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

FRB candidate optimizer for SPOTLIGHT, enabling real-time commensal searches

Kshitij Bane*(1), Jyotirmoy Das(1), Arpan Pal(1), Ujjwal Panda(1), Deepak Bhong(1), Santaji N. Katore(1), Sanjay S. Kudale(1), Mekhala Muley(1), S Harshavardhan Reddy(1), Jayanta Roy(1), Jayaram Chengalur(1), Karel Adamek(2), Kenil Ajudiya(4) , Wes Armour(3), Siddhartha Bhattacharyya(1), Ritavash Debnath(5), Nishant P. Deo(6), Chahat Dudeja(1), Shelton Gnanaraj(1), Ajith Kumar(1), Sangita Kumari(1), Shilkumar Meshram(1), Raghav Wani(4), Bhaswati Bhattacharyya(1), Yash Bhusare(1), Ankita Ghosh(1), Ruta Kale(1), Ajay Kumar(1), Banshi Lal(1), Krishnakumar MA(1), Yogesh Maan(1), Puja Majee(1), Visweshwar Ram Ram Marthi(1), Sapan Kumar Sahoo(1), Saptarshi Sarkar(1), Rahul Sharan(1)
(1) NCRA-TIFR, Pune 411007, India, kshitijbane@gmail.com
(2) Silesian University in Opava, Department of Physics, Bezrucovo nam. 13, Opava, 74601
(3) Oxford e-Research Centre, University of Oxford, Oxford OX1 3QG, United Kingdom
(4) Indian Institute of Science, Bengaluru-560012, India
(5) Indian Institute of Science Education and Research, Pune 411008, India
(6) Indian Institute of Science Education and Research, Kolkata 741246, India

Abstract

The SPOTLIGHT project is a commensal survey for Fast Radio Bursts (FRBs) and pulsars at the uGMRT using real-time HPC and AI for simultaneous time-domain detection and arc-second imaging localisation across 300 to 1460 MHz. FRBs are searched within 2,000 post-correlation beams steered across the Field of View (FoV), with SPOTLIGHT leveraging its PetaFlops of computing power to process PetaBytes of data per day in real time. The AI-powered trigger for genuine astronomical signals enables capturing voltage samples spanning the burst duration to study the spectro-temporal-polarimetric properties at microsecond time resolution as well as high-time resolution visibility for event localisation. With an unprecedented sensitivity, SPOTLIGHT is expected to find hundreds of localised FRBs. This paper presents the FRB candidate optimiser for SPOTLIGHT's FRB detection pipeline, which incorporates candidate clustering, classification, and triggering for visibility and raw voltage recording.

1. Introduction

Fast Radio Bursts (FRBs) are highly energetic transient signals of extragalactic origin lasting only a few milliseconds [1]. Their origins are still debated, making them at the forefront of astronomical research. Precisely localising FRBs in the sky is crucial for understanding their origin [2], characterising their local environments [3, 4], and utilising them as cosmological probes [5, 6]. Traditional survey telescopes with large field-of-view (FoV) are capable of detecting FRBs but lack the resolution to know their precise origins. The uGMRT, being an interferometer, is an ideal telescope that simultaneously offers time-domain detection and imaging localisation.

2. The SPOTLIGHT System

SPOTLIGHT [7] is a real-time multi-beam commensal project that will run commensally with regular uGMRT observations to search for FRBs and pulsars within the FoV. It leverages a state-of-the-art, multi-node GPU-powered compute cluster (funded by the National Supercomputing Mission) for real-time processing, forming visibilities, and generating 2000 post-correlation beams at a 1.3 ms resolution while ingesting data at a rate of 25 GB/s [8]. Correlation and beamforming are performed on 16 servers equipped with 32 Nvidia A100 GPUs, while the beams are transmitted to the transient search cluster, which consists of 24 servers with 48 Nvidia A100 GPUs. A multi-beam multi-node FRB detection pipeline (described below) is developed to run on the transient search cluster.

3. SPOTLIGHT's FRB detection pipeline

The multi-beam FRB detection pipeline for SPOTLIGHT consists of two main stages: Single pulse search and candidate optimiser.

