



The FRB-searching Pipeline of the Tianlai Cylinder Pathfinder Array

Zijie Yu^{1,2}, Furen Deng^{1,2}, Shijie Sun^{1,3}, Chenhui Niu⁴, Jixia Li^{1,3}, Fengquan Wu^{1,3}, Wei-Yang Wang², Yougang Wang^{1,3}, Shifan Zuo^{1,3}, Lin Shu^{5,6}, Jie Hao^{5,6}, Xiaohui Liu^{1,2}, Reza Ansari⁷, Ue-Li Pen^{8,9}, Albert Stebbins¹⁰, Peter Timbie¹¹, and Xuelei Chen^{1,2,3,12,13}

¹ National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; wufq@bao.ac.cn, xuelei@cosmology.bao.ac.cn

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Key Laboratory of Radio Astronomy and Technology, Chinese Academy of Science, Beijing 100101, China

⁴ Institute of Astrophysics, Central China Normal University, Wuhan 430079, China

⁵ Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

⁶ Guangdong Institute of Artificial Intelligence and Advanced Computing, Guangzhou 510535, China

⁷ Université Paris—Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France

⁸ Institute of Astronomy and Astrophysics, Academia Sinica, Roosevelt Road, Taipei 10617, Taiwan, China

⁹ Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 Saint George Street, Toronto, ON M5S 3H8, Canada

¹⁰ Department of Physics, University of Wisconsin—Madison, Madison, WI 53706, USA

¹¹ Fermi National Accelerator Laboratory, P.O. Box 500, Batavia IL 60510–5011, USA

¹² Department of Physics, College of Sciences, Northeastern University, Shenyang 110819, China

¹³ Center of High Energy Physics, Peking University, Beijing 100871, China

Received 2024 March 29; revised 2024 June 11; accepted 2024 June 21; published 2024 July 31

Abstract

This paper presents the design, calibration, and survey strategy of the Fast Radio Burst (FRB) digital backend and its real-time data processing pipeline employed in the Tianlai Cylinder Pathfinder Array. The array, consisting of three parallel cylindrical reflectors and equipped with 96 dual-polarization feeds, is a radio interferometer array designed for conducting drift scans of the northern celestial semi-sphere. The FRB digital backend enables the formation of 96 digital beams, effectively covering an area of approximately 40 square degrees with the 3 dB beam. Our pipeline demonstrates the capability to conduct an automatic search of FRBs, detecting at quasi-real-time and classifying FRB candidates automatically. The current FRB searching pipeline has an overall recall rate of 88%. During the commissioning phase, we successfully detected signals emitted by four well-known pulsars: PSR B0329+54, B2021+51, B0823+26, and B2020+28. We report the first discovery of an FRB by our array, designated as FRB 20220414A. We also investigate the optimal arrangement for the digitally formed beams to achieve maximum detection rate by numerical simulation.

Key words: instrumentation: interferometers – methods: data analysis – (stars:) pulsars: general – telescopes

1. Introduction

Fast Radio Bursts (FRBs) have become a focus of current research in radio astronomy since the initial discovery of the Lorimer burst (Lorimer et al. 2007; Thornton et al. 2013). FRBs are believed to originate mostly from sources beyond our Galaxy, as they exhibit a dispersion measure (DM) that exceeds what is expected from the interstellar medium within our own Galaxy along the same line of sight. This hypothesis was confirmed when FRB 20121102A was localized to a host galaxy with a redshift of $z = 0.193$ (Bassa et al. 2017; Chatterjee et al. 2017; Marcote et al. 2017). Thousands of FRB bursts have been detected by radio telescopes worldwide, and more than 790 FRB sources have been published, with a majority ($\sim 70\%$) by CHIME (CHIME/FRB Collaboration et al. 2021), including 66 repeaters. Despite the unknown origin and radiation mechanism of FRBs, these observations have significantly advanced our understanding of this enigma. The Blinkverse database (Xu et al. 2023) gathers the information for about 7900 radio bursts from

these sources. For an overview of FRB properties and the progress in identifying their sources and corresponding astrophysical mechanisms, see, e.g., Caleb & Keane (2021).

FRB signals typically last for several to tens of milliseconds, necessitating telescopes equipped with sub-millisecond sampling backends for detection. Beam forming techniques are also commonly employed in FRB searches to enhance sky coverage and localization. As a consequence, a significant volume of data is generated. There are two approaches to tackling this issue. One option is to allocate an adequate number of storage servers. Alternatively, real-time data processing can be implemented, storing only the data containing potential FRB candidates while discarding the remaining data.

This paper presents the design and commissioning results of the digital FRB backend for the Tianlai Cylinder Pathfinder Array. The Tianlai experiment consists of a dish pathfinder array (Wu et al. 2021) and a cylinder pathfinder array (Zhang et al. 2016; Zuo et al. 2019; Li et al. 2020, 2021;

Table 1
Basic Parameters of the Tianlai Cylinder Pathfinder Array

Parameters	Values
Number of cylinders	3
Reflector E–W diameter	15 m
Reflector N–S length	40 m
Number of receivers	96
f/D	0.32
Average SEFD	67.18 kJy
Latitude	44.15°N
Longitude	91.80°E
Current Observing frequency	685–810 MHz

Sun et al. 2022), which aim to test key technologies for cosmological surveys using the 21-cm intensity mapping method (Chen 2012; Xu et al. 2014). The cylinder pathfinder array consists of three reflectors, and is equipped with a total of 96 dual-polarization receivers. The basic parameters and configuration of the array are shown in Table 1 and Figure 1, respectively. Initially, these arrays were equipped with correlators that provided visibilities with a second-level sampling rate (Niu et al. 2019), which is sufficient for sky map reconstruction (Zuo et al. 2021). However, the cylinder array has a relatively wide field of view (FoV), covering approximately 96 square degrees in its -3 dB primary beam, which allows it to be used for a blind search of FRBs, once a high-speed digital backend is equipped. In this paper we introduce the design, implementation and commissioning test of the digital transient backend of the cylinder array, which is added to run in parallel with the original digital correlator backend. Similarly, a digital backend for the dish array has also been built and installed, though it handles fewer inputs and forms only 16 beams (Yu et al. 2022b).

This paper is organized as follows. In Section 2, after a brief review of the antenna and analog electronic system of the Tianlai Cylinder Pathfinder Array, we describe the beam forming principle and its hardware implementation, followed by the transient searching pipeline. System operation and performance are described in Section 3, including calibration and test of digital beam forming efficiency, and a study of the sensitivity of the system. We have characterized the system and the detection pipeline by observing bright pulsars and through mock FRB injection. Beam arrangement optimization is discussed in Section 4 and we report the detection of one FRB event during our test observation campaign in Section 5. Finally, we conclude in Section 6.

2. System Design

2.1. Analog System of the Array

The analog part of the cylinder array is unchanged, and a detailed description has been given in Li et al. (2020). In Table 6

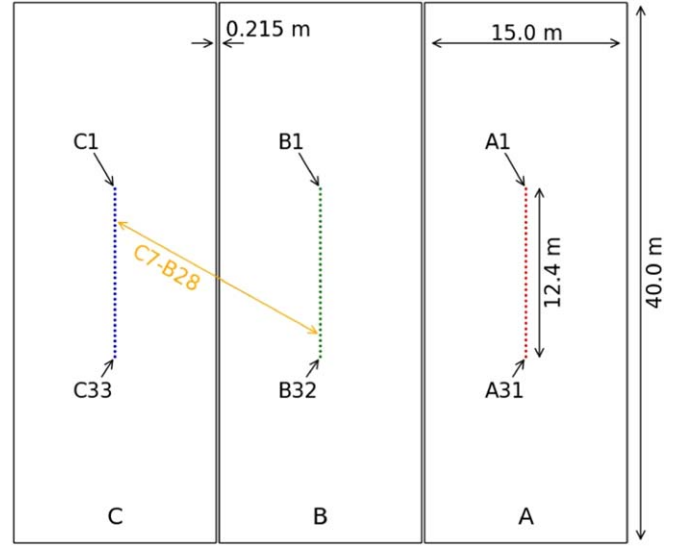


Figure 1. The configuration of Tianlai Cylinder Pathfinder reflectors and the layout of receiver feed numbers (Li et al. 2020). Each rectangle represents a cylinder reflector. The geometric parameters, including the length (40 m), width (15 m) of the reflector and the gap between them (0.215 m), are annotated. The dots in the center of each cylinder mark the position of a dual-polarization feed, which are denoted by the letters A, B, C with a sequential number. A baseline is labeled by its pair of feeds, e.g., the baseline made by feed C7 and B28 is labeled as C7-B28.

