URSI AP-RASC, Sydney, Australia, 17 – 22 August 2025

Pulsar Searching with MeerKAT

Yunpeng Men*(1), Ewan Barr(1), MMGPS Collaboration
(1) Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany,
email:ypmen@mpifr-bonn.mpg.de

Abstract

Pulsar searches are a critical means by which natural laboratories for testing fundamental physics can be found. Pulsar discoveries enable the study of nanohertz-frequency gravitational waves, the limits of general relativity and alternative theories of gravity, and investigations into extreme states of matter. In this paper, we present the pulsar search instrumentation and pipeline developed for the MPIfR-MeerKAT Galactic Plane Survey (MMGPS). As a pathfinder project for the Square Kilometre Array (SKA), the MMGPS survey provides a valuable baseline for future SKA pulsar searches. We will present results and their implications for SKA, including pulsar discoveries, data facility, and suggested pipeline optimisations.

1 Introduction

Pulsars are highly magnetized, rapidly rotating neutron stars that emit beams of electromagnetic radiation. Their precise rotation periods make them ideal for studying fundamental physics, including: (1) nanohertz-frequency gravitational wave detection via pulsar timing arrays (PTAs), which are networks of highly stable millisecond pulsars (MSPs) that measure correlated pulse arrival times; (2) tests of gravity theories like Einstein's general relativity, using pulsars in binary systems with extreme gravitational fields; and (3) studies of the extreme states of matter within neutron stars, owing to pulsars' extremely high densities.

The aforementioned studies benefit significantly from the discovery of specific classes of pulsar; highly stable millisecond pulsars (MSPs), pulsars in compact binary systems, and massive pulsars. Dedicated pulsar surveys, such as Galactic plane surveys and globular cluster surveys, are conducted to identify these systems. Radio interferometers with large fields of view (FoV) and high sensitivity can substantially accelerate pulsar survey speeds. However, the large number of antennas and wide bandwidths of instruments like MeerKAT generates extremely high data rates $\sim$TB/s, which, when coupled with the high angular resolution of such arrays, poses significant data processing challenges.

This paper describes the data processing used in the MMGPS survey and discusses its implications for pulsar surveys in the upcoming SKA era.

2 Data processing facility

To support the pulsar survey with MeerKAT, we designed two high-performance computing clusters: the filterbanking beamformer user-supplied equipment (FBFUSE) for multi-beam beamforming, and the accelerated pulsar search user-supplied equipment (APSUSE) for pulsar searching, as shown in Figure 1.

The FBFUSE cluster, comprising 32 processing nodes, enables the generation of the many total-power beam required to make use of MeerKATs wide FoV. Each node is equipped with two Intel Xeon Gold 6134 CPUs, two 40GbE NICs, and two Nvidia GTX 1080 Ti GPUs, enabling an aggregate ingest data rate of 1.8 Tb/s. FBFUSE can generate up to 768 coherent beams and one incoherent beam (the sum of the detected antenna powers), with modelling of the coherent beams used for determining optimal placement [1]. The cluster also includes a transient buffer capable of storing 56 seconds of baseband voltage data from the full MeerKAT array, facilitating fast radio burst (FRB) localization and polarization analysis.

The APSUSE cluster, dedicated to pulsar searching, comprises 8 data capture nodes and 60 processing nodes. Each capture node is equipped with two Intel Xeon Gold 6136 CPUs, a 40GbE NIC, and a 56Gb/s InfiniBand (IB) NIC, enabling a data capture rate of 280 Gb/s from the FBFUSE cluster. The processing nodes each feature two Intel Xeon Silver 4116 CPUs, a 56 Gb/s IB NIC, and two Nvidia GTX 1080 Ti GPUs. Both the FBFUSE and APSUSE clusters share a 3.5-PB distributed filesystem which uses BeeGFS and provides read/write speeds up to 40 GB/s.

3 Pulsar searching pipeline

The MMGPS pulsar search pipeline involves several stages: radio frequency interference (RFI) mitigation, dedispersion, acceleration search, candidate folding, and machine learning-based candidate classification.

3.1 RFI mitigation

Radio signals are often contaminated by human-made RFI signals, such as satellite and communication signals. This contamination leads to false positive detections and reduces

Figure 1. The APSUSE cluster, a high-performance computing system dedicated to pulsar searching. The cluster is located at the site of the MeerKAT radio telescope in the Karoo region of South Africa.

the signal-to-noise ratio (S/N) of genuine astronomical signals. Therefore, RFI mitigation is a crucial step in pulsar searching. In the MMGPS processing pipeline, we preprocess the beamformed input data using FILTOOL from the PulsarX package [2], applying several filters: (1) a Skewness-Kurtosis Filter (SKF), which removes narrowband RFI based on the skewness and kurtosis distribution across frequency channels; (2) a Zero-DM Matched Filter (ZDMF), effective in removing long-timescale RFI and baseline variations; and (3) a Kadane Filter (KF), which removes narrowband RFI with slow variations. As a recent upgrade to the pipeline, we also implemented an incoherent beam subtraction method, described in more detail below, following [3] to further mitigate baseline variations and near-field RFI .

