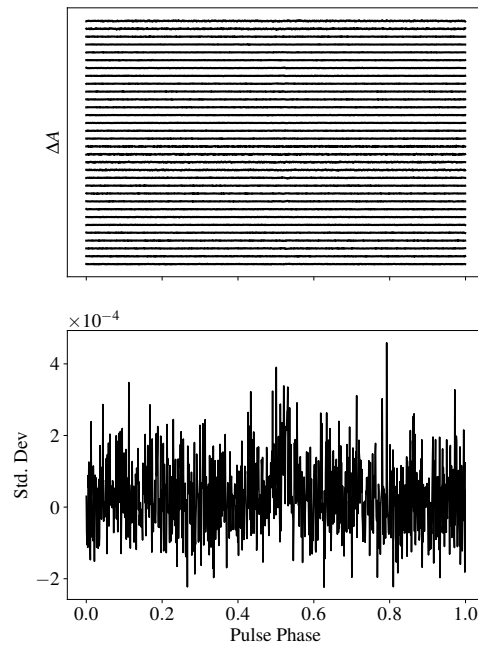


Figure 6.6: MHM plot for the data set used by [Hotan et al. \(2004a\)](#). The integration time for each epoch is five minutes. The top panel is the difference amplitudes stacked and the bottom panel is the standard deviation per phase bin across all epochs.

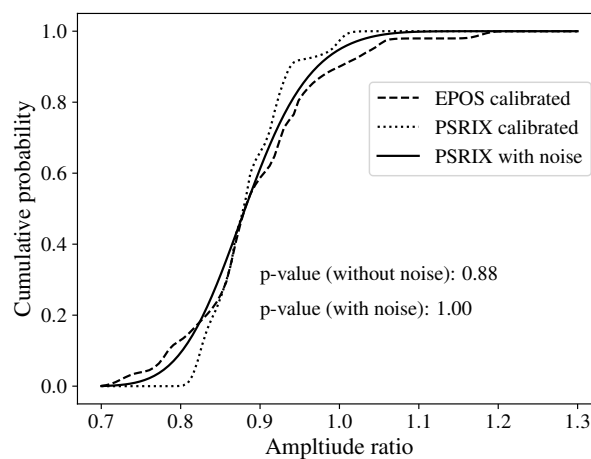


Figure 6.7: Comparison of cumulative distribution functions of amplitude ratios between PSRIX, EPOS data and PSRIX data with white noise.

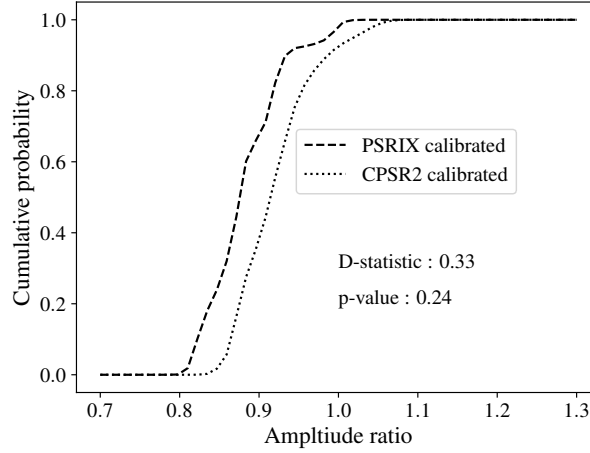


Figure 6.8: Comparison of cumulative distribution functions of amplitude ratios between PSRIX and CPSR2 data after calibration.

pulse profiles. Liu et al. (2015) used simulations based on real profile evolution across frequency and showed that scintillation can introduce significant profile variations for a large bandwidth (150-200 MHz). If the profile shows different structures across different observing frequencies, scintillation can lead to variations in the integrated profile. We calculated the expected scintillation bandwidth for this pulsar using the NE2001 model (Cordes & Lazio, 2002). This is 57 MHz at 1.4 GHz which is of the order of the bandwidth used in our analyses. The scintillation timescale from the model is close to 15 minutes.

We investigated the impact of diffractive scintillation on CPSR2 and PSRIX data. We generated median profiles for six frequency sub bands based on observations from all epochs for a particular backend. We then modelled the gain variation across the observing bandwidth as a sinusoid and shifted it in phase in steps of a quarter wavelength across the band to create multiple templates. We overlaid these templates across the bandwidth. This would weight each subband differently based on the sinusoidal template used. We finally applied the PO for the frequency scrunched profile based on each sinusoidal template. We further extended this procedure for multiple sinusoidal frequencies ranging from 0.5 to 2.5 wavelengths in steps of 0.25. We then calculated the maximum possible scatter value that can be obtained by varying the sinusoidal frequency as well as shifting the template in phase.

We observe that the phase difference has a maximum variation of 0.0008 as compared to 0.0025 as seen in our analysis with PO for PSRIX data (see Figure 6.2). Thus, the variations from diffractive scintillation accounts for roughly up to 30 *per cent* of the variation observed with PSRIX. Similarly, the amplitude ratio has a maximum variation close to 25 *per cent* of the total variation observed for PSRIX data. We carried out the same analysis with the CPSR2 data and the maximum variation in phase difference and amplitude ratios account up to 15 *per cent* and 20 *per cent*, respectively. These

results indicate that scintillation cannot explain the observed instability.

6.8 Discussion

We were able to reproduce the result from [Kramer et al. \(1999\)](#) for the scatter seen in the amplitude ratios across time. Though the scatter observed in [Kramer et al. \(1999\)](#) was across an observation of a few hours, our analysis showed that integrated profiles of the order of 30 min still show the same behaviour. The EPOS backend specifically shows a higher degree of instability compared to the other data sets (about 20 *per cent* higher scatter). This is especially apparent from the KS test analysis where introducing a constant noise source to the PSRIX data fits the model for EPOS. At the beginning of the observation, the channel gains for EPOS were set manually on dials trying to keep the total power readings for each channel within a prescribed range typically at a medium to low elevation and higher system noise. To make sure that the manual setup is not causing any artefacts, we varied the gains of each cross polarisation up to 10 dB and reconstructed the profile for various gain ratios. The analysis showed that the profile shape change is negligible compared to the scale of changes noticed for pulse profiles across time. The dynamic range of the EPOS system was limited and interference at times saturated one of the polarisation channels and not the other. A possibility is that the first component of the pulse could saturate the receiver causing the system to be unstable and in turn leading to varying amplitudes in the second component. However, pulsar signals observed with EPOS were usually not strong enough to saturate the receiver. Additionally, if this were the case, we would not expect to see the profile shape changing for the other backends. We reproduced the results from [Hotan et al. \(2004a\)](#) for 5 min integrations of the pulse profile (as shown in the Appendix). Though we notice no trend in the difference amplitudes, the standard deviation plot reveals a hint of instability that is further confirmed when applying MHM on the 5 min integrations. This slight difference in the results in comparison with [Hotan et al. \(2004a\)](#) is most likely due to a slight difference in the epochs used for analysing. However, it is well evident from MHM and the original method used in [Hotan et al. \(2004a\)](#) that the profile is unstable at longer timescales on the CPSR2 data. We also explored the possibility of profile evolution with frequency playing a role in shape changes. While there is clear evidence of profile evolution with frequency for this pulsar ([Ramachandran & Kramer, 2003](#); [Shao & You, 2016](#)), this cannot directly explain the profile changes seen across time if we observe across the same range of observing frequencies. One can expect profile evolution with frequency to have an impact when a 'scintle' passes through the observing band. We thus modulated the band with a wide range of simulated scintillation frequencies and phases to determine the maximum variation that could be produced. We find that scintillation can only account for 25 *per cent* of the observed variability.

The KS test analysis on the CPSR2 data set revealed that calibration introduces a slight offset but does not play a significant role in distorting the pulse shape. In addition to the offset, we observed a scatter in the range of amplitude ratios for the

CPSR2 data similar in scatter range observed by [Kramer et al. \(1999\)](#) indicating that some unexplained causes of instability remain. The KS test between EPOS and PSRIX indicated the strong possibility of an increased noise floor with EPOS. Previous tests have been conducted comparing the non-dedispersing backend with the standard pulsar backend of EPOS, using a standard noise diode signal. There was a significant increase (up to a maximum factor of 4) observed, which explains the KS test result.

Previous studies have suggested several phenomena intrinsic to the pulsar which could explain the changes seen in the profile. [Kramer et al. \(1999\)](#) in their analysis suggested that mode changes alone cannot explain the variability seen. We note, however, that since this initial discussion, pulsars have been discovered to also exhibit significant profile changes on much longer timescales (see [Lyne et al. \(2010\)](#) or [Karastergiou et al. \(2011\)](#)). For the current analyses, we have used pulse profiles of the order close to 10^5 pulses where one expects the pulse profile to be stabilised if “classical mode changing” is happening (i.e. typically of the order of tens of minutes at most). However, the MKM analysis (see Figure 6.2) does not show bimodality in the metrics which most likely indicates that there is indeed no “classical” mode switching.

We note that we do not observe a monotonic variation in the profile across time unlike effects due to spin precession. The timescale of variation observed is also much lower compared to the typical timescales of profile evolution due to precession effects. [Kramer et al. \(1999\)](#) argued that a propagation effect through the pulsar magnetosphere would most likely explain the changing components of the profile. This was reiterated by the study of [Ramachandran & Kramer \(2003\)](#) where they analysed the polarisation properties of the pulsar. They observed a jump in the polarisation angle swing in a position coinciding with the leading component and suggested that the model proposed by [Hibschman & Arons \(2001\)](#) best fits the data. [Liu et al. \(2015\)](#) suggested that the two components could arise from different emission heights but would require further investigation. While these models help explain the emission mechanism of this pulsar, observing and timing PSR J1022+1001 with telescopes like MeerKAT and the upcoming SKA offers better sensitivity to study these subtle variations and further constrain the emission theories previously proposed.

6.9 Conclusions

We were able to demonstrate results indicating that PSR J1022+1001 has profile shape changes at long integration time scales. We modified methods suggested in preceding studies to make them more robust. We applied these methods not only on previously used data sets but also on new data from the latest Effelsberg pulsar backend. The EPOS data set specifically shows a higher degree of instability compared to the other data sets. We demonstrated that this is due to greater degree of noise in the back end chain of EPOS compared to PSRIX. The profile shape changes seen with CPSR2 show that instrumental errors cannot significantly change the pulse profile. We demonstrate that the calibration error variation is a tiny fraction of the total variations seen (up to maximum of 1 *per cent*). Thus, there are additional instabilities that cannot be ex-

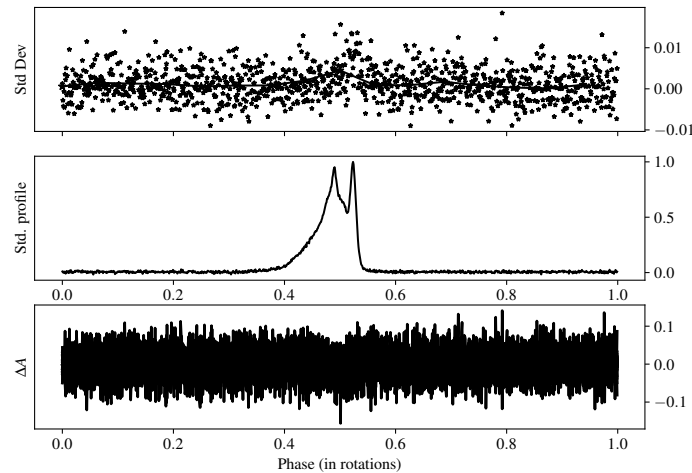


Figure 6.9: Results from applying the original Hotan et al. method on CPSR2 (uncalibrated) data with 5 min long integrations as used by Hotan et al.

plained fully by miscalibration. With the CPSR2 and PSRIX data, we tested the effect of interstellar scintillation across frequency and showed that the scale of this variation can at most account for 25 *per cent* of the observed variations. The results strongly suggest that intrinsic emission factors are the major contribution to the observed profile variation.

Appendix A: Original method analysis

We present here the results of applying the method described by Hotan et al. (2004a) on CPSR2 data. Results described in Figure 6.9 is obtained by applying the method on 5 min integration lengths of data as used by Hotan et al. (2004a). The bottom panel shows superposed difference profiles of the pulsar with a standard high S/N reference profile. The middle panel shows the mean profile. The top panel represents the standard deviation of per phase bin (marked as stars) in the difference profiles displayed in the bottom panel. The mean value of the standard deviation for every 64 bins is connected by lines to show trends in the data. We extended the analysis to the expanded CPSR2 dataset (calibrated and uncalibrated) with 20-30 min integration lengths. Figures 6.10 and 6.11 show the results for uncalibrated and calibrated data respectively which show a much more significant trend for variability as compared to the 5 min data.

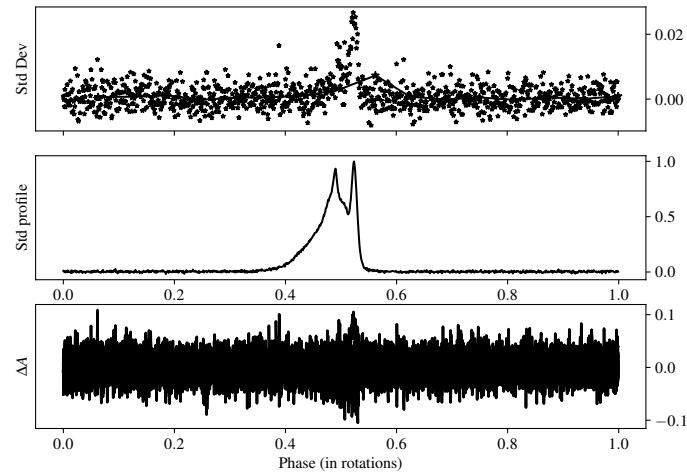


Figure 6.10: Results from applying the original Hotan et al. method on CPSR2 (uncalibrated) data with 20-30 min long integrations.

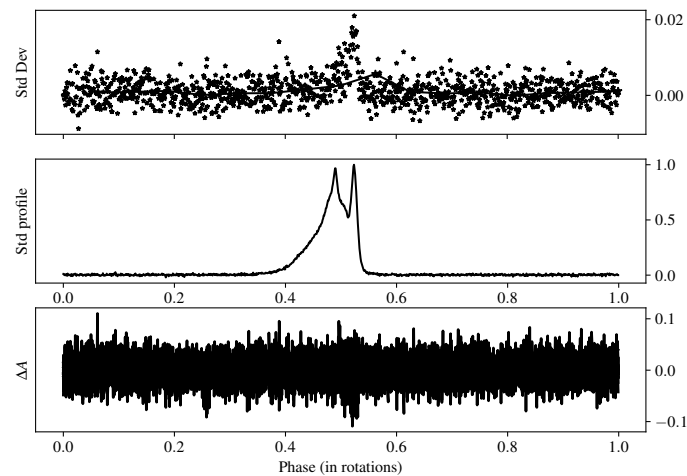


Figure 6.11: Results from applying the original Hotan et al. method on CPSR2 (calibrated) data with 20-30 min long integrations.