we list its main parameters. Three cylinder reflectors, designated as the A, B, and C cylinder, are adjacent to each other, with axis in the North-South (N–S) direction. Receiver feeds are arranged in a uniformly spaced linear array along the focus line of each cylinder, designated A1, A2, ... C33, with the same maximum distance between the two ends of the array (12.4 meters), but with slightly different number of feeds on each cylinder (31, 32, 33 respectively), to suppress grating lobes of the array (Zhang et al. 2016). An interferometric baseline can be specified by the designation of the pair of feeds (Figure 1). The 96 dual linear polarization receiver feeds generate a total of 192 outputs of analog radio signals. These are then amplified by low noise amplifiers (LNAs) located on the feeds. The amplified signals are converted to optical signals and transported via optical fiber cable to a station where the digital facilities are housed. At the digital facilities, the optical signals are converted back to radio signals and undergo bandpass filtering to select a 100 MHz bandwidth, which is currently set to 700–800 MHz. The signals are then down-converted to an intermediate frequency (IF) range of 135–235 MHz using a mixer. Subsequently, each of the IF signals is split into two outputs, then amplified to compensate for the reduced power, and fed to the original digital correlator and the new digital transient backend. Both the original digital correlator and the new FRB backend actually digitize the IF signal within 125–250 MHz independently, which are mapped to its original frequency in the data processing.

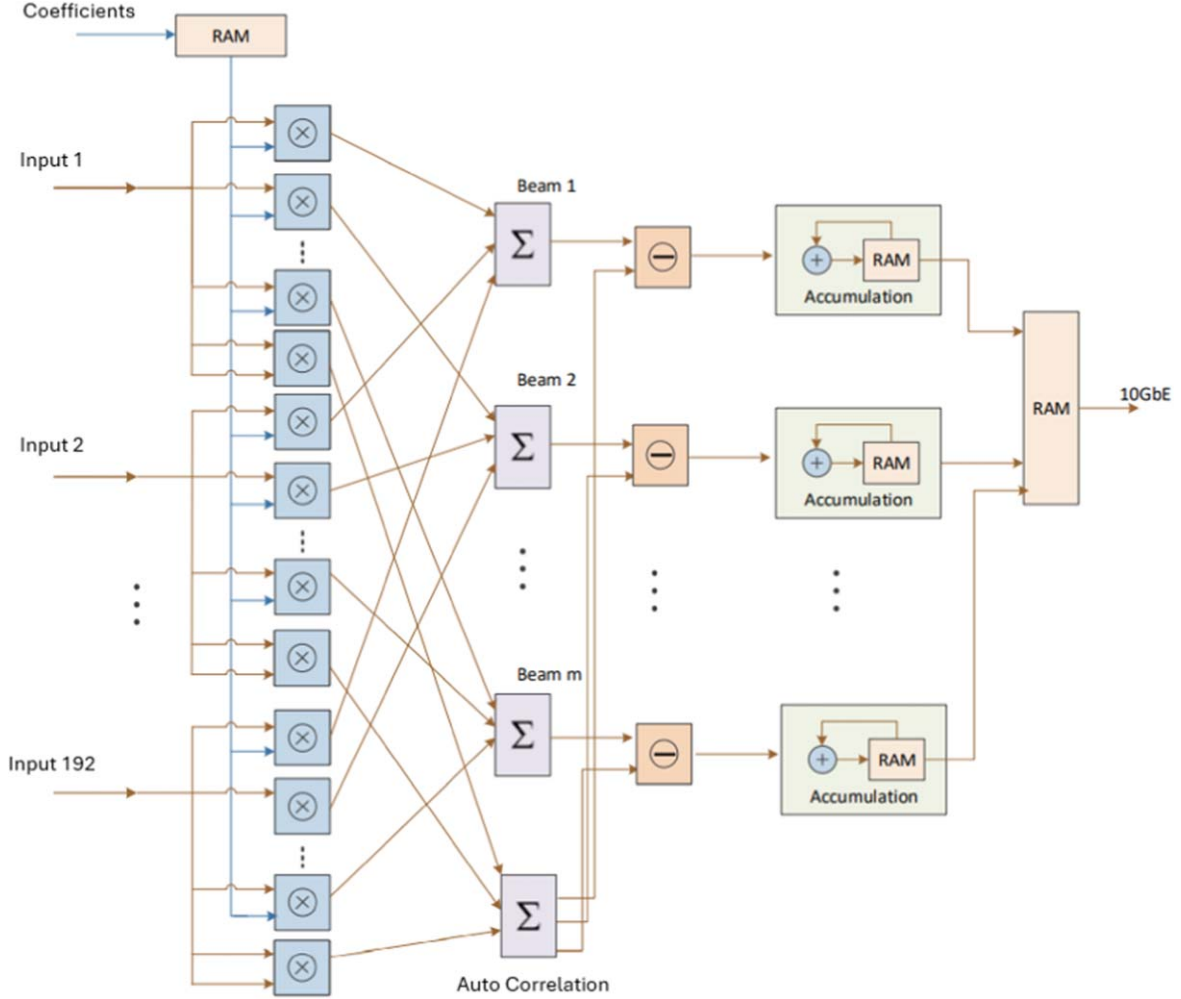


Figure 2. The principle of beam forming. Input signal is first digitized and then multiplied with complex coefficients, then added up to form beams. The auto-correlation of all inputs is summed and can be subtracted from beams to reduce noise. The beam data are accumulated according to the required time resolution.

2.2. Beam forming Principle

Figure 2 illustrates the principle of beam forming. To form a digital beam, we can add up the voltage output from each array element with a complex weight,

$$J(\mathbf{k}) = \sum_a w_a(\mathbf{k}) \mathcal{E}_a, \quad (1)$$

where $w_a(\mathbf{k})$ denotes the complex weight of antenna a for beam direction \mathbf{k} . The voltage of element a is given by

$$\mathcal{E}_a = g_a \int e^{-2\pi j \mathbf{n} \cdot \mathbf{u}_a} A_a(\mathbf{n}) E(\mathbf{n}) d^2 \mathbf{n} + \eta_a, \quad (2)$$

where $A_a(\mathbf{n})$ describes the voltage response for antenna a , $\mathbf{u}_a \equiv \mathbf{x}_a/\lambda$ is the position vector of the antenna in unit of wavelength, the integration is over sky directions, g_a is the complex gain of the instrument for unit a , and η_a represents the noise for that channel. We write $g_a = |g_a| e^{j\phi_a}$ to extract the

overall delay of the signal transport chain. The complex weights are set as

$$w_a(\mathbf{k}) = \frac{1}{|g_a|} e^{-j\phi_a} e^{2\pi j \mathbf{k} \cdot \mathbf{u}_a}, \quad (3)$$

then the phases of all antennas are equal for the direction $\mathbf{n} = \mathbf{k}$. The power of this beam is given by

$$\begin{aligned} S(\mathbf{k}) &\propto \langle |J(\mathbf{k})|^2 \rangle \\ &= \sum_{a,b} \int A_a^*(\mathbf{n}) A_b(\mathbf{n}) I(\mathbf{n}) e^{-2\pi j (\mathbf{n}_a - \mathbf{n}_b) \cdot \mathbf{u}_{ab}} d^2 \mathbf{n} + \sum_a |g_a|^{-2} \langle |\eta_a|^2 \rangle, \end{aligned} \quad (4)$$

where $\mathbf{u}_{ab} = \mathbf{u}_b - \mathbf{u}_a$, and $I(\mathbf{n})$ is the sky radiation intensity in direction \mathbf{n} . If a burst occurs in the direction \mathbf{q} at some moment, and assuming that the sky in other directions and the receiver noise remain nearly constant at this time, then the received