3.2 Dedispersion

As radio waves propagate through the cold plasma of the interstellar medium, dispersion causes frequency-dependent delays. Dedispersion, which corrects these delays across frequency channels, is essential for maximizing the S/N of astronomical signals. The dispersion delay in cold plasma with a given dispersion measure (DM) is described by:

$$\tau_d = \frac{e^2}{2\pi m_e c}\mathrm{DM}\left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right), \tag{1}$$

where e, m_e, and c are the elementary charge, electron mass, and speed of light in a vacuum, respectively. While a brute-force dedispersion algorithm can be used, more computationally efficient methods exist, such as sub-band dedispersion and the Fast Discrete Dispersion Measure Transform (FDMT). In our pipeline, we employ the GPU-based DEDISP library, which implements an efficient parallel brute-force approach. To accelerate processing, we implement a dedispersion plan that downsamples the raw data by over different DM ranges according to the minimum dispersion-induced temporal smearing in that range. The MMGPS search covers a DM range from 0 to $3000\,\mathrm{pc\,cm^{-3}}$.

3.3 Acceleration search

After dedispersion, a time series is generated for each DM trial, followed by a Fast Fourier Transform (FFT) with dereddening and birdie exclusion to search for periodic signals. However, the orbital motion of pulsars in binary systems introduces Doppler shifts that modulate their periods, making them difficult to detect using a simple FFT. Two primary methods address this challenge: Time Domain Acceleration Search (TDAS) and Fourier Domain Acceleration Search (FDAS). TDAS resamples the time series data with different acceleration values before performing the FFT, effectively demodulating the signal. The resampled time series at the time t', relates to the original time t, by

$$\begin{aligned} t' - t'_0 &= (t - t_0) + \frac{a}{2c}(t - t_0)^2 \\ t_0 &= \frac{1}{2}T_{\mathrm{obs}} \\ t'_0 &= \frac{1}{2}T_{\mathrm{obs}} - \frac{a}{8c}T_{\mathrm{obs}}^2, \end{aligned} \tag{2}$$

where T_{obs} is the observation length of the time series, and a is an acceleration trial. FDAS, on the other hand, applies matched filters to the Fourier series using templates derived from different acceleration values, after the FFT is performed on the raw time series. Because pulsar profiles are not purely sinusoidal, a harmonic summation step is crucial to recover signal power distributed over higher spin harmonics. Our pipeline uses the highly optimized, GPU-based acceleration search software PEASOUP [4], which implements the TDAS algorithm. In our search, we apply accelerations ranging from $-50\,\mathrm{m/s^2}$ to $50\,\mathrm{m/s^2}$, and sum up to 32 harmonics.

3.4 Candidate folding

The acceleration search yields candidate pulsars, characterized by their DM, period, period derivative, and S/N. To classify these candidates, folding the data using their respective parameters is essential to generate diagnostic plots, including pulse profiles, dynamic spectra, and time-resolved profile variations. This candidate folding process involves several key steps: (1) The raw data is subband-dedispersed to reduce its volume, thereby accelerating the folding step; (2) The subbanded data is then folded using the candidate's period and period derivative, creating an archive data file for each candidate; (3) Finally, the candidate's DM, period, and period derivative are refined by maximizing its S/N. To reduce computational load, our pipeline filters out candidates below an S/N threshold of 8 prior to folding, resulting in approximately 50 candidates per beam. This folding is performed using the highly optimized PSRFOLD_FIL program within PulsarX [2], reducing the folding time from months to days, a significant improvement over other software.