Summary and Future Prospects

Contents

7.1	Summary	191
7.2	Future work and prospects	192
7.2.1	Enhancements to the TRAPUM computing infrastructure	192
7.2.2	Pulsar searches of Terzan 5 using TRAPUM	194
7.2.3	Continuation of the MGPS survey	194
7.2.4	Expansion of candidate datasets and classifier models for SAP Data Intelligence	195
7.2.5	Extension of profile instability analyses and implications on timing	195
7.3	Preparing for the SKA	196

7.1 Summary

The goal of this thesis is to present novel techniques for pulsar data analysis that have implications in pulsar searching and timing. In Chapter 3, I described a new and unique computing infrastructure setup for large scale pulsar searching using the MeerKAT radio telescope. This setup exploits latest advancements in container orchestration technology and message brokering using open source tools like `Docker`, `Kubernetes` and `RabbitMQ` to provide enhanced and efficient data management and processing capabilities. In the second half of Chapter 3, I demonstrated a use-case for this system using observations of the globular cluster Terzan 5. While the acceleration searches did not reveal new pulsars in Terzan 5, the overall pipeline performance is encouraging to conduct similar searches in other observed epochs. The multibeam data offers near uniform sensitivity across a wide area of the cluster which is not achievable with single dish telescopes like GBT. This boosts the chances of finding pulsars in the outer realms of Terzan 5 that have possibly been flung away from the core of the cluster. Regular timing of future discoveries and known pulsars in Terzan 5 would provide further insight into the environment of this globular cluster. In Chapter 4, I introduced the MPIfR Galactic Plane Survey (MGPS) which is the most sensitive galactic plane survey being conducted in the Southern Hemisphere. The survey encompasses a unique arrangement where multiple key science projects including pulsar searching, Galactic magnetism as well as star formation studies are covered using the

same observed pointings within the survey region. Besides providing a system description and survey strategy, I also provided an overview of searches conducted with a pilot survey that has demonstrated a reasonable sensitivity to detecting known pulsars and also provided feedback for improvements moving forward in conducting the survey. I demonstrated a need for optimised survey beam tiling in order to ensure that pulsars are not missed due to poor coverage. Once the commissioning phase for S-Band receivers are complete, the S-Band component of the survey promises unprecedented sensitivity along the Galactic plane including the Galactic centre thus providing the best possibility yet to discover a pulsar orbiting Sgr A* and relativistic binary systems similar to the Double pulsar. In Chapter 5, I introduced SAP Data Intelligence, an industrial platform that provides data analytics tools and functionalities allowing for streamlined data processing. I demonstrated a use case for SAP Data Intelligence by employing existing machine learning classifier models on millions of pulsar candidates generated from the HTRU South low-lat survey. I also presented a comparative study between models using evaluation metrics used in determining the robustness of a typical machine learning classifier. In Chapter 6, I delved into characterising the long term pulse profile stability of PSR J1022+1001. Pulsar timing methods hinge on the assumption that average pulse profiles of pulsars generated from integrating up to 100,000 pulses are stable across long time periods (timescales of years) and are hence suitable for generating reference templates for further timing analyses. I demonstrated an anomaly in this assumption via PSR J1022+1001 which is a part of two pulsar timing arrays and has a rich history of conflicting opinions about the pulse profile stability over different timescales. I introduced a novel technique for characterising the profile stability by modifying previously existing methods for the same. I showed that the profile changes are primarily due to intrinsic instabilities rather than extrinsic effects. This analysis provides a platform for improving the understanding of pulse profile stability for a given pulsar. This can in turn give an insight not only about the pulsar but also about its impact on the sensitivity of pulsar timing arrays whose ultimate goal is to detect nanohertz gravitational waves.

7.2 Future work and prospects

7.2.1 Enhancements to the TRAPUM computing infrastructure

At the time of writing of this thesis, the TRAPUM survey has been running for nearly one year with 150 hours of the allotted telescope time complete. The APSUSE cluster (see Section 3.2.3) currently has a 3.5 PB storage capacity. Assuming 288 tied-array beams are recorded, each 4 hour observation records ~ 200 TB of data. This implies that the cluster can record up to 72 hours straight. Thus, upgrades in storage would help increase the observing cadence. Apart from storage, the RAM capability also needs to be enhanced to enable larger FFT operations (2^{28} and above) with the current pipeline setup (see Section 3.4.3.2).

While an end-to-end computing infrastructure is in place, there is scope for improving multiple components within the system. In the first year of operation, many

sources of RFI have been identified and excision methods (see Section 3.4.3.3 from Chapter 3) have accordingly been implemented to reduce number of candidates generated as well as inspected. However, there is scope for further examination of current methods. There exist a vast number of unidentified RFI sources that show up as candidates and would require careful examination to discern their origin. Besides this, the current clustering scheme for multi-beam candidate filtering (see Section 3.4.3.3) would benefit from using characterisation of the point spread function (PSF) of the tied-array beams. This ensures that the power of the main lobe as well as side lobes are accounted for. The *Mosaic* package (Chen et al., submitted) provides this information and can be incorporated as an input for candidate filtering. Eventually, multi-beam candidate filtering could potentially be treated as a machine learning problem where a candidate generated showing certain signatures for a set of handcrafted features would help segregate the pulsars from non-pulsar candidates (e.g. Connor & van Leeuwen, 2018). Extensive quality assurance testing of the existing pipeline would help refine and converge on the exact features to be used in this scenario. Currently the acceleration search pipeline (see Section 3.4.3.2) uses a time domain resampling algorithm. The *AstroAccelerate*¹ library which implements a Fourier domain acceleration algorithm for GPUs (Dimoudi et al., 2018; Carels et al., 2019) is a potential replacement for the current PEASOUP algorithm in the long term. Besides this, the *CHEETAH*² and *PANDA*³ platforms that are already used in the infrastructure for MeerTRAP (Rajwade et al., 2021) could be incorporated into the search setup for TRAPUM. This could minimise possible overheads due to file handling while searching the data.

On the post-folding candidate sifting aspect, an ensemble of machine learning classifiers can be used for improved decision making. The existing PICS model can be improved by retraining the model with candidates accumulated from the survey. A similar scheme can be applied to new classifiers (existing and in development) that are incorporated into the setup. One such classifier is a semi-supervised general adversarial network (Balakrishnan et al., 2020) which uses a mixture of labelled and unlabelled data for training. This network has been shown to outperform supervised machine learning classifiers for a limited number of candidates (order of hundreds to early thousands). Given that the TRAPUM survey is at an early stage, training sets can be prepared based on existing candidates from TRAPUM to generate a model within a limited time. In the long-term, the aim is to inject pulsar signals into TRAPUM data and retrain existing classifiers to enhance the performance of candidate sifting. This setup allows the injection of pulsar signals with different spin period and duty cycles and thus improving the robustness of classification across the known pulsar parameter space.

¹<https://github.com/AstroAccelerateOrg/astro-accelerate>

²<https://gitlab.com/SKA-TDT/cheetah>

³<https://gitlab.com/SKA-TDT/panda>

7.2.2 Pulsar searches of Terzan 5 using TRAPUM

The acceleration search conducted on Terzan 5 with TRAPUM (see Section 3.5) revealed several weak candidates that are yet to be confirmed. The aim is to continue searching on other observing runs to retrieve potential new discoveries and confirm existing candidates. Terzan 5 houses the highest number of confirmed millisecond pulsars for a globular cluster (39). These discoveries are a consequence of several innovative approaches that have been trialed on this cluster (see Section 3.5.2). The boost in sensitivity provided by MeerKAT complemented by the unique capabilities of the TRAPUM software and hardware improves the chances of a new discovery using the existing approaches.

Apart from already implemented acceleration searches, compact binary pulsars can be searched using jerk searches (Andersen & Ransom, 2018) and coherent Keplerian orbital searches (Allen et al., 2013). Apart from extending the dimensions of the searches, refined searches can be conducted using known information from imaging studies. For example, the recent work by Urquhart et al. (2020) has revealed three promising eclipsing redback candidates. Using the localisation capability of TRAPUM, certain tied array beams can be searched after incorporating the constraints on the binary pulsar parameters like orbital period and spin period as reported by Urquhart et al. (2020). Any new discoveries would also be followed up with extensive timing campaigns with MeerKAT and other telescopes. The parameters obtained from timing can also help update the solutions to the cluster parameters as given by Prager et al. (2017).

7.2.3 Continuation of the MGPS survey

The MGPS has been shown to be more sensitive than previous southern sky galactic plane surveys in terms of the radiometer equation estimates (see Section 4.4). However, these estimates do not account for impact of red-noise and RFI. Thus, I plan to inject synthetic pulsar signals into the data and retrieve their S/N as detected by the pipeline, similar to the work done by Lazarus et al. (2015). This gives a realistic feedback of the sensitivity across the searched parameter space under different RFI and noise conditions. With continued observations and processing of the MGPS, new discoveries would help in multiple ways. They would help investigate the difference between expected and actual observed pulsar populations. They would also help draw a direct comparison with previous surveys with Parkes like PMPS and HTRU South low-lat, particularly where the regions overlap with MGPS. The short integration times for the surveys (maximum of 20 min) allow for minimal loss in sensitivity to compact binary pulsars. This boosts the opportunity for finding relativistic binaries like PSR J1757-1854 and the elusive pulsar-black hole binaries. With a significant fraction of time allotted for observing the Galactic Centre, the MGPS survey also provides the best sensitivity yet to discover pulsars in a tight orbit with Sgr A* and potentially open a new regime of fundamental tests of gravity with pulsars. As stated in Chapter 4, the commensality between different science cases will also be exploited to refine the search

schemes.

7.2.4 Expansion of candidate datasets and classifier models for SAP Data Intelligence

The SAP Data Intelligence database currently holds information on candidates from a fraction of the HTRU South low-lat survey generated from an acceleration search pipeline. It also hosts multiple models from the PICS classifier (Zhu et al., 2014). The system can be expanded to have candidates from different surveys. Given that a memorandum of understanding (MoU) is established between different observatories and SAP, candidates generated from pulsar surveys could be quickly streamed to the SAP platform and put through the existing pipeline schematic. One such example is the candidates generated from the two MeerKAT surveys discussed in this thesis viz. TRAPUM and MGPS. Both surveys are impacted by the large scale candidate sifting problem and would thus benefit from the SAP platform.

The platform can also be expanded to include more classifier models. Some examples include SPINN (Morello et al., 2014) and Semi Supervised General Adversarial Networks (Balakrishnan et al., 2020). The pipeline can also be integrated with human ranking algorithms like PEACE (Lee et al., 2013). This allows for a direct comparison of human vs machine learning models on candidates at a large scale. The visual analytics tool-kit applied on large sets of candidates will also allow for developing new features and better resolve pulsars and RFI in a unique hyper-parameter space.

7.2.5 Extension of profile instability analyses and implications on timing

The profile instability analysis of PSR J1022+1001 (see Chapter 6) can be expanded to other pulsars and particularly those belonging to pulsar timing arrays. This would provide a comprehensive study of which pulsars are significantly susceptible to timing instabilities due to profile variations. This would be crucial in identifying sources of noise and discrepancy that could hamper the possibility of detection of nanohertz gravitational waves.

Furthermore, precision timing of PSR J1022+1001 using the same data sets as used in the analysis in Chapter 6 could provide a deeper insight into the correlation between the profile variations and timing residuals. The nature of the emission of this pulsar can also be better resolved by analysing long and sensitive data sets. A recent paper on single pulse studies of PSR J1022+1001 (Feng et al., 2020) with 20 min of observing with FAST highlights the numerous possibilities to conduct such a careful analysis. Extending the analysis by Feng et al. (2020) for longer integration times and studying the profile evolution across bandwidth could also help improve the current known noise parameters of the pulsar. Besides this, high fidelity polarimetry of this pulsar using data from FAST and MeerKAT could also enable and constrain possible emission models that explain the timescales of the instability observed.

7.3 Preparing for the SKA

The latest generation radio telescopes namely MeerKAT and FAST are currently the most sensitive telescopes in their respective hemispheres. In this thesis, the MeerKAT telescope has been demonstrated as a remarkable prospect for discovering pulsars in terms of targeted searches (TRAPUM) as well as blind searches (MGPS). The telescope has already been responsible for multiple pulsar discoveries⁴ and promises to discover more in the lead up to the Square Kilometre Array (SKA).

The SKA consists of two telescopes namely SKA Low and SKA Mid. SKA Low consists of thousands of dipoles with a frequency coverage spanning 50-350 MHz. The SKA Mid consists of 197 dishes (in the first phase) operating between 350 MHz to 14 GHz, thus making it a factor of approx. 3 higher than the current number of dishes with MeerKAT. This would make SKA the most sensitive radio telescope/interferometer ever built. While the scientific prospects for the SKA include $\sim 30,000$ new potential pulsar discoveries (e.g. Keane et al., 2015) including pulsars beyond the Galaxy and the Magellanic clouds, the practicality of data management and processing are a major concern. Pulsar searches with the SKA will rely on multi-beam beamforming with up to 500 synthesised beams on sky for SKA low and 1500 for SKA Mid (Levin et al., 2018). Besides this, the SKA would operate as a real-time system with observation and processing happening in tandem. It is estimated that the computing demands are at least a factor of 30-40 higher than TRAPUM and MGPS.

MeerKAT being a precursor to the SKA has provided the scope for learning and implementing the necessary procedures for large scale operations in terms of observations as well as data processing. The work in this thesis has provided a platform for understanding different problems that arise as a consequence of running a large scale multi-beam pulsar survey with an interferometer. The large number of beams used in each TRAPUM/MGPS observation has proven useful for spatial filtering, thus reducing the impact of RFI and improving candidate sifting (see Section 2.2.9). Further investment in exploiting multi-beam information is important given that the aim with SKA is to apply such sifting routines after searching (e.g. Levin et al., 2018). This is because real-time multibeam RFI excision prior to searching would not be feasible for thousands of beams. Apart from this, the APSUSE cluster has allowed us to examine how pulsar processing can be efficiently mapped to high performance computing. This includes efficient usage of CPUs and GPUs within a computing node, asynchronous processing, memory management and usage of distributed filesystems like BeeGFS which ensure a significant boost in read/write speeds and reduce I/O overheads. Besides this, the performances of processing pipelines on the cluster have been observed to be primarily I/O bound. These I/O limitations are dependent on the acceleration range. For making a large scale search feasible with the SKA, it is not only important to have a large computing capacity but also invest in methods that can reduce the number of read/write operations. For this reason, the pulsar searches with the SKA should be conducted on RAM throughout. This way, there is a seamless transition between

⁴<http://www.trapum.org/discoveries.html>

different routines within pulsar searching. Furthermore, it also reduces the number of intermediate data products to deal with. A similar approach is also needed for dealing with the pulsar candidate classification problem ensuring a high degree of accuracy with minimal compromise on compute speed. Overall, the hope is that a deep focus on issues similar to the work presented here and elsewhere (see [Levin et al. \(2018\)](#) and references therein) would prove effective for ensuring the smooth operation of the pulsar surveys with SKA once it is fully operational.