3.1. Single pulse search

As an FRB signal travels through ionized plasma along the line of sight, different frequencies propagate at different speeds, causing dispersion. The dispersion measure (DM) quantifies this effect, representing the column density of electrons along the line of sight. To correct for this, a process called dedispersion is applied to the data. However, since the exact dispersion is unknown beforehand, dedispersion is performed across multiple trial DM values. This process involves collapsing the frequency-time intensity data, or dynamic spectrum, into a time series while accounting for frequency-dependent delays at each DM value. The dedispersion module of the

SPOTLIGHT performs ~8500 DM trials covering up to 2000 pc cm^{-3}. The dedispersed time series are searched for single pulses with widths of up to 0.5 seconds using a fast implementation of boxcar match-filtering. The candidates above a signal-to-noise ratio (SNR) threshold are passed through a peak filtering module, which conducts multiple passes using fixed-size stencils over the DM and time-of-arrival (TOA) plane to remove redundant detections. The AstroAccelerate software package provides GPU-optimised implementations of dedispersion, and the single pulse search algorithm [9, 10], and is integrated into the SPOTLIGHT transient search pipeline.

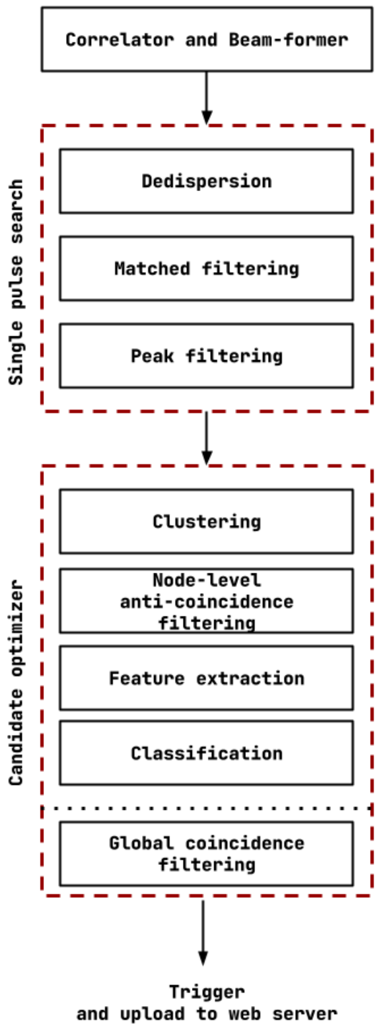


Figure 1. The FRB detection pipeline running on the transient search cluster. Each node processes multiple beams (up to 85), with all steps up to classification executed independently on each node, handling all assigned beams.

3.2. Candidate optimizer

The single pulse search pipeline can generate anywhere from a few thousand to several million candidates within a 10-minute observing span. Many of these candidates will be spurious resulting from Radio Frequency Interference (RFI), even after zero-DM subtraction applied during the single pulse search. This will give false triggers to the subsequent stages which will overload the system. Additionally, SPOTLIGHT's unprecedented sensitivity captures multiple bursts from the same events, clustered in DM-Time space.

Since the SPOTLIGHT system is designed to accept one trigger every 10 minutes for follow-up processing - such as imaging localization (using high-time-resolution visibilities) and spectro-temporal-polarimetric analysis (using Nyquist-sampled baseband data from all antennas) - a robust candidate optimization module is essential to filter out spurious candidates.

3.2.1. Clustering:. We cluster candidates in the DM-time plane for each beam, reducing the number of candidates by a few orders of magnitude. The clustering is performed using DBSCAN [11], which is a density-based spatial clustering technique implemented on multi-node multi-GPUs via `cuml`. The candidates are spread across the DM and time-of-arrival (TOA) space and the closely packed points are grouped together. DBCAN does not require the number of clusters in the data to be known a priori and can find clusters of arbitrary shape. We then select the highest SNR candidate in the given cluster. This reduces the number of candidates from several million to a few thousand, or even a few hundred. Note that the clustering is performed independently for each beam in a node. After clustering, the filtered candidates from each beam in a given node are combined to perform "anti-coincidence" filtering i.e. removal of candidates that are seen in isolated beams, which are most likely due to RFI. This further reduces the number of candidates by a factor of ~5.

3.2.2. Feature extraction: The dedispersed dynamic spectrum and the DM transform (DMT) are characteristic features of FRBs that can distinguish them from RFI. The DM transform is formed by dedispersing the signal over multiple DMs, and stacking the time series to create a 2D array. For real FRB-like signals, it shows a distinctive bow-tie pattern, as seen in Figure 3. `candies`, a GPU-optimised feature extraction program for FRBs, extracts these features for each candidate from its beamformed data, based on its TOA and DM. `candies` is a variant of `your`, optimised for low-frequency observations. The extracted features are stored in an HDF5 file for each candidate. The HDF5 format allows storing and organising large amounts of data.