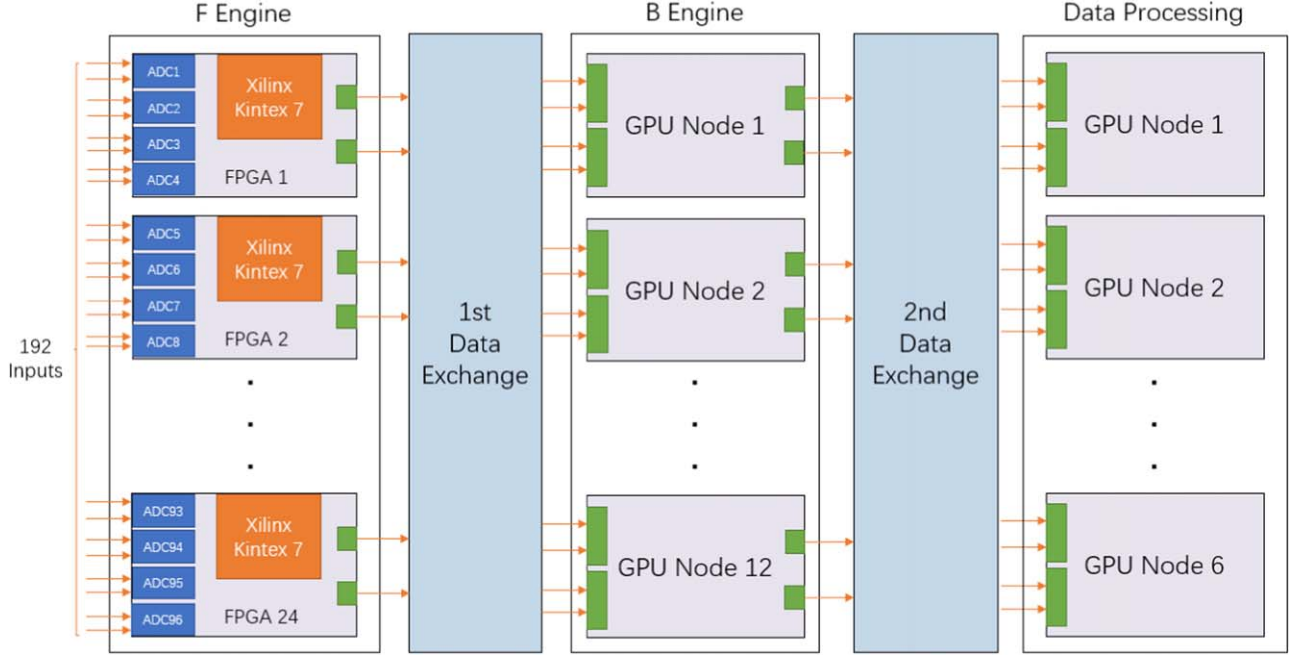


Figure 3. The hardware configuration of the beamformer. The F engine consists of 24 FPGA boards, and each is equipped with four ADCs and a Xilinx Kintex-7 FPGA to digitize eight inputs. Each green box in the figure represents a QSFP port for data transfer. The first data exchange is accomplished with two 48-port switchboards. The B engine consists of 12 GPU nodes, each having two GPU servers. A 48-port switchboard is used to process the second data exchange. The 96 beams are searched using six data processing GPU servers.

signal is given by $S = S_0 + \Delta S$, where

$$\Delta S = \sum_{ab} e^{-2\pi j(q-k) \cdot u_{ab}} A_a^*(q) A_b(q) I(q), \quad (5)$$

and S_0 is the S value without a burst. Multiple beams can be formed at the same time, and for each of the beams, we can search for radio pulses with dispersion delay. This is realized by the beamformer system, which performs the digital beam forming, and the beam data are then streamed to de-dispersion servers, which search for the dispersed radio pulse in the data at real time.

2.3. The Beamformer

The beamformer consists of an F engine which performs Fast Fourier Transform (FFT) to convert the data from the time domain to the frequency domain, and a B engine which forms beams at each frequency by a weighted summation of the signals from different feeds. We have a total of 96 dual linear polarization feeds, generating a total of 192 signal inputs. Each of these are digitized by the F engine and connected to the B engine, but due to limited processing power, currently we are only forming 96 digital beams, and for this we choose to use a single polarization. The choice of signal input can be made in real time by specifying the weight, and we are now usually using the x -polarization, i.e., the one aligned in the N-S

direction, which has less cross-coupling between the feeds. The weights of the unused channels are set to zero.

The hardware setup of the system is illustrated in Figure 3. The F engine is implemented on 24 sets of analog-to-digital (ADC) converter boards, with each digitizing eight analog signals and exporting data through two QSFP ports. The B engine is based on 12 GPU nodes, each handling a block of frequency channels for all beams formed out of the digital signals. The data are transferred from the F engine to the B engine through the 1st Data Exchange which includes two 48-port switchboards, each handling half of the frequencies. The 1st Data Exchange re-packages the Fourier transformed data from different receivers according to their frequency, with the same frequency to the same processing unit in the B engine. The output of the B engine goes through the 2nd Data Exchange, which includes another 48-port switchboard. The 2nd Data Exchange packages the data from different frequencies according to their beam, thus generating 96 beams, which are transferred to the de-dispersion nodes for further data processing.

In Figure 4 we draw the workflow of the beamformer. The digitized data from the ADC are transformed to the frequency domain by FFT. Prior to the FFT, the data are first pre-processed. This includes a shifting of a few blocks of data in time for each signal channel separately to compensate for different time delays, which arise from the long optical fibers

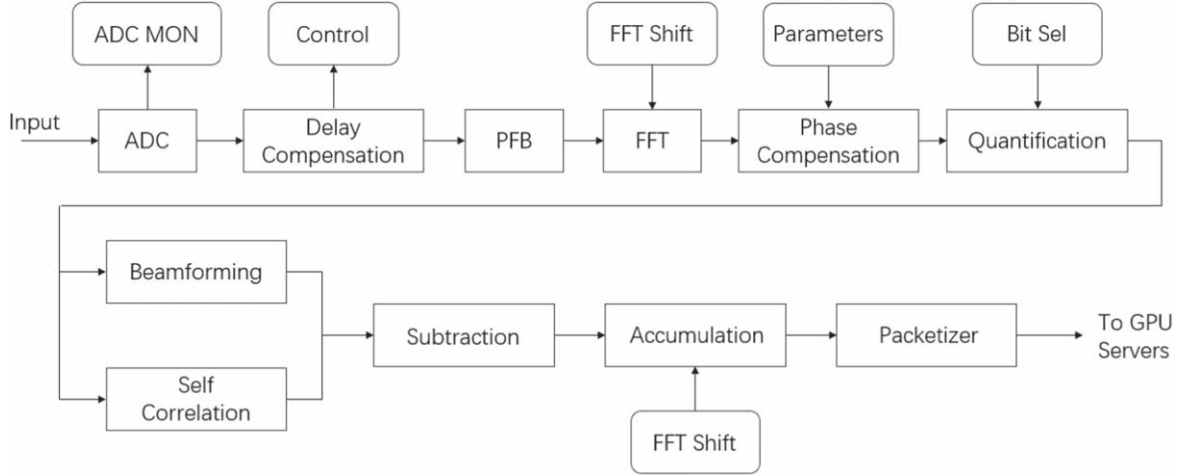


Figure 4. The workflow of the beamformer.

connecting the antenna to the analog and digital electronics in the station house. To mitigate leakage and minimize sidelobes, a polyphase filter bank (PFB) is employed, which uses a Hamming function with four taps and 2048 samples. Following the PFB, we proceed with a 2048-sample FFT on the pre-processed data. After the FFT, a phase compensation for the residue instrument delay is added. The time block shift prior to the PFB and the phase shift after FFT are determined by calibration measurement and set by the system control, the compensating phase can be changed by command. Once these are done, the data are cut off to the required number of bits, then the beams are formed by summing the results with the appropriate complex weighting factors. An auto-correlation for each channel is also calculated, which can be subtracted subsequently to reduce the average output. Each of the beams is then integrated (accumulated) to yield the output at millisecond level, and the results are packaged to be sent to the GPU servers for dispersion and pulse search.