Bibliography

- Abbate, F., Possenti, A., Ridolfi, A., et al. 2018, MNRAS, 481, 627, doi: [10.1093/mnras/sty2298](https://doi.org/10.1093/mnras/sty2298) (Cited on page 95.)
- Abbate, F., Possenti, A., Tiburzi, C., et al. 2020, Nature Astronomy, 4, 704, doi: [10.1038/s41550-020-1030-6](https://doi.org/10.1038/s41550-020-1030-6) (Cited on pages 32 and 95.)
- Abbott, B., Abbott, R., Abbott, T. D., et al. 2017, Phys. Rev. Lett., 119, 161101, doi: [10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101) (Cited on pages 38 and 42.)
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102, doi: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102) (Cited on page 38.)
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020, ApJ, 896, L44, doi: [10.3847/2041-8213/ab960f](https://doi.org/10.3847/2041-8213/ab960f) (Cited on page 38.)
- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010, aap, 524, A75, doi: [10.1051/0004-6361/201014458](https://doi.org/10.1051/0004-6361/201014458) (Cited on page 109.)
- Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, The Astrophysical Journal Supplement Series, 247, 33, doi: [10.3847/1538-4365/ab6bcb](https://doi.org/10.3847/1538-4365/ab6bcb) (Cited on page 94.)
- Ajello, M., Atwood, W. B., Axelsson, M., et al. 2021, Nature Astronomy, doi: [10.1038/s41550-020-01287-8](https://doi.org/10.1038/s41550-020-01287-8) (Cited on page 29.)
- Allen, B., Knispel, B., Cordes, J. M., et al. 2013, The Astrophysical Journal, 773, 91, doi: [10.1088/0004-637x/773/2/91](https://doi.org/10.1088/0004-637x/773/2/91) (Cited on pages 77, 123 and 194.)
- Alpar, M. A., Cheng, A. F., Ruderman, M. A., & Shaham, J. 1982, Nature, 300, 728, doi: [10.1038/300728a0](https://doi.org/10.1038/300728a0) (Cited on pages 29 and 83.)
- Andersen, B. C., & Ransom, S. M. 2018, ApJ, 863, L13, doi: [10.3847/2041-8213/aad59f](https://doi.org/10.3847/2041-8213/aad59f) (Cited on pages 32, 76, 77, 110, 122 and 194.)
- Anderson, D. P., Christensen, C., & Allen, B. 2006, in SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, 33–33, doi: [10.1109/SC.2006.24](https://doi.org/10.1109/SC.2006.24) (Cited on page 77.)
- Anderson, S. B., Gorham, P. W., Kulkarni, S. R., Prince, T. A., & Wolszczan, A. 1990, Nature, 346, 42, doi: [10.1038/346042a0](https://doi.org/10.1038/346042a0) (Cited on page 84.)
- Antoniadis, J. 2014, ApJ, 797, L24, doi: [10.1088/2041-8205/797/2/L24](https://doi.org/10.1088/2041-8205/797/2/L24) (Cited on page 43.)
- Antoniadis, J., Kaplan, D. L., Stovall, K., et al. 2016, The Astrophysical Journal, 830, 36, doi: [10.3847/0004-637x/830/1/36](https://doi.org/10.3847/0004-637x/830/1/36) (Cited on page 41.)

- Antoniadis, J., van Kerkwijk, M. H., Koester, D., et al. 2012, *MNRAS*, 423, 3316, doi: [10.1111/j.1365-2966.2012.21124.x](https://doi.org/10.1111/j.1365-2966.2012.21124.x) (Cited on page 31.)
- Antoniadis, J., Freire, P. C. C., Wex, N., et al. 2013, *Science*, 340, 448, doi: [10.1126/science.1233232](https://doi.org/10.1126/science.1233232) (Cited on pages 21, 36, 37 and 171.)
- Antoniou, V., & Zezas, A. 2016, *MNRAS*, 459, 528, doi: [10.1093/mnras/stw167](https://doi.org/10.1093/mnras/stw167) (Cited on page 95.)
- Archibald, A. M., Stairs, I. H., Ransom, S. M., et al. 2009, *Science*, 324, 1411, doi: [10.1126/science.1172740](https://doi.org/10.1126/science.1172740) (Cited on page 28.)
- Archibald, A. M., Gusinskaia, N. V., Hessels, J. W. T., et al. 2018, *Nature*, 559, 73, doi: [10.1038/s41586-018-0265-1](https://doi.org/10.1038/s41586-018-0265-1) (Cited on page 38.)
- Arzoumanian, Z., Gendreau, K., & NICER Team. 2012, in *American Astronomical Society Meeting Abstracts*, Vol. 219, American Astronomical Society Meeting Abstracts #219, 249.05 (Cited on pages 21 and 36.)
- Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, *ApJ*, 905, L34, doi: [10.3847/2041-8213/abd401](https://doi.org/10.3847/2041-8213/abd401) (Cited on pages 39 and 96.)
- Baade, W., & Zwicky, F. 1934, *Proceedings of the National Academy of Sciences*, 20, 259, doi: [10.1073/pnas.20.5.259](https://doi.org/10.1073/pnas.20.5.259) (Cited on page 20.)
- Backer, D. 1975, *A&A*, 43, 395 (Cited on page 35.)
- Backer, D. C. 1970, *Nature*, 228, 42, doi: [10.1038/228042a0](https://doi.org/10.1038/228042a0) (Cited on pages 30 and 172.)
- Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M., & Goss, W. M. 1982, *Nature*, 300, 615, doi: [10.1038/300615a0](https://doi.org/10.1038/300615a0) (Cited on pages 41, 80 and 83.)
- Bagchi, M., Lorimer, D. R., & Wolfe, S. 2013, *MNRAS*, 432, 1303, doi: [10.1093/mnras/stt559](https://doi.org/10.1093/mnras/stt559) (Cited on page 77.)
- Bahramian, A., Strader, J., Chomiuk, L., et al. 2018, *The Astrophysical Journal*, 864, 28, doi: [10.3847/1538-4357/aad68b](https://doi.org/10.3847/1538-4357/aad68b) (Cited on page 109.)
- Bailes, M., Jameson, A., Flynn, C., et al. 2017, *Publications of the Astronomical Society of Australia*, 34, e045, doi: [10.1017/pasa.2017.39](https://doi.org/10.1017/pasa.2017.39) (Cited on page 86.)
- Bailes, M., Jameson, A., Abbate, F., et al. 2020, *Publications of the Astronomical Society of Australia*, 37, e028, doi: [10.1017/pasa.2020.19](https://doi.org/10.1017/pasa.2020.19) (Cited on pages 90, 110 and 113.)
- Balakrishnan, V., Champion, D., & Barr, E. 2020, arXiv e-prints, arXiv:2010.07457. <https://arxiv.org/abs/2010.07457> (Cited on pages 152, 158, 193 and 195.)

- Bannister, K. W., Shannon, R. M., Macquart, J. P., et al. 2017, *ApJ*, 841, L12, doi: [10.3847/2041-8213/aa71ff](https://doi.org/10.3847/2041-8213/aa71ff) (Cited on page 85.)
- Bansal, K., Taylor, G. B., Stovall, K., & Dowell, J. 2019, *ApJ*, 875, 146, doi: [10.3847/1538-4357/ab0d8f](https://doi.org/10.3847/1538-4357/ab0d8f) (Cited on page 34.)
- Barr, E. D. 2013, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn. <http://hdl.handle.net/20.500.11811/5664> (Cited on pages 48, 52, 55 and 63.)
- Barr, E. D., Freire, P. C. C., Kramer, M., et al. 2017, *MNRAS*, 465, 1711, doi: [10.1093/mnras/stw2947](https://doi.org/10.1093/mnras/stw2947) (Cited on page 41.)
- Barr, E. D., Champion, D. J., Kramer, M., et al. 2013, *MNRAS*, 435, 2234, doi: [10.1093/mnras/stt1440](https://doi.org/10.1093/mnras/stt1440) (Cited on pages 41, 51, 71, 80, 82, 150 and 162.)
- Barsdell, B. R., Bailes, M., Barnes, D. G., & Fluke, C. J. 2012, *MNRAS*, 422, 379, doi: [10.1111/j.1365-2966.2012.20622.x](https://doi.org/10.1111/j.1365-2966.2012.20622.x) (Cited on pages 57 and 78.)
- Barsdell, B. R., Barnes, D. G., & Fluke, C. J. 2010, *MNRAS*, 408, 1936, doi: [10.1111/j.1365-2966.2010.17257.x](https://doi.org/10.1111/j.1365-2966.2010.17257.x) (Cited on pages 57 and 97.)
- Bassa, C. G., Pleunis, Z., & Hessels, J. W. T. 2017a, *Astronomy and Computing*, 18, 40, doi: [10.1016/j.ascom.2017.01.004](https://doi.org/10.1016/j.ascom.2017.01.004) (Cited on pages 57 and 59.)
- Bassa, C. G., Pleunis, Z., Hessels, J. W. T., et al. 2017b, *ApJ*, 846, L20, doi: [10.3847/2041-8213/aa8400](https://doi.org/10.3847/2041-8213/aa8400) (Cited on pages 36 and 58.)
- Bates, S. D., Lorimer, D. R., Rane, A., & Swiggum, J. 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 2893â2902, doi: [10.1093/mnras/stu157](https://doi.org/10.1093/mnras/stu157) (Cited on pages 96 and 135.)
- Bates, S. D., Bailes, M., Barsdell, B. R., et al. 2012, *MNRAS*, 427, 1052, doi: [10.1111/j.1365-2966.2012.22042.x](https://doi.org/10.1111/j.1365-2966.2012.22042.x) (Cited on pages 155 and 157.)
- Baumgardt, H., & Hilker, M. 2018, *MNRAS*, 478, 1520, doi: [10.1093/mnras/sty1057](https://doi.org/10.1093/mnras/sty1057) (Cited on page 109.)
- Baym, G., Pethick, C., & Pines, D. 1969, *Nature*, 224, 673, doi: [10.1038/224673a0](https://doi.org/10.1038/224673a0) (Cited on page 21.)
- Becker, W., Kramer, M., & Sesana, A. 2018, *Space Sci. Rev.*, 214, 30, doi: [10.1007/s11214-017-0459-0](https://doi.org/10.1007/s11214-017-0459-0) (Cited on page 40.)
- Berezina, M. 2020, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn (Cited on page 105.)
- Bethapudi, S., & Desai, S. 2018, *Astronomy and Computing*, 23, 15, doi: [10.1016/j.ascom.2018.02.002](https://doi.org/10.1016/j.ascom.2018.02.002) (Cited on page 157.)

- Bhat, N. D. R., Cordes, J. M., Camilo, F., Nice, D. J., & Lorimer, D. R. 2004, *ApJ*, 605, 759, doi: [10.1086/382680](https://doi.org/10.1086/382680) (Cited on pages 34 and 55.)
- Bhattacharya, D., & van den Heuvel, E. P. J. 1991, *Phys. Rep.*, 203, 1, doi: [10.1016/0370-1573\(91\)90064-S](https://doi.org/10.1016/0370-1573(91)90064-S) (Cited on page 28.)
- Bhattacharyya, B., Cooper, S., Malenta, M., et al. 2016, *ApJ*, 817, 130, doi: [10.3847/0004-637X/817/2/130](https://doi.org/10.3847/0004-637X/817/2/130) (Cited on page 85.)
- Bochenek, C. D., Ravi, V., Belov, K. V., et al. 2020, A fast radio burst associated with a Galactic magnetar. <https://arxiv.org/abs/2005.10828> (Cited on page 43.)
- Bogdanov, S., Bahramian, A., Heinke, C. O., & Ransom, S. M. 2020, arXiv e-prints, arXiv:2012.12944. <https://arxiv.org/abs/2012.12944> (Cited on page 109.)
- Bower, G., Chatterjee, S., Cordes, J., et al. 2019, *BAAS*, 51, 438 (Cited on page 40.)
- Britton, M. C. 2000, *ApJ*, 532, 1240, doi: [10.1086/308595](https://doi.org/10.1086/308595) (Cited on pages 24 and 175.)
- Broekema, P. C., Mol, J. J. D., Nijboer, R., et al. 2018, *Astronomy and Computing*, 23, 180, doi: [10.1016/j.ascom.2018.04.006](https://doi.org/10.1016/j.ascom.2018.04.006) (Cited on page 86.)
- Burgay, M., D'Amico, N., Possenti, A., et al. 2003, *Nature*, 426, 531â533, doi: [10.1038/nature02124](https://doi.org/10.1038/nature02124) (Cited on pages 28, 38 and 77.)
- Burns, W. R., & Clark, B. G. 1969, *A&A*, 2, 280 (Cited on page 80.)
- Caballero, R. N., Guo, Y. J., Lee, K. J., et al. 2018, *MNRAS*, 481, 5501, doi: [10.1093/mnras/sty2632](https://doi.org/10.1093/mnras/sty2632) (Cited on page 40.)
- Cadelano, M., Ransom, S. M., Freire, P. C. C., et al. 2018, *The Astrophysical Journal*, 855, 125, doi: [10.3847/1538-4357/aaac2a](https://doi.org/10.3847/1538-4357/aaac2a) (Cited on pages 110, 113, 118 and 122.)
- Cameron, A. D., Barr, E. D., Champion, D. J., Kramer, M., & Zhu, W. W. 2017, *MNRAS*, 468, 1994, doi: [10.1093/mnras/stx589](https://doi.org/10.1093/mnras/stx589) (Cited on page 78.)
- Cameron, A. D., Champion, D. J., Kramer, M., et al. 2018, *MNRAS*, 475, L57, doi: [10.1093/mnrasl/sly003](https://doi.org/10.1093/mnrasl/sly003) (Cited on pages 38, 71, 75 and 82.)
- Cameron, A. D., Champion, D. J., Bailes, M., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 493, 1063â1087, doi: [10.1093/mnras/staa039](https://doi.org/10.1093/mnras/staa039) (Cited on pages 75, 82, 133, 160 and 164.)
- Camilo, F., Lorimer, D. R., Freire, P., Lyne, A. G., & Manchester, R. N. 2000, *apjl*, 535, 975, doi: [10.1086/308859](https://doi.org/10.1086/308859) (Cited on pages 70, 71 and 84.)
- Camilo, F., Nice, D. J., Shrauner, J. A., & Taylor, J. H. 1996, *ApJ*, 469, 819, doi: [10.1086/177829](https://doi.org/10.1086/177829) (Cited on pages 172 and 173.)
- Camilo, F., Ransom, S. M., Chatterjee, S., Johnston, S., & Demorest, P. 2012, *ApJ*, 746, 63, doi: [10.1088/0004-637X/746/1/63](https://doi.org/10.1088/0004-637X/746/1/63) (Cited on page 30.)