3.2.3 Classification:. The extracted features are then fed to `FETCH` (Fast Extragalactic Transient Candidate Hunter) [12] - a convolutional neural network (CNN) based binary classifier. FETCH classifies each candidate as either an FRB or RFI, and labels them accordingly as 0 (for RFI) or 1 (for FRB). A "coincidence filter" is then applied to all the true candidates, in order to filter out the candidates which appear in multiple beams. Candidates are grouped based on their TOAs, with the highest SNR candidate selected for each TOA. The current implementation of the classifier and coincidence filter reduces the number of candidates to approximately 20 within a 10-minute observing run. The final selected candidates are uploaded to the SPOTLIGHT web server [13]. For each of these candidates, we trigger the recording of 1.3 ms visibilities, and Nyquist-sampled raw voltage data.

4. Testing and results

The SPOTLIGHT system is currently undergoing rigorous functional testing as various modules are being commissioned. For example, we carried out a search of giant pulses from the Crab Pulsar (B0531+21) over the full SPOTLIGHT frequency spectrum (300 - 1460 MHz). The SPOTLIGHT system was used to form 800 beams and each beam's data was recorded for ~500 s. This data was processed using the offline variant of the pipeline, which ran across 8 nodes, with each node handling 100 beams.

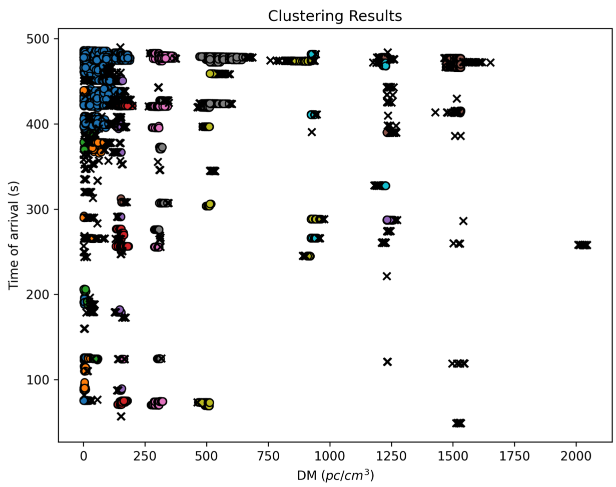
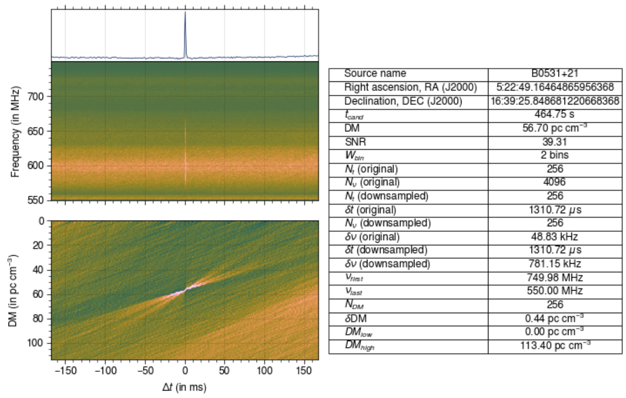


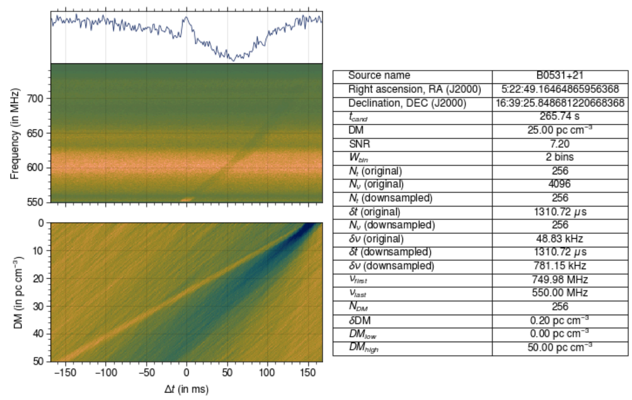
Figure 2. Results of clustering for Crab Pulsar data in Band 4 (550-750 MHz) for one beam out of 800. Different colours show the different clusters formed by DBSCAN in the DM-time plane. The candidates marked "x" are the outliers.