3.5 Candidate classification

Pulsar surveys generate a large number of false positives, including persistent RFI and noise signals, particularly due to the use of hundreds of beams. Traditional pulsar candidate classification relies on manual inspection or empirical

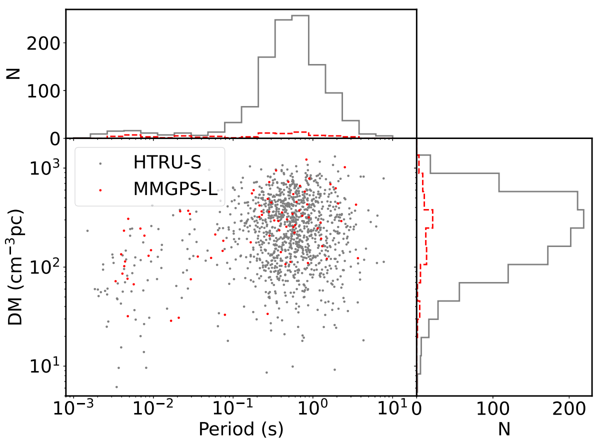

Figure 2. Period-DM diagram comparing pulsars found in the MMGPS survey with those from the HTRU-S survey with Parkes in the south hemisphere.

feature extraction. To improve classification efficiency, machine learning-based classifiers have been developed, such as the convolutional neural network (CNN) based PICS [5]. We employ a PICS score threshold of 0.1, which reduces the number of candidates requiring manual review by a factor of ten.

4 Results and implications for pulsar surveys with SKA

As a pathfinder survey for the SKA, the MMGPS provides valuable insights for future SKA pulsar surveys, including expected pulsar discoveries, data processing facility, and pulsar processing pipeline.

4.1 Pulsar discoveries

The MMGPS survey encompasses L-band, S-band and Galactic center surveys [6]. The MMGPS L-band survey concluded with the discovery of 78 pulsars, including 17 confirmed binary systems, two of which are potential double neutron star systems. This higher fraction of binary systems compared to the known pulsar population can be attributed to the short 10-minute integration time, which enhances the sensitivity of acceleration searches to binaries. Because of its high sensitivity, as expected, our survey also detected more pulsars with higher DMs than the High Time Resolution Universe (HTRU) survey, as shown in Figure 2.

In the MMGPS survey, we utilized only the MeerKAT core array to cover the FoV, minimizing the number of beams needed with short baselines. In comparison, the sensitivity of the SKA-mid array (phase 1) core array is approximately twice that of the MeerKAT core array, which will include additional dishes with slightly larger size. Under a similar survey configuration, the observable volume of the Galactic plane is estimated to be three times larger than that of the MMGPS survey. Assuming a complete sky coverage, a threefold increase in known pulsar detection is expected, potentially yielding 2000 new pulsars. However, the SKA surveys require real-time pulsar searching on FPGAs and GPU. This strictly limits the parameter space that can be explored with the SKA, with searches for binary systems only possible over 600 DM trials. This approach can reduce the sensitivity to pulsar binary systems by a factor of 2.

4.2 Extensible infrastructure

The MeerKAT data processing facility was built with flexibility and extensibility in mind. Through the User Supplied Equipment (USE) programme, SARAO provides access to MeerKAT voltages, allowing teams to build plug-in instrumentation that can expand MeerKATs scientific capabilities. Examples include the aforementioned FBFUSE and APSUSE instruments, but also TUSE (real-time transient search), BLUSE (SETI search), PTUSE (pulsar timing), and FBF-BVRUSE (full-array baseband recording). While some of these were anticipated from the outset, many were not. The FBF-BVRUSE extension, for instance, was completely unforeseen: it provides a full-array voltage recording system, enabling the most sensitive pulsar search ever conducted in the Galactic center. By contrast, the SKA currently employs a more rigid design that does not include such “spigot” technology. While such a design may allow the SKA to perform observations efficiently, the prospects for extending the instruments to meet the needs of emerging scientific goals are limited.

4.3 Auto-correlation subtraction

The MeerKAT pulsar search pipeline uses a range of techniques for RFI mitigation. In addition to the use of the PULSARX package described above, the results of the searches are use as feedback to create masks for frequency bands with weak but persistent RFI that cannot be identified by automatic zapping algorithms and to generate lists of known periodic RFI that can be removed in the Fourier domain prior to acceleration searching. Such capability will also be required for the SKA. However, the most impactful change to our RFI mitigation has come in the last year with the introduction of auto-correlation subtraction from the FBFUSE beams. This technique, outlined in [3], acts to remove correlated rednoise and RFI signals by subtracting the incoherent beam (the sum of the array auto-correlations) from each coherent beam synthesized by FBFUSE. This upgrade to the signal processing chain was made possible by the flexibility built into the FBFUSE design, which uses commodity-off-the-shelf hardware with consumer GPUs. This capability has proven invaluable for recent MeerKAT observations, reducing the number of spurious candidates and greatly increasing our sensitivity to longer period pulsars. It is strongly recommended that the SKA adopt such a mode in their wide-FoV beamformer for transient searching. Figure 3 demonstrates the power of the technique when observing slow pulsars.