- Camilo, F., Ransom, S. M., Halpern, J. P., & Reynolds, J. 2007, *The Astrophysical Journal*, 666, L93, doi: [10.1086/521826](https://doi.org/10.1086/521826) (Cited on page 29.)
- Camilo, F., Ransom, S. M., Halpern, J. P., et al. 2006, *Nature*, 442, 892â895, doi: [10.1038/nature04986](https://doi.org/10.1038/nature04986) (Cited on page 29.)
- Camilo, F., Kerr, M., Ray, P. S., et al. 2015, *ApJ*, 810, 85, doi: [10.1088/0004-637X/810/2/85](https://doi.org/10.1088/0004-637X/810/2/85) (Cited on pages 41 and 94.)
- Canizares, C. R., & Neighbours, J. E. 1975, *ApJ*, 199, L97, doi: [10.1086/181857](https://doi.org/10.1086/181857) (Cited on page 83.)
- Carels, C., Adámek, K., Novotný, J., & Armour, W. 2019, arXiv e-prints, arXiv:1912.07704. <https://arxiv.org/abs/1912.07704> (Cited on page 193.)
- Cerda-Duran, P., & Elias-Rosa, N. 2018, *Astrophysics and Space Science Library*, 1â56, doi: [10.1007/978-3-319-97616-7_1](https://doi.org/10.1007/978-3-319-97616-7_1) (Cited on page 21.)
- Chadwick, J. 1932, *Nature*, 129, 312, doi: [10.1038/129312a0](https://doi.org/10.1038/129312a0) (Cited on page 20.)
- Champion, D. J., Ransom, S. M., Lazarus, P., et al. 2008, *Science*, 320, 1309, doi: [10.1126/science.1157580](https://doi.org/10.1126/science.1157580) (Cited on pages 31 and 41.)
- Chandler, A. M. 2003, PhD thesis, CALIFORNIA INSTITUTE OF TECHNOLOGY (Cited on pages 76 and 110.)
- Chandrasekhar, S. 1931, *ApJ*, 74, 81, doi: [10.1086/143324](https://doi.org/10.1086/143324) (Cited on page 20.)
- Chen, S., Sesana, A., & Conselice, C. J. 2019, *MNRAS*, 488, 401, doi: [10.1093/mnras/stz1722](https://doi.org/10.1093/mnras/stz1722) (Cited on page 38.)
- Chen, Z.-C., Yuan, C., & Huang, Q.-G. 2020, *Phys. Rev. Lett.*, 124, 251101, doi: [10.1103/PhysRevLett.124.251101](https://doi.org/10.1103/PhysRevLett.124.251101) (Cited on page 38.)
- Chennamangalam, J., Lorimer, D., Mandel, I., & Bagchi, M. 2013, *Mon. Not. Roy. Astron. Soc.*, 431, 874, doi: [10.1093/mnras/stt205](https://doi.org/10.1093/mnras/stt205) (Cited on page 109.)
- CHIME/ FRB Collaboration, Amiri, M., Bandura, K., et al. 2019, *Nature*, 566, 230, doi: [10.1038/s41586-018-0867-7](https://doi.org/10.1038/s41586-018-0867-7) (Cited on page 78.)
- CHIME/FRB Collaboration, Amiri, M., Bandura, K., et al. 2018, *ApJ*, 863, 48, doi: [10.3847/1538-4357/aad188](https://doi.org/10.3847/1538-4357/aad188) (Cited on page 57.)
- CHIME/FRB Collaboration, Andersen, B. C., Bandura, K. M., et al. 2020, *Nature*, 587, 54, doi: [10.1038/s41586-020-2863-y](https://doi.org/10.1038/s41586-020-2863-y) (Cited on page 43.)
- Clark, G. W. 1975, *ApJ*, 199, L143, doi: [10.1086/181869](https://doi.org/10.1086/181869) (Cited on page 83.)
- Clark, G. W., Markert, T. H., & Li, F. K. 1975, *ApJ*, 199, L93, doi: [10.1086/181856](https://doi.org/10.1086/181856) (Cited on page 83.)

- Clifton, T. R., & Lyne, A. G. 1986, *Nature*, 320, 43, doi: [10.1038/320043a0](https://doi.org/10.1038/320043a0) (Cited on pages 80 and 81.)
- Clifton, T. R., Lyne, A. G., Jones, A. W., McKenna, J., & Ashworth, M. 1992, *MNRAS*, 254, 177, doi: [10.1093/mnras/254.2.177](https://doi.org/10.1093/mnras/254.2.177) (Cited on page 81.)
- Coenen, T., van Leeuwen, J., Hessels, J. W. T., et al. 2014, *A&A*, 570, A60, doi: [10.1051/0004-6361/201424495](https://doi.org/10.1051/0004-6361/201424495) (Cited on page 71.)
- Cognard, I., Guillemot, L., Johnson, T. J., et al. 2011, *The Astrophysical Journal*, 732, 47, doi: [10.1088/0004-637x/732/1/47](https://doi.org/10.1088/0004-637x/732/1/47) (Cited on page 94.)
- Connor, L., & van Leeuwen, J. 2018, *AJ*, 156, 256, doi: [10.3847/1538-3881/aae649](https://doi.org/10.3847/1538-3881/aae649) (Cited on page 193.)
- Cooley, J. W., & Tukey, J. W. 1965, *Math. Comput.*, 19, 297, doi: [10.1090/S0025-5718-1965-0178586-1](https://doi.org/10.1090/S0025-5718-1965-0178586-1) (Cited on pages 49, 56 and 58.)
- Cordes, J. 2013a, *Classical and Quantum Gravity*, 30, 224002, doi: [10.1088/0264-9381/30/22/224002](https://doi.org/10.1088/0264-9381/30/22/224002) (Cited on page 171.)
- Cordes, J. M. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 278, *Single-Dish Radio Astronomy: Techniques and Applications*, ed. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter, 227–250 (Cited on page 52.)
- Cordes, J. M. 2013b, *The Astrophysical Journal*, 775, 47, doi: [10.1088/0004-637x/775/1/47](https://doi.org/10.1088/0004-637x/775/1/47) (Cited on page 172.)
- Cordes, J. M., & Lazio, T. J. W. 2002, arXiv e-prints, astro. <https://arxiv.org/abs/astro-ph/0207156> (Cited on pages 33, 54 and 186.)
- Cordes, J. M., & Lazio, T. J. W. 2002, NE2001.I. A New Model for the Galactic Distribution of Free Electrons and its Fluctuations. <https://arxiv.org/abs/astro-ph/0207156> (Cited on page 135.)
- Cordes, J. M., & Shannon, R. M. 2008, *The Astrophysical Journal*, 682, 1152â1165, doi: [10.1086/589425](https://doi.org/10.1086/589425) (Cited on page 30.)
- Cordes, J. M., Freire, P. C. C., Lorimer, D. R., et al. 2006, *ApJ*, 637, 446, doi: [10.1086/498335](https://doi.org/10.1086/498335) (Cited on pages 108, 133, 157, 162, 165 and 168.)
- Cromartie, H. T., Camilo, F., Kerr, M., et al. 2016, *apjl*, 819, 34, doi: [10.3847/0004-637X/819/1/34](https://doi.org/10.3847/0004-637X/819/1/34) (Cited on page 94.)
- Cromartie, H. T., Fonseca, E., Ransom, S. M., et al. 2019, *Nature Astronomy*, 4, 72â76, doi: [10.1038/s41550-019-0880-2](https://doi.org/10.1038/s41550-019-0880-2) (Cited on page 37.)
- D’Amico, N., Lyne, A. G., Manchester, R. N., Possenti, A., & Camilo, F. 2001, *ApJ*, 548, L171, doi: [10.1086/319096](https://doi.org/10.1086/319096) (Cited on page 84.)

- Damour, T., & Deruelle, N. 1985, *Ann. Inst. Henri Poincaré Phys. Théor.*, 43, 107 (Cited on page 38.)
- D'Angelo, C. 2013, in *Neutron Stars and Pulsars: Challenges and Opportunities after 80 years*, ed. J. van Leeuwen, Vol. 291, 155–158, doi: [10.1017/S1743921312023460](https://doi.org/10.1017/S1743921312023460) (Cited on page 29.)
- Davidson, K., & Ostriker, J. P. 1973, *ApJ*, 179, 585, doi: [10.1086/151897](https://doi.org/10.1086/151897) (Cited on page 28.)
- Davies, J. G., & Large, M. I. 1970, *MNRAS*, 149, 301, doi: [10.1093/mnras/149.4.301](https://doi.org/10.1093/mnras/149.4.301) (Cited on page 80.)
- Davies, J. G., Lyne, A. G., & Seiradakis, J. H. 1977, *Monthly Notices of the Royal Astronomical Society*, 179, 635, doi: [10.1093/mnras/179.4.635](https://doi.org/10.1093/mnras/179.4.635) (Cited on page 81.)
- De, K., & Gupta, Y. 2016, *Experimental Astronomy*, 41, 67, doi: [10.1007/s10686-015-9476-8](https://doi.org/10.1007/s10686-015-9476-8) (Cited on page 54.)
- De, K., Kasliwal, M. M., Ofek, E. O., et al. 2018, *Science*, 362, 201, doi: [10.1126/science.aas8693](https://doi.org/10.1126/science.aas8693) (Cited on pages 41 and 42.)
- DeCesar, M. E., Ransom, S. M., Kaplan, D. L., Ray, P. S., & Geller, A. M. 2015, *apjl*, 807, L23, doi: [10.1088/2041-8205/807/2/L23](https://doi.org/10.1088/2041-8205/807/2/L23) (Cited on pages 84 and 95.)
- Deneva, J. S., Cordes, J. M., & Lazio, T. J. W. 2009, *ApJ*, 702, L177, doi: [10.1088/0004-637X/702/2/L177](https://doi.org/10.1088/0004-637X/702/2/L177) (Cited on page 40.)
- Deneva, J. S., Stovall, K., McLaughlin, M. A., et al. 2013, *ApJ*, 775, 51, doi: [10.1088/0004-637X/775/1/51](https://doi.org/10.1088/0004-637X/775/1/51) (Cited on page 41.)
- Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, *mnras*, 458, 3341, doi: [10.1093/mnras/stw483](https://doi.org/10.1093/mnras/stw483) (Cited on page 96.)
- Desvignes, G., Eatough, R. P., Pen, U. L., et al. 2018, *The Astrophysical Journal*, 852, L12, doi: [10.3847/2041-8213/aaa2f8](https://doi.org/10.3847/2041-8213/aaa2f8) (Cited on page 40.)
- Desvignes, G., Kramer, M., Lee, K., et al. 2019, *Science*, 365, 1013, doi: [10.1126/science.aav7272](https://doi.org/10.1126/science.aav7272) (Cited on pages 22, 24 and 172.)
- Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, *Proceedings of the IEEE*, 97, 1482, doi: [10.1109/JPROC.2009.2021005](https://doi.org/10.1109/JPROC.2009.2021005) (Cited on page 126.)
- Dewey, R. J., Taylor, J. H., Weisberg, J. M., & Stokes, G. H. 1985, *ApJ*, 294, L25, doi: [10.1086/184502](https://doi.org/10.1086/184502) (Cited on page 113.)
- Dhurandhar, S. V., & Vecchio, A. 2001, *Phys. Rev. D*, 63, 122001, doi: [10.1103/PhysRevD.63.122001](https://doi.org/10.1103/PhysRevD.63.122001) (Cited on page 77.)

- Dimoudi, S., Adamek, K., Thiagaraj, P., et al. 2018, *ApJS*, 239, 28, doi: [10.3847/1538-4365/aabe88](https://doi.org/10.3847/1538-4365/aabe88) (Cited on pages 73, 158 and 193.)
- Dodson, R., Harris, C., Pal, S., & Wayth, R. 2010, in ISKAF2010 Science Meeting, assen, the netherlands edn., ed. P. Science, Vol. 1 (Proceedings of Science), 64 (Cited on page 57.)
- Donner, J. Y., Verbiest, J. P. W., Tiburzi, C., et al. 2020, *A&A*, 644, A153, doi: [10.1051/0004-6361/202039517](https://doi.org/10.1051/0004-6361/202039517) (Cited on page 33.)
- Duncan, R. C., & Thompson, C. 1992, *ApJ*, 392, L9, doi: [10.1086/186413](https://doi.org/10.1086/186413) (Cited on page 29.)
- Eatough, R. P., Keane, E. F., & Lyne, A. G. 2009, *MNRAS*, 395, 410, doi: [10.1111/j.1365-2966.2009.14524.x](https://doi.org/10.1111/j.1365-2966.2009.14524.x) (Cited on page 51.)
- Eatough, R. P., Kramer, M., Lyne, A. G., & Keith, M. J. 2013a, *MNRAS*, 431, 292, doi: [10.1093/mnras/stt161](https://doi.org/10.1093/mnras/stt161) (Cited on pages 71, 74, 76, 77 and 82.)
- Eatough, R. P., Molkenhain, N., Kramer, M., et al. 2010, *MNRAS*, 407, 2443, doi: [10.1111/j.1365-2966.2010.17082.x](https://doi.org/10.1111/j.1365-2966.2010.17082.x) (Cited on pages 82, 151, 155, 156 and 157.)
- Eatough, R. P., Falcke, H., Karuppusamy, R., et al. 2013b, *Nature*, 501, 391, doi: [10.1038/nature12499](https://doi.org/10.1038/nature12499) (Cited on pages 29 and 40.)
- Eddington, A. S. 1920, *Nature*, 106, 14, doi: [10.1038/106014a0](https://doi.org/10.1038/106014a0) (Cited on page 19.)
- Einstein, A. 1915, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* (Berlin), 844 (Cited on page 37.)
- Faucher-Giguère, C.-A., & Kaspi, V. 2005, *JRASC*, 99, 137 (Cited on page 26.)
- FaucherGiguère, C., & Kaspi, V. M. 2006, *The Astrophysical Journal*, 643, 332â355, doi: [10.1086/501516](https://doi.org/10.1086/501516) (Cited on page 135.)
- Faulkner, A. J., Stairs, I. H., Kramer, M., et al. 2004, *MNRAS*, 355, 147, doi: [10.1111/j.1365-2966.2004.08310.x](https://doi.org/10.1111/j.1365-2966.2004.08310.x) (Cited on pages 75, 82 and 151.)
- Feng, Y., Hobbs, G., Li, D., et al. 2020, arXiv e-prints, arXiv:2012.06709. <https://arxiv.org/abs/2012.06709> (Cited on page 195.)
- Ferdman, R. D., Stairs, I. H., Kramer, M., et al. 2013, *ApJ*, 767, 85, doi: [10.1088/0004-637X/767/1/85](https://doi.org/10.1088/0004-637X/767/1/85) (Cited on page 28.)
- Ferraro, F. R., Dalessandro, E., Mucciarelli, A., et al. 2009, *Nature*, 462, 483â486, doi: [10.1038/nature08581](https://doi.org/10.1038/nature08581) (Cited on page 109.)
- Fonseca, E., Cromartie, H. T., Pennucci, T. T., et al. 2021, arXiv e-prints, arXiv:2104.00880. <https://arxiv.org/abs/2104.00880> (Cited on page 36.)

- Foster, R. S., & Backer, D. C. 1990, *ApJ*, 361, 300, doi: [10.1086/169195](https://doi.org/10.1086/169195) (Cited on pages 38 and 171.)
- Fowler, R. H. 1926, *MNRAS*, 87, 114, doi: [10.1093/mnras/87.2.114](https://doi.org/10.1093/mnras/87.2.114) (Cited on page 20.)
- Freire, P. C. C., & Tauris, T. M. 2013, *Monthly Notices of the Royal Astronomical Society: Letters*, 438, L86âL90, doi: [10.1093/mnrasl/s1t164](https://doi.org/10.1093/mnrasl/s1t164) (Cited on page 43.)
- Freire, P. C. C., Bassa, C. G., Wex, N., et al. 2011, *MNRAS*, 412, 2763, doi: [10.1111/j.1365-2966.2010.18109.x](https://doi.org/10.1111/j.1365-2966.2010.18109.x) (Cited on pages 41 and 43.)
- Freire, P. C. C., Wex, N., Esposito-Farèse, G., et al. 2012, *MNRAS*, 423, 3328, doi: [10.1111/j.1365-2966.2012.21253.x](https://doi.org/10.1111/j.1365-2966.2012.21253.x) (Cited on page 38.)
- Fruchter, A. S., & Goss, W. M. 1990, *apjl*, 365, L63, doi: [10.1086/185889](https://doi.org/10.1086/185889) (Cited on page 110.)
- Fruchter, A. S., & Goss, W. M. 2000, *The Astrophysical Journal*, 536, 865â874, doi: [10.1086/308948](https://doi.org/10.1086/308948) (Cited on page 110.)
- Fruchter, A. S., Stinebring, D. R., & Taylor, J. H. 1988, *Nature*, 333, 237, doi: [10.1038/333237a0](https://doi.org/10.1038/333237a0) (Cited on pages 41, 94 and 111.)
- Gajjar, V., Joshi, B. C., & Kramer, M. 2012, *Monthly Notices of the Royal Astronomical Society*, 424, 1197, doi: [10.1111/j.1365-2966.2012.21296.x](https://doi.org/10.1111/j.1365-2966.2012.21296.x) (Cited on page 30.)
- Gajjar, V., Siemion, A., Croft, S., et al. 2019, in *Bulletin of the American Astronomical Society*, Vol. 51, 223. <https://arxiv.org/abs/1907.05519> (Cited on page 90.)
- Giacconi, R., Murray, S., Gursky, H., et al. 1974, *ApJS*, 27, 37, doi: [10.1086/190288](https://doi.org/10.1086/190288) (Cited on page 83.)
- Ginzburg, V. L. 1964, *Soviet Physics Doklady*, 9, 329 (Cited on page 25.)
- Gold, T. 1968, *Nature*, 218, 731, doi: [10.1038/218731a0](https://doi.org/10.1038/218731a0) (Cited on page 20.)
- Goldreich, P., & Julian, W. H. 1969, *ApJ*, 157, 869, doi: [10.1086/150119](https://doi.org/10.1086/150119) (Cited on page 22.)
- Gorham, P. W., Ray, P. S., Anderson, S. B., Kulkarni, S. R., & Prince, T. A. 1996, *ApJ*, 458, 257, doi: [10.1086/176808](https://doi.org/10.1086/176808) (Cited on page 94.)
- Gravity Collaboration, Abuter, R., Amorim, A., et al. 2019, *A&A*, 625, L10, doi: [10.1051/0004-6361/201935656](https://doi.org/10.1051/0004-6361/201935656) (Cited on page 128.)
- Green, D. A. 2014, A catalogue of 294 Galactic supernova remnants. <https://arxiv.org/abs/1409.0637> (Cited on page 94.)