2.3.1. F engine

Each ADC board is equipped with a Xilinx Kintex-7 FPGA, four TI ADS5407 ADCs, a 72 MB QDR, two QSFP ports for high-speed data transmission, and an Ethernet port for control and configuration. The ADCs on each board sample signals from two feeds at a sampling rate of 250 Msps.

To achieve synchronization for the ADC boards, a `sync` signal is sent to all boards. The rising edge of the `sync` signal triggers all the boards to simultaneously reset their status and clock. Shortly after this, the control starts transmitting a `trigger` signal, with its rising edge triggering the boards to commence writing data into a first-in-first-out (FIFO) at the next rising edge of the `syncout` signal. When the falling edge of the `trigger` signal is detected, the boards output the data from the FIFO. Thus, the output data are synchronized.

2.3.2. Switchboard

To establish the 1st Data Exchange system with 96 ports, two switchboards are utilized, and each is equipped with 48 SFP28 10 GbE ports and eight QSFP28 100 GbE ports. These switchboards are interconnected through the use of five QSFP28 ports. Prior to transmission to the B engine, data from the F engines undergo rearrangement. Specifically, data from different input feeds with identical frequencies are grouped together, enabling the servers within the B engine to form beams at each frequency with all the data needed.

Similarly, the data outputted by the B engine are also rearranged by the 2nd Data Exchange system, which is realized with one switchboard of 48 ports, so that the different frequencies of the same beam are directed to the same processing node for de-dispersion and pulse search.

2.3.3. B Engine

Each GPU node consists of two servers and each server is equipped with an NVIDIA RTX 3060 and 8 GB VRAM. The workflow of the B engine is illustrated in Figure 2. Within this workflow, data from different frequency channels and feeds are weighted by a coefficient matrix, followed by summation to generate the 96 beams. Subsequently, the data are accumulated to achieve a time resolution of 0.1 ms. Lastly, the B engine sends the data to the de-dispersing nodes.

2.4. Transient Searching Pipeline

The beamformed stream data are processed on six GPU servers, with each being equipped with two Intel Xeon E5-2699 V3 CPUs, two NVIDIA RTX 3080 GPUs, and 314 GB RAM. Each GPU server is connected to the beamformer via four 10 GbE cables, with each cable carrying data for four beams. The workflow of the transient searching pipeline is

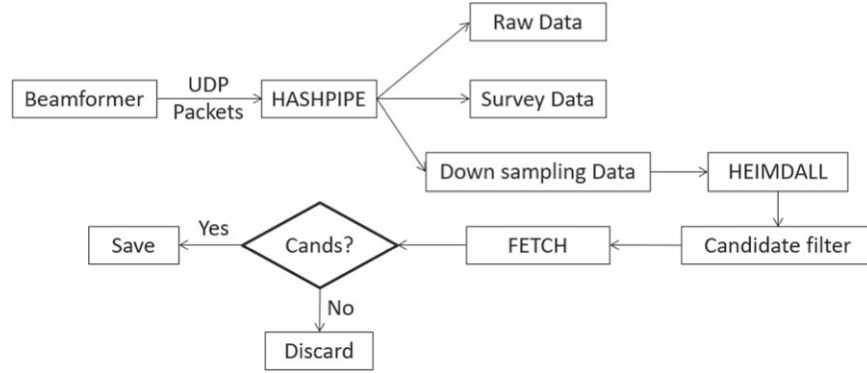


Figure 5. The workflow of the transient searching pipeline. HASHPIPE is utilized to process the UDP packets from beamformer and generate three data streams. The raw data stream stores full time and frequency resolution data. The survey data stream is accumulated to a 10.06 second integration interval for sky map reconstruction. The downsampling data reduce both time and frequency resolution to 1/4 and are for FRB quasi-real-time searching. We utilized the HEIMDALL program, candidate filter and the FETCH program to search and classify FRB candidates.

drawn in Figure 5. The data are transmitted from the beamformer in UDP format. At the GPU server end, the High Availability Shared Pipeline Engine (HASHPIPE)¹⁴ is utilized to receive data packets and write them into SIGPROC¹⁵ Filterbank format files, with each file containing 81.2 s of data from one beam.

Once a data file is generated, we employ HEIMDALL,¹⁶ a GPU-accelerated single pulse-searching software (Barsdell et al. 2012), to perform de-dispersion and identify transient candidates within the file. The candidate events identified by HEIMDALL are then subjected to a candidate filter, which applies multiple criteria to eliminate obvious radio frequency interference (RFI). The remaining candidates are classified using FETCH,¹⁷ a deep learning-based transient classifier (Agarwal et al. 2020). If any of the candidates are classified as potential FRBs, the pipeline saves data files from all beams to disk and sends notification emails to administrators. In the absence of any candidates, the pipeline discards the raw data and candidate searching files to free up disk space. The processing time may vary slightly but is generally shorter than the observing time covered by the file.

Note that this part of the pipeline works independently from the beam forming pipeline, so in principle its hardware or software can also be replaced, as long as the alternative system can handle the data and complete the pulse-searching task in comparable or shorter time. It is conceivable that in the future, by employing faster algorithm, e.g., the Hough transform method (Zuo & Chen 2020), the processing cycle time could be further shortened.

2.4.1. Data Stream

Four instances of the HASHPIPE program are created on each server to process data from one cable. Each instance generates three child threads. The net child thread reads data from the packets and writes them into ring buffer 1. The cal thread reads data from ring buffer 1, reshapes the data array into SIGPROC’s Filterbank format, and writes them into ring buffer 2. The output thread reads data from ring buffer 2 and writes them into Filterbank format data files.

The search for pulses in beamformed raw data, with a time and frequency resolution of 0.098304 ms and 0.122 MHz respectively, incurs significant computational load. To reduce this load, the output thread divides the raw data stream from the beamformers into three separate data streams, each serving a different purpose. The first data stream is dedicated to storing the raw data to preserve all information in case of an FRB detection. The second data stream undergoes downsampling by a factor of 4 in both time and frequency, resulting in data with a resolution of 0.393 ms and 0.488 MHz. All quasi-real-time processes mentioned below are performed on this downsampled data stream. The third data stream is accumulated to a time resolution of 10.06 s, which can be utilized for imaging with this backend.

2.4.2. Calibrator Noise Source

The Tianlai array utilizes a noise source to perform relative phase calibration. This calibrator noise source is activated periodically for a few seconds every few minutes, depending on the specific settings. However, the presence of these noise signals leads to the identification of numerous candidates when using HEIMDALL for single pulse searches. This significantly increases the workload of the pipeline and obscures genuine signals that may be present in the vicinity of the noise. To mitigate this interference, we apply a zero DM removal

¹⁴ <https://casper.astro.berkeley.edu/wiki/HASHPIPE>

¹⁵ <https://sigproc.sourceforge.net/>

¹⁶ <https://sourceforge.net/projects/heimdall-astro/>

¹⁷ <https://github.com/devanshkv/fetch>

Table 2
HEIMDALL Parameters

Parameters	Values
DM range	0–1500 pc cm ⁻³
SNR loss tolerance	1.25
DM trials	377
Maximum boxcar width	128 samples
Pulse width	0.393–50.3 ms
Zap channels	0–140, 246–256
Masked frequency	741.6–810.0, 685.0–689.9 MHz

procedure to the data, by subtracting the frequency-averaged value at each sample point.

2.4.3. De-dispersion and Pulse Searching

Two parallel HEIMDALL threads are executed on each GPU server to process data from 16 beams. The HEIMDALL configuration is presented in Table 2. Burst searches are conducted over a range of DM from 0 to 1500 pc cm⁻³, with a default Signal-to-Noise Ratio (SNR) loss tolerance of 1.25. This results in a total of 377 DM trials. The output time interval is 0.393 ms, while the maximum width of the boxcar filter (rectangular in time domain) is set to 128 samples, corresponding to a pulse width of 50.3 ms. According to the Blinkverse FRB catalog (Xu et al. 2023), the width of most FRBs falls within this range: among 3418 recorded FRB pulse widths (including FAST and CHIME samples), only 15 of them are smaller than 0.393 ms, and seven of them are larger than 50.3 ms.