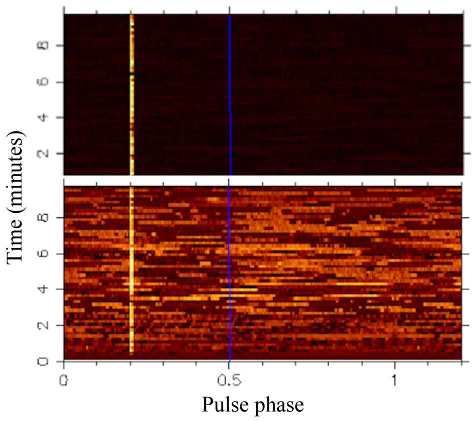

Figure 3. A simultaneous observation of the 8-s pulsar PSR J2144−3933 with two FBFUSE beams. Each panel shows the total intensity of the FBFUSE beam as a function of pulse phase and observation time, with the top panel generated from a beam using auto-correlation subtraction and the bottom panel from a beam without auto-correlation subtraction.

4.4 Pipeline testing and sensitivity estimation

The current instrumentation at MeerKAT allows for the recording of long ($\sim$ 10-hour) observations with hundreds of beams. This enables MeerKAT to not only perform incredibly deep observations of targets of interest but it enables detailed offline investigation of the beamformed data which is critical to understanding and addressing issues in the search processing. To do this, we use signal injection testing, where realistic pulsar signals are injected into real recorded beams offline before they are processed. This technique allows for the true sensitivity profile of our pipelines, across the complete parameter space and in the presence of real RFI, to be determined and gives the benchmarks against which changes to the pipelines can be evaluated. By moving to a fully “online” search system the SKA will lack the ability to perform deep wide-FoV time domain observations and pipeline testing as is performed at MeerKAT will become significantly more complex. We see high value in SKA incorporating, at least in an engineering mode, the kind of recording capability available at MeerKAT in order to retain these benefits.

5 Summary

In this paper, we have presented the data processing facility and pipeline developed for pulsar searching with MeerKAT, a precursor to the SKA. We have also discussed the implications for pulsar searching in the SKA era. We already have an operational pulsar search system on MeerKAT, and it would be a missed opportunity not to use this in parallel with the first SKA array build that supports pulsar searching. This would allow us to directly compare the products of both systems while retaining the ability to record large numbers of beamformed products. In general, we are concerned that much of the pulsar search and survey capability will be lost when MeerKAT is integrated into SKA. It is clear that, in the long term, SKA will be a phenomenal instrument for pulsar discovery. However, just like with MeerKAT, it will take time to tune classifiers and RFI pipelines to their optimal performance. With overlap between the two observatories, this transition period could be shortened, enabling a smoother progression from pulsar searches on MeerKAT to those on the SKA.

6 Acknowledgements

The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. SARAO acknowledges the ongoing advice and calibration of GPS systems by the National Metrology Institute of South Africa (NMISA) and the time space reference systems department of the Paris Observatory. TRAPUM observations used the FBFUSE and APSUSE computing clusters for data acquisition, storage and analysis. These clusters were funded and installed by the Max-Planck-Institut für Radioastronomie and the Max-Planck-Gesellschaft. YPM and EB acknowledge continuing support from the Max Planck society.

References

[1] Chen, W. et al., “Wide Field Beamformed Observation with MeerKAT”, <i>Journal of Astronomical Instrumentation</i>, vol. 10, no. 3, Art. no. 2150013-178, WSPC, 2021. doi:10.1142/S2251171721500136.

[2] Men, Y. et al., “PulsarX: A new pulsar searching package. I. A high performance folding program for pulsar surveys”, <i>Astronomy and Astrophysics</i>, vol. 679, Art. no. A20, 2023. doi:10.1051/0004-6361/202347356.

[3] Roy, J. et al., “A Post-correlation Beamformer for Time-domain Studies of Pulsars and Transients”, <i>The Astrophysical Journal</i>, vol. 864, no. 2, Art. no. 160, IOP, 2018. doi:10.3847/1538-4357/aad815.

[4] Barr, E., “Peasoup: C++/CUDA GPU pulsar searching library”, <i>Astrophysics Source Code Library</i>, Art. no. ascl:2001.014, 2020. ascl:2001.014.

[5] Zhu, W. W. et al., “Searching for Pulsars Using Image Pattern Recognition”, <i>The Astrophysical Journal</i>, vol. 781, no. 2, Art. no. 117, IOP, 2014. doi:10.1088/0004-637X/781/2/117.

[6] Padmanabh, P. V. et al., “The MPIfR-MeerKAT Galactic Plane Survey - I. System set-up and early results”, <i>Monthly Notices of the Royal Astronomical Society</i>, vol. 524, no. 1, OUP, pp. 1291–1315, 2023. doi:10.1093/mnras/stad1900.