- Guo, P., Duan, F., Wang, P., et al. 2017, arXiv e-prints, arXiv:1711.10339. <https://arxiv.org/abs/1711.10339> (Cited on page 158.)
- Gupta, Y., Rickett, B. J., & Lyne, A. G. 1994, MNRAS, 269, 1035, doi: [10.1093/mnras/269.4.1035](https://doi.org/10.1093/mnras/269.4.1035) (Cited on page 35.)
- H. E. S. S. Collaboration, Abdalla, H., Abramowski, A., et al. 2018, aap, 612, A1, doi: [10.1051/0004-6361/201732098](https://doi.org/10.1051/0004-6361/201732098) (Cited on page 94.)
- Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, aap, 117, 137 (Cited on page 24.)
- Hamilton, T. T., Helfand, D. J., & Becker, R. H. 1985, AJ, 90, 606, doi: [10.1086/113767](https://doi.org/10.1086/113767) (Cited on page 83.)
- Han, J. L., & Manchester, R. N. 2001, MNRAS, 320, L35, doi: [10.1046/j.1365-8711.2001.04146.x](https://doi.org/10.1046/j.1365-8711.2001.04146.x) (Cited on page 22.)
- Han, J. L., Manchester, R. N., Lyne, A. G., Qiao, G. J., & van Straten, W. 2006, The Astrophysical Journal, 642, 868â881, doi: [10.1086/501444](https://doi.org/10.1086/501444) (Cited on page 23.)
- Han, J. L., Manchester, R. N., & Qiao, G. J. 1999, Monthly Notices of the Royal Astronomical Society, 306, 371, doi: [10.1046/j.1365-8711.1999.02544.x](https://doi.org/10.1046/j.1365-8711.1999.02544.x) (Cited on page 23.)
- Han, J. L., Manchester, R. N., van Straten, W., & Demorest, P. 2018, The Astrophysical Journal Supplement Series, 234, 11, doi: [10.3847/1538-4365/aa9c45](https://doi.org/10.3847/1538-4365/aa9c45) (Cited on page 23.)
- Hankins, T. H., & Rickett, B. J. 1975, Methods in Computational Physics, 14, 55, doi: [10.1016/B978-0-12-460814-6.50007-3](https://doi.org/10.1016/B978-0-12-460814-6.50007-3) (Cited on page 54.)
- Harding, A. K. 2017, Proceedings of the International Astronomical Union, 13, 52â57, doi: [10.1017/s1743921318000509](https://doi.org/10.1017/s1743921318000509) (Cited on page 22.)
- Haskell, B., Zdunik, J. L., Fortin, M., et al. 2018, A&A, 620, A69, doi: [10.1051/0004-6361/201833521](https://doi.org/10.1051/0004-6361/201833521) (Cited on page 36.)
- Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, , 47, 1 (Cited on pages 133 and 145.)
- Hearst, M. A. 1998, IEEE Intelligent Systems, 13, 18â28, doi: [10.1109/5254.708428](https://doi.org/10.1109/5254.708428) (Cited on page 151.)
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, ApJ, 591, 288, doi: [10.1086/375341](https://doi.org/10.1086/375341) (Cited on page 31.)
- Heinke, C. O. 2010, in American Institute of Physics Conference Series, Vol. 1314, International Conference on Binaries: in celebration of Ron Webbink's 65th Birthday, ed. V. Kalogera & M. van der Sluys, 135–142, doi: [10.1063/1.3536355](https://doi.org/10.1063/1.3536355) (Cited on page 31.)

- Heinke, C. O., Wijnands, R., Cohn, H. N., et al. 2006, *The Astrophysical Journal*, 651, 1098â1111, doi: [10.1086/507884](https://doi.org/10.1086/507884) (Cited on page 109.)
- H.E.S.S. Collaboration, Abramowski, A., Acero, F., et al. 2011, *A&A*, 531, L18, doi: [10.1051/0004-6361/201117171](https://doi.org/10.1051/0004-6361/201117171) (Cited on page 109.)
- Hessels, J. W. T., Ransom, S. M., Stairs, I. H., et al. 2006, *Science*, 311, 1901, doi: [10.1126/science.1123430](https://doi.org/10.1126/science.1123430) (Cited on pages 32, 36, 73, 84, 95, 109, 110 and 115.)
- Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, *Nature*, 217, 709, doi: [10.1038/217709a0](https://doi.org/10.1038/217709a0) (Cited on pages 20, 32 and 80.)
- Hewish, A., & Okoye, S. E. 1965, *Nature*, 207, 59, doi: [10.1038/207059a0](https://doi.org/10.1038/207059a0) (Cited on page 20.)
- Heywood, I., Camilo, F., Cotton, W. D., et al. 2019, *Nature*, 573, 235, doi: [10.1038/s41586-019-1532-5](https://doi.org/10.1038/s41586-019-1532-5) (Cited on page 129.)
- Hirschman, J. A., & Arons, J. 2001, *The Astrophysical Journal*, 546, 382, doi: [10.1086/318224](https://doi.org/10.1086/318224) (Cited on page 188.)
- Hobbs, G., Faulkner, A., Stairs, I. H., et al. 2004, *MNRAS*, 352, 1439, doi: [10.1111/j.1365-2966.2004.08042.x](https://doi.org/10.1111/j.1365-2966.2004.08042.x) (Cited on page 82.)
- Hobbs, G., et al. 2010, *Class. Quant. Grav.*, 27, 084013, doi: [10.1088/0264-9381/27/8/084013](https://doi.org/10.1088/0264-9381/27/8/084013) (Cited on page 39.)
- Hotan, A., Bailes, M., & Ord, S. M. 2005, *MNRAS*, 362, 1267, doi: [10.1111/j.1365-2966.2005.09389.x](https://doi.org/10.1111/j.1365-2966.2005.09389.x) (Cited on pages 54 and 171.)
- Hotan, A. W. 2005, *Astronomical Society of the Pacific Conference Series*, Vol. 328, High-Precision Baseband Timing at Parkes, ed. F. A. Rasio & I. H. Stairs, 377 (Cited on page 175.)
- Hotan, A. W., Bailes, M., & Ord, S. M. 2004a, *MNRAS*, 355, 941, doi: [10.1111/j.1365-2966.2004.08376.x](https://doi.org/10.1111/j.1365-2966.2004.08376.x) (Cited on pages 12, 172, 173, 174, 175, 176, 178, 181, 183, 185, 187 and 189.)
- Hotan, A. W., van Straten, W., & Manchester, R. N. 2004b, *Publications of the Astronomical Society of Australia*, 21, 302, doi: [10.1071/as04022](https://doi.org/10.1071/as04022) (Cited on page 175.)
- Hu, H., Kramer, M., Wex, N., Champion, D. J., & Kehl, M. S. 2020, *MNRAS*, 497, 3118, doi: [10.1093/mnras/staa2107](https://doi.org/10.1093/mnras/staa2107) (Cited on page 37.)
- Hughes, G. 1968, *IEEE Transactions on Information Theory*, 14, 55, doi: [10.1109/TIT.1968.1054102](https://doi.org/10.1109/TIT.1968.1054102) (Cited on page 157.)
- Hulse, R., & Taylor, J. H. 1974, *ApJ*, 191, L59, doi: [10.1086/181548](https://doi.org/10.1086/181548) (Cited on page 81.)

- Hulse, R. A., & Taylor, J. . 1975a, ApJ, 195, L51, doi: [10.1086/181708](https://doi.org/10.1086/181708) (Cited on page 81.)
- Hulse, R. A., & Taylor, J. H. 1975b, ApJ, 201, L55, doi: [10.1086/181941](https://doi.org/10.1086/181941) (Cited on pages 28, 38 and 81.)
- Hund, F. 1936, *Materie unter sehr hohen Drucken und Temperaturen* (Berlin, Heidelberg: Springer Berlin Heidelberg), 189–228, doi: [10.1007/BFb0111968](https://doi.org/10.1007/BFb0111968) (Cited on page 20.)
- Israel, G. L., & Stella, L. 1996, ApJ, 468, 369, doi: [10.1086/177697](https://doi.org/10.1086/177697) (Cited on page 60.)
- Jackson, J. D. 1962, *Classical Electrodynamics* (Cited on page 25.)
- Janssen, G. H., & van Leeuwen, J. 2004, aap, 425, 255, doi: [10.1051/0004-6361:20041062](https://doi.org/10.1051/0004-6361:20041062) (Cited on page 30.)
- Jaodand, A., Hessels, J. W. T., & Archibald, A. 2018, in *Pulsar Astrophysics the Next Fifty Years*, ed. P. Weltevrede, B. B. P. Perera, L. L. Preston, & S. Sanidas, Vol. 337, 47–51, doi: [10.1017/S1743921317010407](https://doi.org/10.1017/S1743921317010407) (Cited on page 28.)
- Jessner, A. 1996, in *Large Antennas in Radio Astronomy*, ed. C. G. M. van't Klooster & A. van Ardenne, 185 (Cited on page 175.)
- Jiang, L., Li, X.-D., Dey, J., & Dey, M. 2015, ApJ, 807, 41, doi: [10.1088/0004-637X/807/1/41](https://doi.org/10.1088/0004-637X/807/1/41) (Cited on page 43.)
- Johnston, H. M., & Kulkarni, S. R. 1991, ApJ, 368, 504, doi: [10.1086/169715](https://doi.org/10.1086/169715) (Cited on page 70.)
- Johnston, S., & Karastergiou, A. 2019, MNRAS, 485, 640, doi: [10.1093/mnras/stz400](https://doi.org/10.1093/mnras/stz400) (Cited on page 40.)
- Johnston, S., & Kerr, M. 2017, *Monthly Notices of the Royal Astronomical Society*, 474, 4629, doi: [10.1093/mnras/stx3095](https://doi.org/10.1093/mnras/stx3095) (Cited on pages 23 and 40.)
- Johnston, S., Kramer, M., Lorimer, D. R., et al. 2006, MNRAS, 373, L6, doi: [10.1111/j.1745-3933.2006.00232.x](https://doi.org/10.1111/j.1745-3933.2006.00232.x) (Cited on page 40.)
- Johnston, S., Manchester, R. N., Lyne, A. G., et al. 1992, ApJ, 387, L37, doi: [10.1086/186300](https://doi.org/10.1086/186300) (Cited on page 68.)
- Johnston, S., Smith, D. A., Karastergiou, A., & Kramer, M. 2020, MNRAS, 497, 1957, doi: [10.1093/mnras/staa2110](https://doi.org/10.1093/mnras/staa2110) (Cited on page 95.)
- Jones, S., Röpke, F. K., Pakmor, R., et al. 2016, A&A, 593, A72, doi: [10.1051/0004-6361/201628321](https://doi.org/10.1051/0004-6361/201628321) (Cited on page 41.)
- Jouteux, S., Ramachandran, R., Stappers, B. W., Jonker, P. G., & van der Klis, M. 2002, A&A, 384, 532, doi: [10.1051/0004-6361:20020052](https://doi.org/10.1051/0004-6361:20020052) (Cited on page 75.)

- Kaplan, D., Escoffier, R., Lacasse, R., et al. 2005, *Publications of the Astronomical Society of the Pacific*, 117, 643. <http://www.jstor.org/stable/10.1086/430368> (Cited on pages 110 and 113.)
- Karastergiou, A., Roberts, S. J., Johnston, S., et al. 2011, *MNRAS*, 415, 251, doi: [10.1111/j.1365-2966.2011.18697.x](https://doi.org/10.1111/j.1365-2966.2011.18697.x) (Cited on page 188.)
- Kaspi, V. M., & Beloborodov, A. M. 2017, *ARA&A*, 55, 261, doi: [10.1146/annurev-astro-081915-023329](https://doi.org/10.1146/annurev-astro-081915-023329) (Cited on page 29.)
- Keane, E., Bhattacharyya, B., Kramer, M., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 40. <https://arxiv.org/abs/1501.00056> (Cited on page 196.)
- Keane, E. F., & Kramer, M. 2008, *MNRAS*, 391, 2009, doi: [10.1111/j.1365-2966.2008.14045.x](https://doi.org/10.1111/j.1365-2966.2008.14045.x) (Cited on pages 31 and 95.)
- Keane, E. F., Kramer, M., Lyne, A. G., Stappers, B. W., & McLaughlin, M. A. 2011, *MNRAS*, 415, 3065, doi: [10.1111/j.1365-2966.2011.18917.x](https://doi.org/10.1111/j.1365-2966.2011.18917.x) (Cited on page 30.)
- Keane, E. F., Barr, E. D., Jameson, A., et al. 2018, *MNRAS*, 473, 116, doi: [10.1093/mnras/stx2126](https://doi.org/10.1093/mnras/stx2126) (Cited on page 105.)
- Keith, M., Johnston, S., Ray, P. S., et al. 2011, *mnras*, 414, 1292, doi: [10.1111/j.1365-2966.2011.18464.x](https://doi.org/10.1111/j.1365-2966.2011.18464.x) (Cited on pages 71 and 94.)
- Keith, M. J., Eatough, R. P., Lyne, A. G., et al. 2009, *MNRAS*, 395, 837, doi: [10.1111/j.1365-2966.2009.14543.x](https://doi.org/10.1111/j.1365-2966.2009.14543.x) (Cited on pages 82 and 151.)
- Keith, M. J., Jameson, A., van Straten, W., et al. 2010, *MNRAS*, 409, 619, doi: [10.1111/j.1365-2966.2010.17325.x](https://doi.org/10.1111/j.1365-2966.2010.17325.x) (Cited on pages 80, 82, 133, 157, 158, 160, 162 and 168.)
- Kerr, M., Reardon, D. J., Hobbs, G., et al. 2020, *Publications of the Astronomical Society of Australia*, 37, e020, doi: [10.1017/pasa.2020.11](https://doi.org/10.1017/pasa.2020.11) (Cited on page 96.)
- Kloumann, I. M., & Rankin, J. M. 2010, *Monthly Notices of the Royal Astronomical Society*, 408, 40, doi: [10.1111/j.1365-2966.2010.17114.x](https://doi.org/10.1111/j.1365-2966.2010.17114.x) (Cited on page 30.)
- Knispel, B. 2011, PhD thesis, Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany ; Institut für Gravitationsphysik, Leibniz Universität Hannover, D-30167 Hannover, Germany (Cited on page 77.)
- Knispel, B., Eatough, R. P., Kim, H., et al. 2013, *ApJ*, 774, 93, doi: [10.1088/0004-637X/774/2/93](https://doi.org/10.1088/0004-637X/774/2/93) (Cited on pages 77 and 82.)
- Knispel, B., Lyne, A. G., Stappers, B. W., et al. 2015, *ApJ*, 806, 140, doi: [10.1088/0004-637X/806/1/140](https://doi.org/10.1088/0004-637X/806/1/140) (Cited on page 41.)