After the single pulse search, the number of candidates/beam was 10,000. After clustering this number reduced to ~100. Figure 2 shows the clustering output from DBSCAN. Only the highest SNR candidate from each cluster was forwarded for the next stages i.e. feature extraction and classification. After classification, there were approximately 5 Label 1 candidates per beam, resulting in a total of ~2000 candidates from 500 seconds of data in all the beams. These candidates were seen in most of the beams simultaneously. Applying the coincidence filter gave 5 bursts as the final output. Figure 3 shows an example of Label 0 and a Label 1 candidate.

As it can be seen Label 1 candidates are true Crab Giant Pulses.



(a) Label 1 candidate



(b) Label 0 candidate

Figure 3. An example of classification outcomes from Crab Pulsar observation in Band 4 (550-750 MHz). Label 1 is the true candidate, Label 0 is a spurious candidate.

5. Summary

We are commissioning the FRB detection pipeline of the SPOTLIGHT system while conducting functional validation with 800 simultaneous beams. During the initial test runs we were able to detect single pulses from multiple pulsars including giant pulses from Crab pulsar. The candidate optimizer is able to reduce the number of candidates by a factor of ~200. This significantly reduces the number of false triggers. We have deployed a fully functional end-to-end pipeline in offline mode, which is now being migrated to real-time operation.

6. Acknowledgements

We acknowledge the funding of the Param Rudra data centre at the GMRT under NSM Phase 3. We also acknowledge support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.02-0700 for the contributions towards the overhead cost. The GMRT is run by the National Centre for Radio Astrophysics of the Tata Institute of Fundamental Research, India. The data centre was jointly built by team members from C-DAC and the GMRT. We thank the GMRT telescope operators for supporting the observations during the software prototyping and commissioning phases.

References

[1] D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, and F. Crawford, "A Bright Millisecond Radio Burst of Extragalactic Origin," Science, vol. 318, no. 5851, pp. 777–780, Nov. 2007, doi: 10.1126/science.1147532.

[2] The CHIME/FRB Collaboration et al., "A bright millisecond-duration radio burst from a Galactic magnetar," arXiv e-prints, p. arXiv:2005.10324, May 2020.

[3] S. P. Tendulkar et al., "The 60 pc Environment of FRB 20180916B," The Astrophysical Journal, vol. 908, p. L12, Feb. 2021, doi: 10.3847/2041-8213/abdb38.

[4] H. Xu et al., "A fast radio burst source at a complex magnetised site in a barred galaxy," arXiv e-prints, p. arXiv:2111.11764, Nov. 2021.

[5] J.-P. Macquart et al., "A census of baryons in the Universe from localized fast radio bursts," Nature, vol. 581, no. 7809, pp. 391–395, May 2020, doi: 10.1038/s41586-020-2300-2.

[6] Z.-W. Zhao et al., "Cosmological Parameter Estimation for Dynamical Dark Energy Models with Future Fast Radio Burst Observations," ApJ, vol. 903, no. 2, p. 83, Nov. 2020, doi: 10.3847/1538-4357/abb8ce.

[7] J. Roy, J.N. Chengalur, and the SPOTLIGHT team, "SPOTLIGHT: A Probe of the Fast Radio Transient Sky", URSI-RCRS 2024

[8] H. Reddy and the SPOTLIGHT team, "A multi-beam backend for SPOTLIGHT, enabling fast transient surveys with the GMRT," Abstract, URSI AP-RASC, 2025.

[9] J. Novotný, K. Adámek, M. A. Clark, M. Giles, and W. Armour, "Accelerating Dedispersion Using Many-core Architectures," ApJS, vol. 269, no. 1, p. 29, Nov. 2023, doi: 10.3847/1538-4365/acfef6.

[10] K. Adámek and W. Armour, "Single-pulse Detection Algorithms for Real-time Fast Radio Burst Searches Using GPUs," The Astrophysical Journal Supplement Series, vol. 247, p. 56, Apr. 2020, doi: 10.3847/1538-4365/ab7994.

[11] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in Second International Conference on Knowledge Discovery and Data Mining (KDD'96). Proceedings of a conference held August 2-4, 1996, pp. 226–331.

[12] D. Agarwal, K. Aggarwal, S. Burke-Spolaor, D. R. Lorimer, and N. Garver-Daniels, "FETCH: A deep-learning based classifier for fast transient classification," Monthly Notices of the Royal Astronomical Society, vol. 497, no. 2, pp. 1661–1674, Sep. 2020, doi: 10.1093/mnras/staa1856.

[13] S. N. Katore and the SPOTLIGHT team, "A web-based tool listing discovery database with transient alert generation for the SPOTLIGHT," URSI-RCRS 2024.