If a burst occurs near the end of the data block, the delayed low frequency part of the burst may not fall within the present data block, then it could be missed in the search of the present block. We solve this end-of-file problem by copying the 8 s of data at the end of the file to the start of the next data file, so that such a missed pulse can still be detected at the next block. To mitigate the impact of RFIs, frequency channels that are heavily contaminated are masked using the “zap channel” option of HEIMDALL. The masked frequency is set empirically based on past observations and is fixed in FRB-search during all observations, while data within these bands will still be saved for off-line studies.

After the search, HEIMDALL produces a candidate file containing the identified potential candidates, each accompanied by information such as the SNR, candidate time, DM, and boxcar width. Upon completion of processing for all beams, the HEIMDALL-provided *coincider* script is utilized to summarize the candidates. This script groups together candidates with matching times and similar DM values from different beams, effectively consolidating them into a single candidate. Additionally, the *coincider* script records the

number of beams in which a candidate is detected and identifies the beam with the highest SNR for each candidate.

2.4.4. Candidates Filtering

HEIMDALL produces a substantial number of candidates, ranging from hundreds to thousands per minute. However, the majority of these candidates are attributed to background noise fluctuations, the noise source used for calibration, and RFI. We implement a filtering process to reject the non-astronomical candidates based on several criteria.

First, we discarded faint candidates below a certain threshold, at present this threshold is set at SNR = 8. Additionally, candidates with a boxcar width of 0 or exceeding 7 are also discarded, as they are likely to be RFIs with pulse widths either too small (<0.393 ms) or too large (>50.3 ms) compared to typical FRBs at 750 MHz.

Second, candidates detected in more than 20 beams are excluded, as astronomical signals are expected to be detected only in beams pointing toward nearby directions. Strictly speaking, this criterion should depend on how the beams are placed, as the digitally formed beams can be directed at any direction, and they could in principle overlap with each other, where the number of detected beam may not be a good criterion. However, currently we have placed the beams apart from each other in our observations, and this criterion mostly rejects nearby RFI which affects all beams.

In cases where a pulse is detected in more than one beam, HEIMDALL reports multiple candidates, one for each beam, which unnecessarily increases the computational cost during classification. Thus, when such simultaneous candidates are present, the one with the highest SNR is used for candidate classification as described below. Note that if a pulse is classified as real, the data corresponding to that time frame are saved, and in the later off-line processing the detection in multiple beams could be utilized in analysis.

2.4.5. Candidate Classification

We currently use an artificial intelligence program *FETCH* to further classify the remaining candidates after the filtering process. The *FETCH* program uses different combinations of Convolutional Neural Network (CNN) models to process the Frequency-Time (FT) and DM-Time (DMT) matrices, and replaces the top classification layer of both models with a dense layer to generate ensemble models, borrowing techniques developed in the image processing community (Agarwal et al. 2020).

The training data set for *FETCH* is chosen from three different telescope backends, consisting of simulated FRBs, real pulsar bursts, and real RFIs. The validation data set includes 6664 real pulsar bursts and 7319 RFIs. Real FRBs and RFIs observed with ASKAP, Parkes and BL are used to

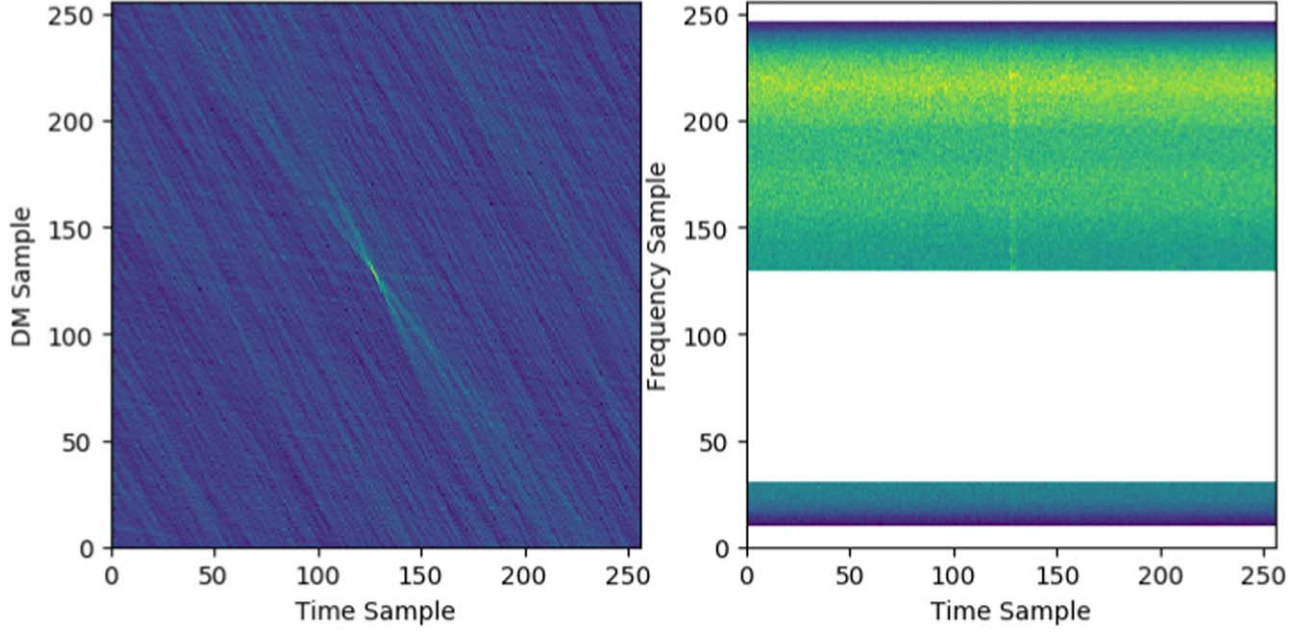


Figure 6. An example of FETCH input using the data of FRB 20220414A. The left panel displays the DMT matrix, and the right panel features the FT matrix, for the time frame which contains the burst. The figure shows a total of 256 DM values and 256 time values, centered at the detected event. The 256 frequencies are downsampled from our original 1024 frequency channels. The white region is contaminated by RFIs and masked.

Table 3
FETCH Models and their Performance Tested with Real ASKAP and Parkes FRBs and RFIs (Agarwal et al. 2020)

Label	FT Model	DMT Model	k	ASKAP FRBs (/33)	Mislabeled ASKAP RFIs (/10639)	Mislabeled Parkes RFIs (/478)
a	DenseNet121 (4)	Xception (21)	256	33	2	0
b	DenseNet121 (4)	VGG16 (2)	32	28	5	48
c	DenseNet169 (11)	Xception (21)	112	33	16	6
d	DenseNet201 (7)	Xception (21)	32	33	12	29
e	VGG19 (4)	Xception (21)	128	33	16	7
f	DenseNet169 (11)	VGG16 (2)	512	33	2	1
g	VGG19 (4)	VGG16 (2)	128	29	1	10
h	DenseNet201 (7)	InceptionResNetV2 (34)	160	33	43	40
i	DenseNet201 (7)	VGG16 (2)	32	33	15	52
j	VGG19 (4)	InceptionResNetV2 (34)	512	33	33	3
k	DenseNet121 (4)	InceptionV3 (31)	64	33	7	70

Note. The number in brackets after each model is the number of unfrozen layers. k is the fusion hyperparameter. Models a, c, and f are used in this work.

evaluate the model performance. FETCH Models and their performance are listed in Table 3.

FETCH provides a script (`candmaker.py`) which reads the candidate files generated by our pipeline, and generates two data matrices: one for DMT and another for FT, as illustrated in Figure 6. The resulting data matrices are then stored in an HDF5 file. Subsequently, the `FETCH predict.py` script reads all the HDF5 files in this batch, then utilizes the selected models to assess the likelihood of each candidate being an astronomical signal, and outputs the classification results in a CSV file.

We combine the FETCH models a, c and f to classify the candidates by summing their predictions. These three models have a 100% recall rate when tested with the real ASKAP FRBs and have a relatively small amount of mislabeled RFIs. Models b and g missed several ASKAP FRBs, while models d, h, i, j, k performed poorly against RFIs, so these are not used. Between models e and f, we choose model f to introduce diversity in the DMT model.