- Kocz, J., Bailes, M., Barnes, D., Burke-Spolaor, S., & Levin, L. 2012, MNRAS, 420, 271, doi: [10.1111/j.1365-2966.2011.20029.x](https://doi.org/10.1111/j.1365-2966.2011.20029.x) (Cited on page 51.)
- Kocz, J., Briggs, F. H., & Reynolds, J. 2010, AJ, 140, 2086, doi: [10.1088/0004-6256/140/6/2086](https://doi.org/10.1088/0004-6256/140/6/2086) (Cited on page 51.)
- Konar, S., & Deka, U. 2019, Journal of Astrophysics and Astronomy, 40, doi: [10.1007/s12036-019-9608-z](https://doi.org/10.1007/s12036-019-9608-z) (Cited on page 30.)
- Kouwenhoven, M. L. A., & Voûte, J. L. L. 2001, A&A, 378, 700, doi: [10.1051/0004-6361:20011226](https://doi.org/10.1051/0004-6361:20011226) (Cited on page 133.)
- Kramer, M. 1998, ApJ, 509, 856, doi: [10.1086/306535](https://doi.org/10.1086/306535) (Cited on page 172.)
- Kramer, M., Lyne, A. G., O'Brien, J. T., Jordan, C. A., & Lorimer, D. R. 2006a, Science, 312, 549, doi: [10.1126/science.1124060](https://doi.org/10.1126/science.1124060) (Cited on page 30.)
- Kramer, M., Xilouris, K. M., Camilo, F., et al. 1999, The Astrophysical Journal, 520, 324, doi: [10.1086/307449](https://doi.org/10.1086/307449) (Cited on pages 172, 173, 174, 175, 176, 178, 179, 187 and 188.)
- Kramer, M., Bell, J. F., Manchester, R. N., et al. 2003, MNRAS, 342, 1299, doi: [10.1046/j.1365-8711.2003.06637.x](https://doi.org/10.1046/j.1365-8711.2003.06637.x) (Cited on page 82.)
- Kramer, M., Stairs, I. H., Manchester, R. N., et al. 2006b, Science, 314, 97, doi: [10.1126/science.1132305](https://doi.org/10.1126/science.1132305) (Cited on pages 29, 38 and 171.)
- Krishnakumar, M. A., Joshi, B. C., & Manoharan, P. K. 2017, ApJ, 846, 104, doi: [10.3847/1538-4357/aa7af2](https://doi.org/10.3847/1538-4357/aa7af2) (Cited on page 34.)
- Krishnan, V. V., Bailes, M., van Straten, W., et al. 2019, The Astrophysical Journal, 873, L15, doi: [10.3847/2041-8213/ab0a03](https://doi.org/10.3847/2041-8213/ab0a03) (Cited on page 172.)
- Landau, L. 1938, Nature, 141, 333, doi: [10.1038/141333b0](https://doi.org/10.1038/141333b0) (Cited on page 20.)
- Lanzoni, B., Ferraro, F. R., Dalessandro, E., et al. 2010, The Astrophysical Journal, 717, 653, doi: [10.1088/0004-637x/717/2/653](https://doi.org/10.1088/0004-637x/717/2/653) (Cited on page 109.)
- Large, M. I., Vaughan, A. E., & Mills, B. Y. 1968a, Nature, 220, 340, doi: [10.1038/220340a0](https://doi.org/10.1038/220340a0) (Cited on page 20.)
- Large, M. I., Vaughan, A. E., & Wielebinski, R. 1968b, Nature, 220, 753, doi: [10.1038/220753a0](https://doi.org/10.1038/220753a0) (Cited on page 81.)
- Lattimer, J. M., & Prakash, M. 2001, ApJ, 550, 426, doi: [10.1086/319702](https://doi.org/10.1086/319702) (Cited on pages 36 and 171.)
- Lazarus, P., Karuppusamy, R., Graikou, E., et al. 2016, Monthly Notices of the Royal Astronomical Society, 458, 868, doi: [10.1093/mnras/stw189](https://doi.org/10.1093/mnras/stw189) (Cited on pages 54 and 175.)

- Lazarus, P., Brazier, A., Hessels, J. W. T., et al. 2015, *ApJ*, 812, 81, doi: [10.1088/0004-637X/812/1/81](https://doi.org/10.1088/0004-637X/812/1/81) (Cited on page 194.)
- LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539) (Cited on page 151.)
- Lee, K. J., Guillemot, L., Yue, Y. L., Kramer, M., & Champion, D. J. 2012, *MNRAS*, 424, 2832, doi: [10.1111/j.1365-2966.2012.21413.x](https://doi.org/10.1111/j.1365-2966.2012.21413.x) (Cited on pages 27 and 28.)
- Lee, K. J., Stovall, K., Jenet, F. A., et al. 2013, *MNRAS*, 433, 688, doi: [10.1093/mnras/stt758](https://doi.org/10.1093/mnras/stt758) (Cited on pages 151 and 195.)
- Lense, J., & Thirring, H. 1918, *Physikalische Zeitschrift*, 19, 156 (Cited on page 38.)
- Levin, L. 2012, PhD thesis, Swinburne University of Technology (Cited on pages 55 and 56.)
- Levin, L., Bailes, M., Bates, S., et al. 2010, *The Astrophysical Journal*, 721, L33, doi: [10.1088/2041-8205/721/1/L33](https://doi.org/10.1088/2041-8205/721/1/L33) (Cited on page 29.)
- Levin, L., Armour, W., Baffa, C., et al. 2018, in *Pulsar Astrophysics the Next Fifty Years*, ed. P. Weltevrede, B. B. P. Perera, L. L. Preston, & S. Sanidas, Vol. 337, 171–174, doi: [10.1017/S1743921317009528](https://doi.org/10.1017/S1743921317009528) (Cited on pages 159, 196 and 197.)
- Li, D., Wang, P., Qian, L., et al. 2018, *IEEE Microwave Magazine*, 19, 112, doi: [10.1109/MMM.2018.2802178](https://doi.org/10.1109/MMM.2018.2802178) (Cited on page 81.)
- Li, L., Tong, H., Yan, W. M., et al. 2014, *ApJ*, 788, 16, doi: [10.1088/0004-637X/788/1/16](https://doi.org/10.1088/0004-637X/788/1/16) (Cited on page 30.)
- Lin, L., Göğüş, E., Kaneko, Y., & Kouveliotou, C. 2013, *ApJ*, 778, 105, doi: [10.1088/0004-637X/778/2/105](https://doi.org/10.1088/0004-637X/778/2/105) (Cited on page 29.)
- Liu, K., Eatough, R. P., Wex, N., & Kramer, M. 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 3115, doi: [10.1093/mnras/stu1913](https://doi.org/10.1093/mnras/stu1913) (Cited on page 38.)
- Liu, K., Keane, E. F., Lee, K. J., et al. 2012, *MNRAS*, 420, 361, doi: [10.1111/j.1365-2966.2011.20041.x](https://doi.org/10.1111/j.1365-2966.2011.20041.x) (Cited on pages 172 and 174.)
- Liu, K., Karuppusamy, R., Lee, K. J., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 1158, doi: [10.1093/mnras/stv397](https://doi.org/10.1093/mnras/stv397) (Cited on pages 172, 174, 186 and 188.)
- Lorimer, D., Pol, N., Rajwade, K., et al. 2019, *BAAS*, 51, 261. <https://arxiv.org/abs/1903.06526> (Cited on pages 40 and 125.)
- Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, *Science*, 318, 777, doi: [10.1126/science.1147532](https://doi.org/10.1126/science.1147532) (Cited on page 43.)

- Lorimer, D. R., & Kramer, M. 2012, Handbook of Pulsar Astronomy (Cited on pages 23, 24, 25, 33, 35, 58, 59, 68, 133 and 169.)
- Lorimer, D. R., Kramer, M., Müller, P., et al. 2000, A&A, 358, 169. <https://arxiv.org/abs/astro-ph/9910569> (Cited on page 81.)
- Lorimer, D. R., Lyne, A. G., McLaughlin, M. A., et al. 2012, ApJ, 758, 141, doi: [10.1088/0004-637X/758/2/141](https://doi.org/10.1088/0004-637X/758/2/141) (Cited on page 30.)
- Lorimer, D. R., Faulkner, A. J., Lyne, A. G., et al. 2006, MNRAS, 372, 777, doi: [10.1111/j.1365-2966.2006.10887.x](https://doi.org/10.1111/j.1365-2966.2006.10887.x) (Cited on pages 40, 82 and 135.)
- Lorimer, D. R., Esposito, P., Manchester, R. N., et al. 2015, MNRAS, 450, 2185, doi: [10.1093/mnras/stv804](https://doi.org/10.1093/mnras/stv804) (Cited on pages 40, 135 and 136.)
- Lovelace, R. V. E., S. J. M., & Craft, jun., H. D. 1968 (Cited on page 20.)
- Lower, M. E., Shannon, R. M., Johnston, S., & Bailes, M. 2020, The Astrophysical Journal, 896, L37, doi: [10.3847/2041-8213/ab9898](https://doi.org/10.3847/2041-8213/ab9898) (Cited on page 29.)
- Lundgren, S. C., Foster, R. S., & Camilo, F. 1996, Astronomical Society of the Pacific Conference Series, Vol. 105, Hubble Space Telescope Observations of Millisecond Pulsar Companions: Constraints on Evolution, ed. S. Johnston, M. A. Walker, & M. Bailes, 497 (Cited on page 172.)
- Lyne, A., Hobbs, G., Kramer, M., Stairs, I., & Stappers, B. 2010, Science, 329, 408, doi: [10.1126/science.1186683](https://doi.org/10.1126/science.1186683) (Cited on pages 172 and 188.)
- Lyne, A. G., Biggs, J. D., Brinklow, A., Ashworth, M., & McKenna, J. 1988, Nature, 332, 45, doi: [10.1038/332045a0](https://doi.org/10.1038/332045a0) (Cited on page 84.)
- Lyne, A. G., Brinklow, A., Middleditch, J., Kulkarni, S. R., & Backer, D. C. 1987, Nature, 328, 399, doi: [10.1038/328399a0](https://doi.org/10.1038/328399a0) (Cited on page 83.)
- Lyne, A. G., & Manchester, R. N. 1988, MNRAS, 234, 477, doi: [10.1093/mnras/234.3.477](https://doi.org/10.1093/mnras/234.3.477) (Cited on page 22.)
- Lyne, A. G., Manchester, R. N., & Taylor, J. H. 1985, Monthly Notices of the Royal Astronomical Society, 213, 613, doi: [10.1093/mnras/213.3.613](https://doi.org/10.1093/mnras/213.3.613) (Cited on page 40.)
- Lyne, A. G., Mankelov, S. H., Bell, J. F., & Manchester, R. N. 2000, Mon. Not. Roy. Astron. Soc., 316, 491, doi: [10.1046/j.1365-8711.2000.03517.x](https://doi.org/10.1046/j.1365-8711.2000.03517.x) (Cited on page 110.)
- Lyne, A. G., Manchester, R. N., D'Amico, N., et al. 1990, Nature, 347, 650, doi: [10.1038/347650a0](https://doi.org/10.1038/347650a0) (Cited on pages 77, 95 and 110.)
- Lyne, A. G., Burgay, M., Kramer, M., et al. 2004, Science, 303, 1153, doi: [10.1126/science.1094645](https://doi.org/10.1126/science.1094645) (Cited on page 38.)

- Lyne, A. G., Stappers, B. W., Freire, P. C. C., et al. 2017, *ApJ*, 834, 72, doi: [10.3847/1538-4357/834/1/72](https://doi.org/10.3847/1538-4357/834/1/72) (Cited on page 30.)
- Lyon, R. J. 2016, PhD thesis. <https://search.proquest.com/dissertations-theses/why-are-pulsars-hard-find/docview/1917323708/se-2?accountid=105561> (Cited on page 151.)
- Lyon, R. J. ., Brooke, J. M., Knowles, J. D., & Stappers, B. W. 2014, arXiv e-prints, arXiv:1405.2278. <https://arxiv.org/abs/1405.2278> (Cited on page 158.)
- Lyon, R. J., Brooke, J. M., Knowles, J. D., & Stappers, B. W. 2013, arXiv e-prints, arXiv:1307.8012. <https://arxiv.org/abs/1307.8012> (Cited on page 158.)
- Lyon, R. J., Stappers, B. W., Cooper, S., Brooke, J. M., & Knowles, J. D. 2016, *Monthly Notices of the Royal Astronomical Society*, 459, 1104â1123, doi: [10.1093/mnras/stw656](https://doi.org/10.1093/mnras/stw656) (Cited on pages 65, 86, 97, 150, 155, 156, 158 and 167.)
- Lyon, R. J., Stappers, B. W., Levin, L., Mickaliger, M. B., & Scaife, A. 2019, *Astronomy and Computing*, 28, 100291, doi: [10.1016/j.ascom.2019.100291](https://doi.org/10.1016/j.ascom.2019.100291) (Cited on pages 159 and 167.)
- Lyubarskii, Y., & Petrova, S. 1998, *Astrophysics and Space Science*, 262, 379â389, doi: [10.1023/a:1001872805645](https://doi.org/10.1023/a:1001872805645) (Cited on page 23.)
- Macqueen, J. 1967, in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 281–297 (Cited on page 151.)
- Magro, A., Karastergiou, A., Salvini, S., et al. 2011, *MNRAS*, 417, 2642, doi: [10.1111/j.1365-2966.2011.19426.x](https://doi.org/10.1111/j.1365-2966.2011.19426.x) (Cited on pages 56 and 57.)
- Malov, I. F. 2021, *MNRAS*, 502, 809, doi: [10.1093/mnras/stab007](https://doi.org/10.1093/mnras/stab007) (Cited on page 31.)
- Malov, O. I., & Malofeev, V. M. 2010, *Astronomy Reports*, 54, 210, doi: [10.1134/S1063772910030030](https://doi.org/10.1134/S1063772910030030) (Cited on page 22.)
- Manchester, R., Lyne, A., Camilo, F., et al. 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 17â35, doi: [10.1046/j.1365-8711.2001.04751.x](https://doi.org/10.1046/j.1365-8711.2001.04751.x) (Cited on pages 71, 82 and 157.)
- Manchester, R. N. 2008, in *American Institute of Physics Conference Series*, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, ed. C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi, 584–592, doi: [10.1063/1.2900303](https://doi.org/10.1063/1.2900303) (Cited on page 39.)
- Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, *The Astronomical Journal*, 129, 1993â2006, doi: [10.1086/428488](https://doi.org/10.1086/428488) (Cited on pages 27, 46, 82, 133, 146 and 156.)
- Manchester, R. N., Lyne, A. G., Robinson, C., et al. 1991, *Nature*, 352, 219, doi: [10.1038/352219a0](https://doi.org/10.1038/352219a0) (Cited on pages 84 and 95.)

- Manchester, R. N., Lyne, A. G., Taylor, J. H., et al. 1978, MNRAS, 185, 409, doi: [10.1093/mnras/185.2.409](https://doi.org/10.1093/mnras/185.2.409) (Cited on pages 80, 81 and 150.)
- Manchester, R. N., Lyne, A. G., D'Amico, N., et al. 1996, Monthly Notices of the Royal Astronomical Society, 279, 1235, doi: [10.1093/mnras/279.4.1235](https://doi.org/10.1093/mnras/279.4.1235) (Cited on page 80.)
- Marais, N. 2015, in 15th International Conference on Accelerator and Large Experimental Physics Control Systems, MOPGF067, doi: [10.18429/JACoW-ICALEPCS2015-MOPGF067](https://doi.org/10.18429/JACoW-ICALEPCS2015-MOPGF067) (Cited on page 88.)
- Martinez, J. G., Stovall, K., Freire, P. C. C., et al. 2015, ApJ, 812, 143, doi: [10.1088/0004-637X/812/2/143](https://doi.org/10.1088/0004-637X/812/2/143) (Cited on page 41.)
- Massey. 1951, Journal of the American Statistical Association, 46, 68, doi: [10.1080/01621459.1951.10500769](https://doi.org/10.1080/01621459.1951.10500769) (Cited on page 179.)
- McLaughlin, M. A., Lyne, A. G., Lorimer, D. R., et al. 2006, Nature, 439, 817, doi: [10.1038/nature04440](https://doi.org/10.1038/nature04440) (Cited on page 29.)
- Middleditch, J., & Kristian, J. 1984, ApJ, 279, 157, doi: [10.1086/161876](https://doi.org/10.1086/161876) (Cited on page 71.)
- Mitra, D., & Deshpande, A. A. 1999, A&A, 346, 906. <https://arxiv.org/abs/astro-ph/9904336> (Cited on page 22.)
- Morello, V., Barr, E. D., Bailes, M., et al. 2014, MNRAS, 443, 1651, doi: [10.1093/mnras/stu1188](https://doi.org/10.1093/mnras/stu1188) (Cited on pages 155, 157 and 195.)
- Morello, V., Barr, E. D., Stappers, B. W., Keane, E. F., & Lyne, A. G. 2020, Monthly Notices of the Royal Astronomical Society, 497, 4654–4671, doi: [10.1093/mnras/staa2291](https://doi.org/10.1093/mnras/staa2291) (Cited on pages 78, 113 and 133.)
- Morello, V., Barr, E. D., Cooper, S., et al. 2019, MNRAS, 483, 3673, doi: [10.1093/mnras/sty3328](https://doi.org/10.1093/mnras/sty3328) (Cited on pages 55, 58, 59, 60, 79, 97 and 105.)
- Morris, D. J., Hobbs, G., Lyne, A. G., et al. 2002, MNRAS, 335, 275, doi: [10.1046/j.1365-8711.2002.05551.x](https://doi.org/10.1046/j.1365-8711.2002.05551.x) (Cited on page 82.)
- Mottez, F., Bonazzola, S., & Heyvaerts, J. 2013a, A&A, 555, A125, doi: [10.1051/0004-6361/201321182](https://doi.org/10.1051/0004-6361/201321182) (Cited on pages 30 and 172.)
- Mottez, F., Bonazzola, S., & Heyvaerts, J. 2013b, A&A, 555, A126, doi: [10.1051/0004-6361/201321184](https://doi.org/10.1051/0004-6361/201321184) (Cited on page 30.)
- Müller, J., Murphy, T. W., Schreiber, U., et al. 2019, Journal of Geodesy, 93, 2195, doi: [10.1007/s00190-019-01296-0](https://doi.org/10.1007/s00190-019-01296-0) (Cited on page 38.)
- Murphy, T., Chatterjee, S., Kaplan, D. L., et al. 2013, PASA, 30, e006, doi: [10.1017/pasa.2012.006](https://doi.org/10.1017/pasa.2012.006) (Cited on page 85.)

- Ng, C., Champion, D. J., Bailes, M., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 2922, doi: [10.1093/mnras/stv753](https://doi.org/10.1093/mnras/stv753) (Cited on pages 71, 75, 80, 82, 115, 160 and 164.)
- Ng, C. W. Y. 2014, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn. <http://hdl.handle.net/20.500.11811/6215> (Cited on page 73.)
- Nice, D. J., Sayer, R. W., & Taylor, J. H. 1996, *ApJ*, 466, L87, doi: [10.1086/310178](https://doi.org/10.1086/310178) (Cited on page 31.)
- Nice, D. J., Taylor, J. H., & Fruchter, A. S. 1993, *ApJ*, 402, L49, doi: [10.1086/186697](https://doi.org/10.1086/186697) (Cited on page 81.)
- Nieder, L., Clark, C. J., Kandel, D., et al. 2020, *ApJ*, 902, L46, doi: [10.3847/2041-8213/abbc02](https://doi.org/10.3847/2041-8213/abbc02) (Cited on pages 68 and 77.)
- Nyquist, H. 1928, *Transactions of the American Institute of Electrical Engineers*, 47, 617, doi: [10.1109/T-AIEE.1928.5055024](https://doi.org/10.1109/T-AIEE.1928.5055024) (Cited on page 47.)
- Olausen, S. A., & Kaspi, V. M. 2014, *ApJS*, 212, 6, doi: [10.1088/0067-0049/212/1/6](https://doi.org/10.1088/0067-0049/212/1/6) (Cited on page 29.)
- Oppenheimer, J. R., & Volkoff, G. M. 1939, *Phys. Rev.*, 55, 374, doi: [10.1103/PhysRev.55.374](https://doi.org/10.1103/PhysRev.55.374) (Cited on pages 20 and 21.)
- Osłowski, S., van Straten, W., Hobbs, G. B., Bailes, M., & Demorest, P. 2011, *Monthly Notices of the Royal Astronomical Society*, 418, 1258, doi: [10.1111/j.1365-2966.2011.19578.x](https://doi.org/10.1111/j.1365-2966.2011.19578.x) (Cited on pages 170 and 172.)
- Oswald, L. S., Karastergiou, A., Posselt, B., et al. 2021, arXiv e-prints, arXiv:2104.01081. <https://arxiv.org/abs/2104.01081> (Cited on page 34.)
- Pacini, F. 1967, *Nature*, 216, 567, doi: [10.1038/216567a0](https://doi.org/10.1038/216567a0) (Cited on page 20.)
- Papitto, A., Ferrigno, C., Bozzo, E., et al. 2013, *Nature*, 501, 517, doi: [10.1038/nature12470](https://doi.org/10.1038/nature12470) (Cited on page 32.)
- Parfrey, K., & Tchekhovskoy, A. 2017, *ApJ*, 851, L34, doi: [10.3847/2041-8213/aa9c85](https://doi.org/10.3847/2041-8213/aa9c85) (Cited on page 41.)
- Parkinson, P. M. S., Xu, H., Yu, P. L. H., et al. 2016, *The Astrophysical Journal*, 820, 8, doi: [10.3847/0004-637x/820/1/8](https://doi.org/10.3847/0004-637x/820/1/8) (Cited on page 94.)
- Paul, A., Susobhanan, A., Gopakumar, A., et al. 2019, in 2019 URSI Asia-Pacific Radio Science Conference (AP-RASC), 1–1, doi: [10.23919/URSIAP-RASC.2019.8738505](https://doi.org/10.23919/URSIAP-RASC.2019.8738505) (Cited on page 39.)
- Pennucci, T. 2009, in *American Astronomical Society Meeting Abstracts*, Vol. 213, *American Astronomical Society Meeting Abstracts #213*, 436.13 (Cited on page 76.)

- Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, MNRAS, 490, 4666, doi: [10.1093/mnras/stz2857](https://doi.org/10.1093/mnras/stz2857) (Cited on page 39.)
- Perera, B. B. P., Barr, E. D., Mickaliger, M. B., et al. 2019, Monthly Notices of the Royal Astronomical Society, 487, 1025â1039, doi: [10.1093/mnras/stz1180](https://doi.org/10.1093/mnras/stz1180) (Cited on page 40.)
- Phinney, E. S., Evans, C. R., Blandford, R. D., & Kulkarni, S. R. 1988, Nature, 333, 832, doi: [10.1038/333832a0](https://doi.org/10.1038/333832a0) (Cited on page 94.)
- Pleunis, Z., Bassa, C., Hessels, J., et al. 2017, Astrophys. J. Lett., 846, L19, doi: [10.3847/2041-8213/aa83ff](https://doi.org/10.3847/2041-8213/aa83ff) (Cited on page 94.)
- Pooley, D., Lewin, W. H. G., Anderson, S. F., et al. 2003, apjl, 591, L131, doi: [10.1086/377074](https://doi.org/10.1086/377074) (Cited on page 95.)
- Possenti, A., D'Amico, N., Manchester, R. N., et al. 2003, ApJ, 599, 475, doi: [10.1086/379190](https://doi.org/10.1086/379190) (Cited on page 84.)
- Pound, R. V., & Rebka, G. A. 1960, Phys. Rev. Lett., 4, 337, doi: [10.1103/PhysRevLett.4.337](https://doi.org/10.1103/PhysRevLett.4.337) (Cited on page 37.)
- Prager, B. J., Ransom, S. M., Freire, P. C. C., et al. 2017, The Astrophysical Journal, 845, 148, doi: [10.3847/1538-4357/aa7ed7](https://doi.org/10.3847/1538-4357/aa7ed7) (Cited on pages 95, 109 and 194.)
- Prestage, R. M., Bloss, M., Brandt, J., et al. 2015, in 2015 USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), 294–294, doi: [10.1109/USNC-URSI.2015.7303578](https://doi.org/10.1109/USNC-URSI.2015.7303578) (Cited on page 113.)
- Prince, T. A., Anderson, S. B., Kulkarni, S. R., & Wolszczan, A. 1991, apjl, 374, L41, doi: [10.1086/186067](https://doi.org/10.1086/186067) (Cited on page 95.)
- Purver, M. B. 2010, PhD thesis, University of Manchester (Cited on page 173.)
- Qiu, H., Bannister, K. W., Shannon, R. M., et al. 2019, MNRAS, 486, 166, doi: [10.1093/mnras/stz748](https://doi.org/10.1093/mnras/stz748) (Cited on page 85.)
- Radhakrishnan, V., & Cooke, D. J. 1969, Astrophys. Lett., 3, 225 (Cited on page 23.)
- Radhakrishnan, V., & Srinivasan, G. 1982, Current Science, 51, 1096 (Cited on page 83.)
- Rajwade, K., Stappers, B., Williams, C., et al. 2021, arXiv e-prints, arXiv:2103.08410. <https://arxiv.org/abs/2103.08410> (Cited on page 193.)
- Ramachandran, R., & Kramer, M. 2003, Astronomy & Astrophysics, 407, 1085, doi: [10.1051/0004-6361:20031036](https://doi.org/10.1051/0004-6361:20031036) (Cited on pages 172, 173, 174, 183, 187 and 188.)
- Rankin, J. 1993, ApJ, 405, 285, doi: [10.1086/172361](https://doi.org/10.1086/172361) (Cited on page 26.)

- Rankin, J. M. 1983, *ApJ*, 274, 359, doi: [10.1086/161451](https://doi.org/10.1086/161451) (Cited on page 22.)
- Rankin, J. M., Wright, G. A. E., & Brown, A. M. 2013, *Monthly Notices of the Royal Astronomical Society*, 433, 445, doi: [10.1093/mnras/stt739](https://doi.org/10.1093/mnras/stt739) (Cited on page 172.)
- Ransom, S. 2011, PRESTO: Pulsar Exploration and Search TOolkit. <http://ascl.net/1107.017> (Cited on pages 57, 79, 97, 110 and 118.)
- Ransom, S. M. 2001, PhD thesis, Harvard University (Cited on pages 57, 84 and 110.)
- Ransom, S. M., Cordes, J. M., & Eikenberry, S. S. 2003, *ApJ*, 589, 911, doi: [10.1086/374806](https://doi.org/10.1086/374806) (Cited on pages 68, 75 and 133.)
- Ransom, S. M., Eikenberry, S. S., & Middleditch, J. 2002, *AJ*, 124, 1788, doi: [10.1086/342285](https://doi.org/10.1086/342285) (Cited on pages 59, 60, 63, 68, 71, 110 and 122.)
- Ransom, S. M., Hessels, J. W. T., Stairs, I. H., et al. 2005, *Science*, 307, 892, doi: [10.1126/science.1108632](https://doi.org/10.1126/science.1108632) (Cited on pages 71, 75, 84 and 110.)
- Ransom, S. M., Stairs, I. H., Archibald, A. M., et al. 2014, *Nature*, 505, 520, doi: [10.1038/nature12917](https://doi.org/10.1038/nature12917) (Cited on pages 21, 38 and 41.)
- Ridley, J. P., Crawford, F., Lorimer, D. R., et al. 2013, *MNRAS*, 433, 138, doi: [10.1093/mnras/stt709](https://doi.org/10.1093/mnras/stt709) (Cited on page 95.)
- Ridolfi, A., Freire, P. C. C., Gupta, Y., & Ransom, S. M. 2019, *MNRAS*, 490, 3860, doi: [10.1093/mnras/stz2645](https://doi.org/10.1093/mnras/stz2645) (Cited on page 95.)
- Ridolfi, A., Gautam, T., Freire, P. C. C., et al. 2021, *MNRAS*, doi: [10.1093/mnras/stab790](https://doi.org/10.1093/mnras/stab790) (Cited on pages 54, 84, 110 and 120.)
- Riley, T. E., Watts, A. L., Bogdanov, S., et al. 2019, *ApJ*, 887, L21, doi: [10.3847/2041-8213/ab481c](https://doi.org/10.3847/2041-8213/ab481c) (Cited on page 36.)
- Ritchings, R. T. 1976, *Monthly Notices of the Royal Astronomical Society*, 176, 249, doi: [10.1093/mnras/176.2.249](https://doi.org/10.1093/mnras/176.2.249) (Cited on page 30.)
- Roberts, M. S. E. 2013, in *Neutron Stars and Pulsars: Challenges and Opportunities after 80 years*, ed. J. van Leeuwen, Vol. 291, 127–132, doi: [10.1017/S174392131202337X](https://doi.org/10.1017/S174392131202337X) (Cited on pages 41 and 94.)
- Romani, R. W., Narayan, R., & Blandford, R. 1986, *MNRAS*, 220, 19, doi: [10.1093/mnras/220.1.19](https://doi.org/10.1093/mnras/220.1.19) (Cited on page 33.)
- Roy, J., Gupta, Y., & Green, D. A. 2011, in *American Institute of Physics Conference Series*, Vol. 1357, *Radio Pulsars: An Astrophysical Key to Unlock the Secrets of the Universe*, ed. M. Burgay, N. D’Amico, P. Esposito, A. Pellizzoni, & A. Possenti, 54–55, doi: [10.1063/1.3615076](https://doi.org/10.1063/1.3615076) (Cited on page 94.)