Based on the predicted probability, the candidates are classified into three groups: RFI, pulsar, and potential FRB.

Candidates with an average likelihood below a threshold value of 0.4 are discarded as RFIs. This threshold is chosen to recover most of the pulses in the PSR B0329+54 observation. Candidates with an average likelihood above 0.4 and a DM value exceeding 50 are labeled as potential FRBs. Those with DM less than 50 but $\text{SNR} > 100$ are also classified as potential FRBs, as such pulses are not frequently produced by pulsars. Pulses with an average likelihood above 0.4, DM value less than 50, and SNR below 100 are classified as pulsars. In the event of identifying a potential FRB candidate, our pipeline promptly sends a notification email to operators for further human inspection.

3. System Operation and Performance

The digitally formed beams can be pointed at any direction by adjusting the phase of the complex weighting factor. Currently, we have arranged the beams in a pattern similar to our array itself: the 96 beams are put in three columns along the sky meridian circle, i.e., in N-S direction. The separation between columns and the separation within each column are adjustable parameters, which are optimized for maximum detection rate. The area near the zenith would generally have a higher detection rate as it corresponds to the peak of the antenna primary beam.

3.1. Calibration

To obtain the complex gain $g_a = |g_a|e^{j\phi_a}$ needed for beam forming, we make a calibration observation of a bright radio source. We form beams which include only a pair of feeds, with one of these a reference feed, and the other the feed to be calibrated, while the weights for all other feeds are all set to zero; in this way we determine the phase difference of each feed with respect to the reference feed. As the beamformer allows forming multiple beams at the same time, we can determine all the instrument phases in one observation. For feeds C1, ..., C33 and B17, ..., B32, we use A1 as the reference feed, while for feeds B1, ..., B16 and A1, ..., A31 we use C33 as reference feed, so that only the longer baselines, which are more sensitive to phase variation and have less cross-couplings, are used for calibration. The A1-C33 measurement determines the phase difference between these two reference points, and all other phases can then be derived with respect to one of them.

We model the primary beam as a Gaussian function, and the output of the beam is fitted as

$$\begin{aligned} \text{Re } I_{ab}(t) = & c_1^2 [\cos(\Delta\varphi_{ab}(t) + \phi_{ab}) + c_2^2] \\ & \times \exp\left(-\frac{\Delta\varphi_{ab}^2(t)}{2\sigma^2}\right) + Dt, \end{aligned} \quad (6)$$

where $\Delta\varphi_{ab}(t) = \varphi_{ab}(t) - \varphi_{ab}(t_{\text{transit}})$, $\varphi_{ab} = 2\pi(\mathbf{n} - \mathbf{k}) \cdot \mathbf{u}_{ab}$, t_{transit} is the time for the source transit, and c_1, c_2 are some constants. The $D\Delta t$ is a phenomenal term to capture the time

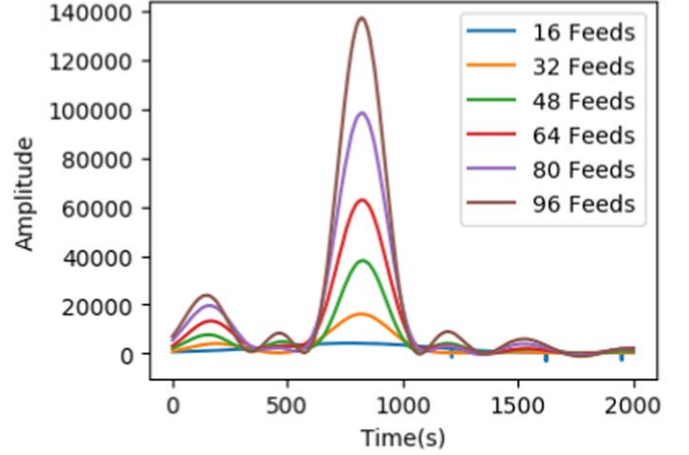


Figure 7. Synthesized beam test: beam amplitude vs. time during a Cas A transit, with each line representing a beam formed with different numbers of inputs as labeled. As we add more feeds, the amplitude of the peak increases and the beam becomes narrower.

variation of the noise bottom during the period of transit observation. The instrument phase difference $\phi_{ab} = \phi_a - \phi_b$ is then determined.

In ideal calibration scenarios, the intensity of the signal of a synthesized beam is n^2 times that of a single feed when the source is positioned at the beam's center. However, practical limitations, such as errors in the complex gain of the feeds, diminish the efficiency of beam forming as the number of feeds n increases. To assess this efficiency, we conduct a test by observing Cassiopeia A (Cas A) after calibrating the complex gains using Cas A as a reference. In this test, the number of feeds used is equal to n , with beam #1 formed using $n = 1$, i.e., only one feed, while the weight for other feeds are all set to zero; beam #2 using $n = 2$, and so forth.

The transit of Cas A for these beams is shown in Figure 7. For the first 32 beams we use the feeds from the same cylinder, while for more than 32 we also include feeds from different cylinders. We can see that with the increase of the number of feeds, the amplitude increases, showing that the beam forming produces higher sensitivity than individual feed measurements. The beams narrowed down significantly when the feeds from the different cylinders were included. There is also a peak prior to the main peak in this transit. It is at partly due to some asymmetric feature in the primary beam response, as we found a similar peak in the transit of other sources such as Cyg A, though the variation of sky intensity may also contribute.

In Figure 8 we show the efficiency of the synthesized beam, i.e., the ratio of the measured peak amplitude and theoretical amplitude, as a function of the number of feeds used. If the amplitude and phase of each feed are exactly the same, or calibrated without error, the peak amplitude would be proportional to the number of feeds being used. However, as

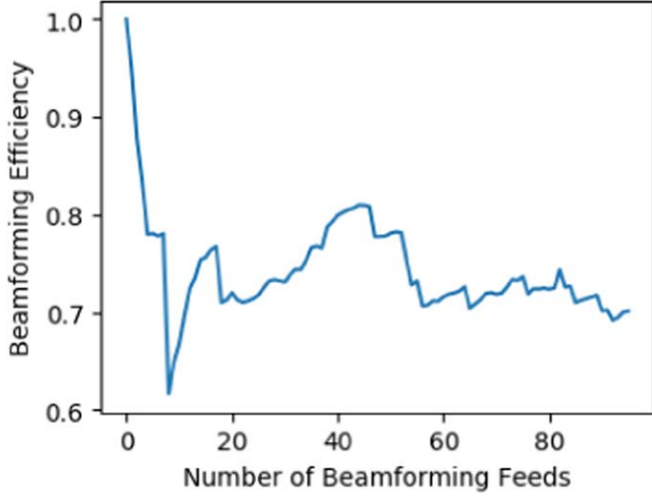


Figure 8. The efficiency of the synthesized beam, i.e., the ratio of the measured amplitude and the theoretical amplitude as a function of the number of feeds.

there are calibration errors, the actual amplitude is less than the theoretical one. In the figure we see once the additional feeds are added, the efficiency drops below 100%, but with the available feeds added in, the efficiency stabilizes at about 70%.

3.2. Sensitivity

The sensitivity of the system can be determined from a calibration observation of a known source, e.g., Cyg A. After masking the RFI, data of 424 channels are used for calibration, ranging from 689.9 to 741.6 MHz, and the same band is used in the FRB search. During its transit, we can see an increase in the received power, induced by the transit of Cyg A, obtained by subtracting the peak value and the background value,

$$\Delta_{\text{cal}} \equiv T_{\text{pk}} - T_{\text{bg}} = 236.92 \sigma_T, \quad (7)$$

where σ_T is estimated for the 689.9–741.6 MHz band (total bandwidth 51.7 MHz), a sampling time of 0.098304 ms, and a beam formed with the 96 feeds. The peak amplitude of the transit is obtained by averaging the data for 5 s, while the background is taken at about 20 minutes before the transit.

The flux of the calibrator source (Cyg A) at this frequency is $S_{\text{cal}} = 2980$ Jy, and $\Delta_{\text{cal}} \propto S_{\text{cal}} B_{\text{cal}}$, where B_{cal} is the primary beam gain in the direction of the calibrator, so the sensitivity at a direction \mathbf{n} for a single integration time is given by

$$S_{n0} = S_{\text{cal}} \frac{B_{\text{cal}}}{B_n} \frac{\sigma_T}{\Delta_{\text{cal}}} = 12.58 \left(\frac{B_{\text{cal}}}{B_n} \right) \text{Jy}, \quad (8)$$

where $B_n = B(\mathbf{n})$ is the primary beam gain in that direction.