- Sanidas, S., Caleb, M., Driessen, L., et al. 2017, Proceedings of the International Astronomical Union, 13, 406â407, doi: [10.1017/S1743921317009310](https://doi.org/10.1017/S1743921317009310) (Cited on page 91.)
- Sanidas, S., Cooper, S., Bassa, C. G., et al. 2019, A&A, 626, A104, doi: [10.1051/0004-6361/201935609](https://doi.org/10.1051/0004-6361/201935609) (Cited on pages 81, 86 and 158.)
- Sauls, J. A. 2019, arXiv e-prints, arXiv:1906.09641. <https://arxiv.org/abs/1906.09641> (Cited on page 21.)
- Scheuer, P. A. G. 1968, Nature, 218, 920, doi: [10.1038/218920a0](https://doi.org/10.1038/218920a0) (Cited on page 33.)
- Shannon, C. E. 1949, IEEE Proceedings, 37, 10, doi: [10.1109/JPROC.1998.659497](https://doi.org/10.1109/JPROC.1998.659497) (Cited on page 47.)
- Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607, doi: [10.1088/0004-637X/725/2/1607](https://doi.org/10.1088/0004-637X/725/2/1607) (Cited on page 170.)
- Shannon, R. M., OsÅowski, S., Dai, S., et al. 2014, Monthly Notices of the Royal Astronomical Society, 443, 1463, doi: [10.1093/mnras/stu1213](https://doi.org/10.1093/mnras/stu1213) (Cited on page 172.)
- Shao, L., Caballero, R. N., Kramer, M., et al. 2013, Classical and Quantum Gravity, 30, 165019, doi: [10.1088/0264-9381/30/16/165019](https://doi.org/10.1088/0264-9381/30/16/165019) (Cited on page 172.)
- Shao, L., Wex, N., & Zhou, S.-Y. 2020, Phys. Rev. D, 102, 024069, doi: [10.1103/PhysRevD.102.024069](https://doi.org/10.1103/PhysRevD.102.024069) (Cited on page 38.)
- Shao, M., & You, X. P. 2016, Acta Astronomica Sinica, 57, 517 (Cited on pages 172, 174, 183 and 187.)
- Shklovskii, I. S. 1968, Soviet Ast., 11, 749 (Cited on page 20.)
- Siemion, A. P. V., Bower, G. C., Foster, G., et al. 2012, ApJ, 744, 109, doi: [10.1088/0004-637X/744/2/109](https://doi.org/10.1088/0004-637X/744/2/109) (Cited on page 85.)
- Slabber, M. J., Manley, J., Mwangama, J., & Ventura, N. 2018, in Software and Cyberinfrastructure for Astronomy V, ed. J. C. Guzman & J. Ibsen, Vol. 10707, International Society for Optics and Photonics (SPIE), 110 – 118, doi: [10.1117/12.2311870](https://doi.org/10.1117/12.2311870) (Cited on page 91.)
- Staelin, D. H. 1969, Proceedings of the IEEE, 57, 724, doi: [10.1109/PROC.1969.7051](https://doi.org/10.1109/PROC.1969.7051) (Cited on pages 77 and 80.)
- Staelin, D. H., & Reifenstein, E. C. 1968, Science, 162, 1481, doi: [10.1126/science.162.3861.1481](https://doi.org/10.1126/science.162.3861.1481) (Cited on page 20.)
- Stappers, B., & Kramer, M. 2018, PoS, MeerKAT2016, 009, doi: [10.22323/1.277.0009](https://doi.org/10.22323/1.277.0009) (Cited on page 94.)
- Stappers, B. W., Kramer, M., Lyne, A. G., D'Amico, N., & Jessner, A. 2006, Chinese Journal of Astronomy and Astrophysics Supplement, 6, 298 (Cited on page 39.)

- Stappers, B. W., Hessels, J. W. T., Alexov, A., et al. 2011, *A&A*, 530, A80, doi: [10.1051/0004-6361/201116681](https://doi.org/10.1051/0004-6361/201116681) (Cited on page 86.)
- Stappers, B. W., Archibald, A. M., Hessels, J. W. T., et al. 2014, *ApJ*, 790, 39, doi: [10.1088/0004-637X/790/1/39](https://doi.org/10.1088/0004-637X/790/1/39) (Cited on page 41.)
- Stovall, K., Lynch, R. S., Ransom, S. M., et al. 2014, *ApJ*, 791, 67, doi: [10.1088/0004-637X/791/1/67](https://doi.org/10.1088/0004-637X/791/1/67) (Cited on pages 81 and 157.)
- Straal, S. M., & van Leeuwen, J. 2019, *A&A*, 623, A90, doi: [10.1051/0004-6361/201833922](https://doi.org/10.1051/0004-6361/201833922) (Cited on page 94.)
- Suresh, A., & Cordes, J. M. 2019, *ApJ*, 870, 29, doi: [10.3847/1538-4357/aaf004](https://doi.org/10.3847/1538-4357/aaf004) (Cited on page 32.)
- Tan, C. M., Lyon, R. J., Stappers, B. W., et al. 2018a, *MNRAS*, 474, 4571, doi: [10.1093/mnras/stx3047](https://doi.org/10.1093/mnras/stx3047) (Cited on page 86.)
- Tan, C. M., Bassa, C. G., Cooper, S., et al. 2018b, *ApJ*, 866, 54, doi: [10.3847/1538-4357/aade88](https://doi.org/10.3847/1538-4357/aade88) (Cited on page 81.)
- Tauris, T. M., & Janka, H.-T. 2019, *ApJ*, 886, L20, doi: [10.3847/2041-8213/ab5642](https://doi.org/10.3847/2041-8213/ab5642) (Cited on page 41.)
- Tauris, T. M., Langer, N., & Kramer, M. 2012, *MNRAS*, 425, 1601, doi: [10.1111/j.1365-2966.2012.21446.x](https://doi.org/10.1111/j.1365-2966.2012.21446.x) (Cited on page 28.)
- Tauris, T. M., & van den Heuvel, E. P. J. 2006, *Formation and evolution of compact stellar X-ray sources*, Vol. 39, 623–665 (Cited on page 28.)
- Tauris, T. M., Kramer, M., Freire, P. C. C., et al. 2017, *ApJ*, 846, 170, doi: [10.3847/1538-4357/aa7e89](https://doi.org/10.3847/1538-4357/aa7e89) (Cited on pages 28, 41 and 42.)
- Taylor, J. H. . 1992, *Philosophical Transactions of the Royal Society of London Series A*, 341, 117, doi: [10.1098/rsta.1992.0088](https://doi.org/10.1098/rsta.1992.0088) (Cited on page 169.)
- Taylor, J. H. 1974, , 15, 367 (Cited on page 56.)
- Taylor, J. H., & Huguenin, G. R. 1969, *Nature*, 221, 816, doi: [10.1038/221816a0](https://doi.org/10.1038/221816a0) (Cited on page 63.)
- Taylor, J. H., & Manchester, R. N. 1977, *ApJ*, 215, 885, doi: [10.1086/155426](https://doi.org/10.1086/155426) (Cited on page 81.)
- Terzan, A. 1968, *Academie des Sciences Paris Comptes Rendus Serie B Sciences Physiques*, 267, 1245 (Cited on page 109.)
- Thompson, C., & Duncan, R. C. 1996, *ApJ*, 473, 322, doi: [10.1086/178147](https://doi.org/10.1086/178147) (Cited on page 29.)

- Thornton, D., Stappers, B., Bailes, M., et al. 2013, *Science*, 341, 53, doi: [10.1126/science.1236789](https://doi.org/10.1126/science.1236789) (Cited on page 78.)
- Titus, N., Stappers, B. W., Morello, V., et al. 2019, *MNRAS*, 487, 4332, doi: [10.1093/mnras/stz1578](https://doi.org/10.1093/mnras/stz1578) (Cited on page 95.)
- Turtle, A. J., & Vaughan, A. E. 1968, *Nature*, 219, 689, doi: [10.1038/219689a0](https://doi.org/10.1038/219689a0) (Cited on page 80.)
- Turyshev, S. G., Shao, M., & Nordtvedt, K., J. 2004, *Astronomische Nachrichten*, 325, 267, doi: [10.1002/asna.200310205](https://doi.org/10.1002/asna.200310205) (Cited on page 38.)
- Urquhart, R., Bahramian, A., Strader, J., et al. 2020, *The Astrophysical Journal*, 904, 147, doi: [10.3847/1538-4357/abb6fc](https://doi.org/10.3847/1538-4357/abb6fc) (Cited on pages 111 and 194.)
- Valenti, E., Ferraro, F. R., & Origlia, L. 2007, *The Astronomical Journal*, 133, 1287, doi: [10.1086/511271](https://doi.org/10.1086/511271) (Cited on page 109.)
- van Cappellen, W. A., & Bakker, L. 2010, in 2010 IEEE International Symposium on Phased Array Systems and Technology, 640–647, doi: [10.1109/ARRAY.2010.5613297](https://doi.org/10.1109/ARRAY.2010.5613297) (Cited on page 86.)
- van den Heuvel, E. P. J., & Habets, G. M. H. J. 1984, *Nature*, 309, 598, doi: [10.1038/309598a0](https://doi.org/10.1038/309598a0) (Cited on page 21.)
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, 556, A2, doi: [10.1051/0004-6361/201220873](https://doi.org/10.1051/0004-6361/201220873) (Cited on page 86.)
- van Straten, W. . 2006, *ApJ*, 642, 1004, doi: [10.1086/501001](https://doi.org/10.1086/501001) (Cited on pages 24 and 174.)
- van Straten, W. 2004, *ApJS*, 152, 129, doi: [10.1086/383187](https://doi.org/10.1086/383187) (Cited on pages 24 and 174.)
- van Straten, W. 2013, *The Astrophysical Journal Supplement Series*, 204, 13, doi: [10.1088/0067-0049/204/1/13](https://doi.org/10.1088/0067-0049/204/1/13) (Cited on pages 24, 172, 173, 174, 175 and 179.)
- van Straten, W., Demorest, P., & Osłowski, S. 2012, *Astronomical Research and Technology*, 9, 237. <https://arxiv.org/abs/1205.6276> (Cited on page 108.)
- Venkatraman Krishnan, V., Bailes, M., van Straten, W., et al. 2020a, *Science*, 367, 577, doi: [10.1126/science.aax7007](https://doi.org/10.1126/science.aax7007) (Cited on pages 29 and 38.)
- Venkatraman Krishnan, V., Flynn, C., Farah, W., et al. 2020b, *MNRAS*, 492, 4752, doi: [10.1093/mnras/staa111](https://doi.org/10.1093/mnras/staa111) (Cited on pages 85 and 86.)
- Verbunt, F., & Freire, P. C. C. 2014, *aap*, 561, A11, doi: [10.1051/0004-6361/201321177](https://doi.org/10.1051/0004-6361/201321177) (Cited on pages 31, 84, 95 and 112.)

- Verbunt, F., & Hut, P. 1987, in *The Origin and Evolution of Neutron Stars*, ed. D. J. Helfand & J. H. Huang, Vol. 125, 187 (Cited on page 109.)
- Voisin, G., Cognard, I., Freire, P. C. C., et al. 2020, *A&A*, 638, A24, doi: [10.1051/0004-6361/202038104](https://doi.org/10.1051/0004-6361/202038104) (Cited on page 38.)
- von Hoensbroech, & Xilouris. 1997, *Astron. Astrophys. Suppl. Ser.*, 126, 121, doi: [10.1051/aas:1997255](https://doi.org/10.1051/aas:1997255) (Cited on pages 175 and 176.)
- Wang, H., & Sinnen, O. 2015, in *2015 International Conference on Field Programmable Technology (FPT)*, 240–243, doi: [10.1109/FPT.2015.7393158](https://doi.org/10.1109/FPT.2015.7393158) (Cited on page 158.)
- Wang, H., Zhu, W., Guo, P., et al. 2019, *Science China Physics, Mechanics, and Astronomy*, 62, 959507, doi: [10.1007/s11433-018-9388-3](https://doi.org/10.1007/s11433-018-9388-3) (Cited on page 158.)
- Wang, P., Li, D., Zhu, W., et al. 2018, *The Astronomer’s Telegram*, 11584, 1 (Cited on pages 94 and 111.)
- Weltevrede, P., Stappers, B. W., Rankin, J. M., & Wright, G. A. E. 2006, *ApJ*, 645, L149, doi: [10.1086/506346](https://doi.org/10.1086/506346) (Cited on page 30.)
- Wex, N., & Kopeikin, S. M. 1999, *ApJ*, 514, 388, doi: [10.1086/306933](https://doi.org/10.1086/306933) (Cited on page 38.)
- Wharton, R. S., Chatterjee, S., Cordes, J. M., Deneva, J. S., & Lazio, T. J. W. 2012, *ApJ*, 753, 108, doi: [10.1088/0004-637X/753/2/108](https://doi.org/10.1088/0004-637X/753/2/108) (Cited on page 40.)
- Wielebinski, R., VAUGHAN, A. E., & LARGE, M. I. 1969, *Nature*, 221, 47, doi: [10.1038/221047a0](https://doi.org/10.1038/221047a0) (Cited on page 81.)
- Will, C. M., & Zaglauer, H. W. 1989, *Astrophys. J.*, 346, 366, doi: [10.1086/168016](https://doi.org/10.1086/168016) (Cited on page 38.)
- Williamson, I. P. 1972, *MNRAS*, 157, 55, doi: [10.1093/mnras/157.1.55](https://doi.org/10.1093/mnras/157.1.55) (Cited on page 33.)
- Xiao, J., Li, X., Lin, H., & Qiu, K. 2019, *Monthly Notices of the Royal Astronomical Society*, 492, 2119, doi: [10.1093/mnras/stz3539](https://doi.org/10.1093/mnras/stz3539) (Cited on page 158.)
- Yao, J. M., Manchester, R. N., & Wang, N. 2017, *ApJ*, 835, 29, doi: [10.3847/1538-4357/835/1/29](https://doi.org/10.3847/1538-4357/835/1/29) (Cited on pages 33, 54, 79 and 138.)
- Yu, M. 2018, *ApJ*, 868, 8, doi: [10.3847/1538-4357/aae51a](https://doi.org/10.3847/1538-4357/aae51a) (Cited on page 63.)
- Zackay, B., & Ofek, E. O. 2017, *ApJ*, 835, 11, doi: [10.3847/1538-4357/835/1/11](https://doi.org/10.3847/1538-4357/835/1/11) (Cited on pages 56 and 97.)
- Zhang, B., Gil, J., & Dyks, J. 2007, *MNRAS*, 374, 1103, doi: [10.1111/j.1365-2966.2006.11226.x](https://doi.org/10.1111/j.1365-2966.2006.11226.x) (Cited on page 30.)


Zhang, B., Harding, A. K., & Muslimov, A. G. 2000, *ApJ*, 531, L135, doi: [10.1086/312542](https://doi.org/10.1086/312542) (Cited on page 27.)

Zhu, W. W., Berndsen, A., Madsen, E. C., et al. 2014, *ApJ*, 781, 117, doi: [10.1088/0004-637X/781/2/117](https://doi.org/10.1088/0004-637X/781/2/117) (Cited on pages 108, 118, 136, 156, 157, 158, 162, 163, 166, 168 and 195.)

Publication that accounts for major parts of the thesis

Refereed:

P. V. Padmanabh, E. D. Barr, D. J. Champion, R. Karuppusamy, M. Kramer, A. Jessner and P. Lazarus “*Revisiting profile instability of PSR J1022+1001*”, 2021, MNRAS, 500, 1, doi: [10.1093/mnras/staa3174](https://doi.org/10.1093/mnras/staa3174)



***I see the stars that etch along the lines of the blue sky,
I believe the universe is just another painting,
Articulatey done.
An epitome of creativity.***

***Stars scattered erratically, and yet symbolizing something.
Every aspect of the universe, filled with splendid colours.
I wonder if the colours of the night sky lie dry on the palette
of the painter.
What if the universe is just another incomplete painting?***

-Sushi