In pulse searching, the detection limit is usually given in fluence, which depends both on the pulse flux and the duration, $F = \int S_\nu(t) dt$, which has a unit of Jy · ms. Denoting the sampling time interval as τ , and duration of the pulse Δt , if

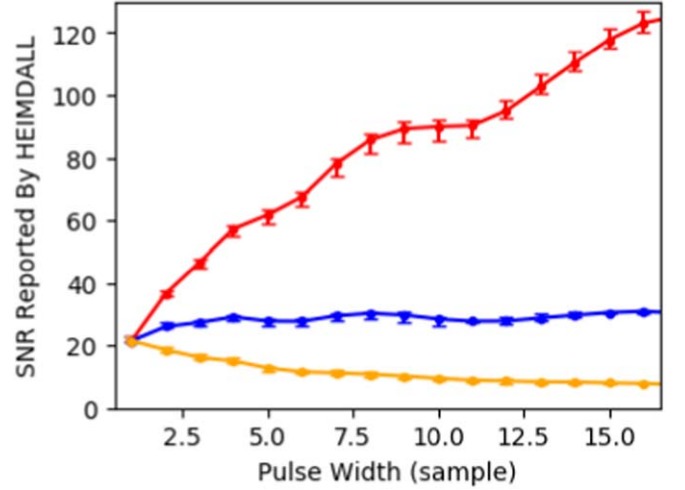


Figure 9. The SNR of mock FRBs as a function for pulse width, for a constant flux density of 280 Jy (red), a constant fluence of 280 Jy ms (orange), and a constant $F/\sqrt{\tau \Delta t}$ of 280 Jy ms^{1/2} (blue). The error bars are derived from five trials.

Δt is less than a single integration time interval τ , then $S_n(\Delta t) = S_{n0}$. For a longer interval, $S_n(\Delta t) = S_{n0} \sqrt{\tau/\Delta t}$, so

$$\text{SNR}(\Delta t) = \begin{cases} \frac{S}{S_{n0}} \frac{\Delta t}{\tau} = \frac{F}{S_{n0} \tau}, & \Delta t < \tau \\ \frac{S}{S_{n0}} \sqrt{\frac{\Delta t}{\tau}} = \frac{F}{S_{n0} \sqrt{\tau \cdot \Delta t}}, & \Delta t \geq \tau \end{cases} \quad (9)$$

In the actual case, most bursts will have $\Delta t > \tau$, and we search through the maximum SNR using a boxcar filter for different widths in time. Pulses with SNR greater than certain threshold are then selected. This is demonstrated in Figure 9, where we show the variation of SNR for different pulse width, for the case with a constant flux (280 Jy), the case with a constant fluence (280 Jy ms), and the case with flux $\times \Delta t^{1/2}$ kept constant (280 Jy ms^{1/2}). As predicted by Equation (9), the observed SNR of a burst is proportional to $F/\sqrt{\tau \Delta t}$.

3.3. Mock FRB Test of Pipeline Performance

To examine our pipeline performance, we inject mock FRB signals into real-time data, and feed these to our transient detection pipeline. The pulses we generate are box-shaped, with flux ranges from 10 to 310 Jy, pulse widths between 0.8 and 24.8 ms, corresponding to 2–62 sample times, and DM values between 70 and 1570 pc cm⁻³. The combination of these parameters produces a total of 7488 mock FRBs.

We inject these mock FRBs into real search data during blind search by HASHPIPE. We inject nine mock FRBs every 81.2 s, separated by approximately 10 s and into different beams. It takes 18.7 hr to inject all 7488 mock FRBs, which

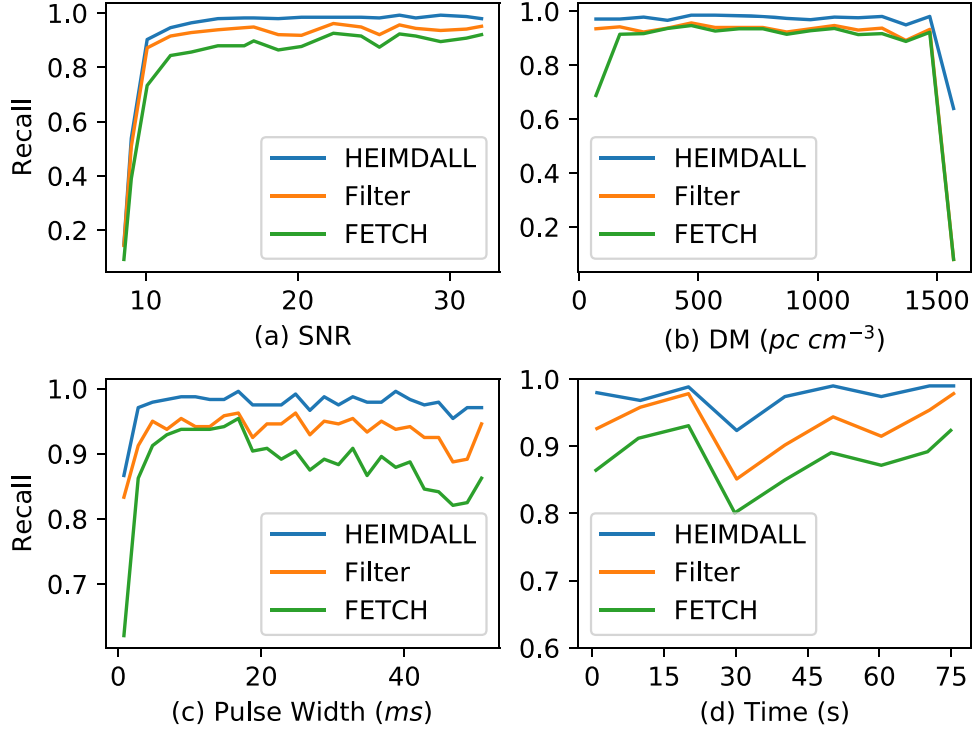


Figure 10. The recall rates of the mock FRBs recovered by our pipeline after each processing step as a function of SNR, DM, pulse width, and candidate time in each data file. For each plot the other parameters are kept at the average value.

allows us to study the influence of day and night. To obtain the performance of each step in the pipeline, we show statistics for the recovered FRBs at three stages: after the HEIMDALLs single pulse searching, after filtering, and after the FETCH classification (final stage of the pipeline). We analyze the result using recall rate and precision, given by Equations (10) and (11) respectively. Recall rate shows our pipeline’s ability to detect potential FRBs, while precision indicates its performance against RFIs.

$$\text{Recall Rate} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (10)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}. \quad (11)$$

Figure 10 illustrates how the recall rate depends on each parameter of the mock FRB, after marginalizing other parameters. The performance of our pipeline as measured by the recall rate depends on the SNR of the event. For FRBs with an SNR below 10, the recall rate is less than 0.5. However, for brighter pulses, the recall rate is above 0.8 and remains stable.

Our pipeline maintained a stable recall rate of approximately 0.9 across a wide range of DM values. Notably, the FETCH has a low performance at the two ends of DM values. In terms of pulse width, the recall rate drops at the narrow pulse end, at a pulse width of 0.8 ms. For longer burst duration, both HEIMDALL and

the filter performed normally, while FETCH exhibited a 5% decrease in performance. Figure 10(d) displays the recall rate of simulated FRBs injected at different candidate times within each data file, which shows that our pipeline consistently delivered stable performance, and we found no noticeable difference between day and night. Note here we have cut off the last 6 s of the data, where the detection efficiency drops due to the reason explained in Section 2.4.3, but this is not a problem from FETCH, and since the same segment of data will be searched at the beginning part of the next time block where such drop would not occur, this does not matter.

In our setting, HEIMDALL typically yields a few hundred to a thousand candidates within the file which contains 81.2 s of data. The majority of these spurious candidates are filtered out by the SNR threshold and lower pulse width limit of the candidate filter. The remaining candidates are classified by FETCH, which reached its maximum precision, equal to unity in this test. The overall recall and precision rates for our test are presented in Table 4, where we excluded those with an SNR below 10 and a DM exceeding 1500 for valid mock FRBs.

In operation, the typical number of candidates selected after filtering is $\sim 10^2$ per day. Instead of the deep learning algorithm FETCH, we can also perform human inspection after filtering at the cost of extra workload and delayed notification for follow-up observations.

Table 4
Overall Recall and Precision of Pipeline

Step	All Mock FRBs		Valid Mock FRBs	
	Recall	Precision	Recall	Precision
HEIMDALL	0.884	<0.01	0.975	<0.01
Filter	0.816	0.502	0.934	0.478
FETCH	0.764	1.00	0.881	1.00

3.4. Pulse-Searching Test

Bright pulsars can be used as target sources to test our pulse-searching pipeline. Table 5 lists bright pulsars, selected from Lorimer et al. (1995) and the ATNF pulsar catalog¹⁸ (Manchester et al. 2005), with mean flux density $S \gtrsim 40$ mJy at 606 MHz (or 350 MHz if it is not available at 606 MHz) in the northern sky ($\delta > 0^\circ$). The objects are listed in descending order of their mean pulse fluence F . Four of these have been detected by the Tianlai Cylinder Pathfinder Array, which are labeled with a * in Table 5. Among these, PSR B0329+54 is the only pulsar that is bright enough to be well above our detection limit and can be re-detected readily each day. PSRs B2021+51, B0823+26 and B2020+28 have also been detected by our array, probably during their bright phases (Chen et al. 2023). Some of the other pulsars listed in Table 5 may also be detectable. PSRs B1133+16 and B0950+08 have very small DM values, which will be filtered out by our pipeline. The table also gives relatively large fluence for PSRs J0341+5711 and J2238+6021, but note these are measured at 350 MHz, if scaled to higher frequency their fluence would not be that large.

We use the bright pulsar PSR B0329+54 as a target source to test and evaluate our pulse-searching pipeline. Figure 11 shows a segment of the data at the peak of transit for this source. The top panels show the raw data, the second panel shows the de-dispersed data, where signal is enhanced; the third panel shows the HEIMDALL candidates (note the higher SNR), the fourth panel shows the candidates after filtering, and finally, the bottom panel shows the candidates classified as real signals (in orange) and RFI (in blue). From these observation data, we derive an average pulse period of 0.7147 s, and a DM value of 27.31 pc cm^{-3} , consistent with those reported in the literature, 0.7145 s and 26.76 pc cm^{-3} respectively (Manchester et al. 2005). These results demonstrate that our pipeline successfully identifies most of the pulses from the pulsar as candidates, especially those with higher SNRs. However, it may miss pulses located near strong RFI sources. Also, as noted in Section 2.4.3, at the end of each data file there is a drop of detection efficiency. This end-of-file problem can be solved by copying the 5 s of data at the end of the file to the

start of the next data file, so that such a missed pulse can still be detected in the next block.

4. Beam Arrangement and Optimization

In the previous sections we considered detection of a radio pulse with a digitally formed beam. The current digital backend allows the simultaneous formation of a total of 96 beams, which can be pointed in different directions in the sky at will. In this section we study how to arrange these beams, in order to achieve the optimal detection rate for FRBs, which occur randomly in the sky, and estimate this rate. We generate a sample of mock FRBs with different fluences and widths, and then check whether they can be detected by one of the digitally formed beams. The arrangement of the beams is varied to find the optimal configuration.

4.1. FRB Distribution

To generate a sample of FRBs, we assume its cumulative fluence distribution follows (CHIME/FRB Collaboration et al. 2021)

$$D_{\text{Flu}}(>F) = R_{\text{sky}, 5 \text{ Jy ms}} \left(\frac{F}{5 \text{ Jy ms}} \right)^\alpha. \quad (12)$$

The all-sky FRB rate $R_{\text{sky}, 5 \text{ Jy ms}} = 525$ per day at the relevant frequency band, and $\alpha = -1.4$. We grid the fluence distribution from 0.25 to 2000.25 Jy ms evenly into 4000 intervals, with $\Delta F = 0.5 \text{ Jy ms}$. The FRB all-sky rate within an interval is then given by

$$\Delta D_{\text{Flu}}(F) = D_{\text{Flu}}(F) - D_{\text{Flu}}(F + \Delta F). \quad (13)$$

For the pulse width we fit a log-normal distribution for the sample from the Blinkverse FRB catalog (Xu et al. 2023), as Figure 12 (left) shows. In that catalog, we exclude the CHIME and FAST FRBs, as the CHIME FRBs correspond to only a few values (Figure 12 right), while the FAST FRBs constitute thousands of bursts from only a few active repeaters, which may not be very representative of our sample (Figure 12 middle). From Figure 12, we can see that the FAST and CHIME samples also follow a normal distribution with similar μ_w , thus excluding them will not cause a major impact on our results. The probability density function of the fitted distribution is given by

$$D_w(W) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left(-\frac{(W - \mu_w)^2}{2\sigma_w^2}\right), \quad (14)$$

where $W = \log_{10}(\Delta t / \text{ms})$, $\mu_w = 0.5386$ and $\sigma_w = 0.3732$. We then grid W from -1.5 to 2.5 into 1000 intervals, with a $\Delta W = 0.004$.

¹⁸ <http://www.atnf.csiro.au/research/pulsar/psrcat>

Table 5
List of Bright Pulsars in the Northern Sky, Ranked by Fluence F

Name	DM (pc cm ⁻³)	S (mJy)	Period (s)	W_{50} (ms)	S_{pulse} (Jy)	F (Jy ms)	F_{th} (Jy ms)
B0329+54*	26.7	1337	0.714	6.6	145	957	100
J0341+5711	101	364.7†	1.89	43	16.0	689	256
J2238+6021	185	111†	3.07	25	13.6	341	195
B1133+16	4.84	144	1.19	5.8	29.5	171	94.1
B2319+60	94.6	49.5	2.26	131	0.85	112	430
B0950+08	2.97	402	0.253	8.6	11.8	102	115
B1919+21	12.4	72.7	1.34	34	2.87	97.6	228
J0302+2252	19.0	70.0†	1.21	50	1.69	84.7	276
B2154+40	71.1	54.7	1.53	39	2.15	83.7	244
B0809+74	5.75	51.0	1.29	41	1.60	65.8	250
B1237+25	9.25	46.8	1.38	45	1.44	64.6	262
B1933+16	159	161	0.359	6.3	9.17	57.8	98.1
B1946+35	129	78.9	0.717	19	2.98	56.6	170
B2021+51*	22.5	95.9	0.529	7.4	6.86	50.8	106
B1929+10	3.18	187	0.227	6.0	7.07	42.4	95.7
J2043+7045	57.6	63.7†	0.588	37	1.01	37.5	238
B0823+26*	19.5	62.4	0.531	5.8	5.71	33.1	94.1
B1859+03	402	47.4	0.655	9.0	3.45	31.0	117
B2020+28*	24.6	59.9	0.343	12	1.71	20.5	135
B2310+42	17.3	46.0	0.349	8.8	1.82	16.1	116
B0531+21	56.7	210	0.0334	3.0	2.34	7.01	67.7
J1022+1001	10.3	75.0†	0.0165	0.97	1.28	1.24	38.5
J0218+4232	61.3	47.0†	0.00232	1.0	0.109	0.109	39.1

Note. The mean flux density S values are mostly measured at 606 MHz, but those labeled with a † are measured at a lower frequency, mostly at 350 MHz. W_{50} is the pulse width at 50% of peak. S_{pulse} is the mean flux density for each single burst, obtained using S and the duty cycle W_{50}/Period . F is the corresponding mean fluence of single burst, and F_{th} is the 10σ fluence detection limit at the primary beam center of Tianlai, calculated using W_{50} and Equation (20). Pulsars labeled with * were detected by the Tianlai Cylinder Pathfinder Array.

4.2. Beam Arrangement

Although a digitally formed beam can be pointed at any direction in the sky, obviously it should be placed within the main lobe of the primary beam to be effective. As the primary beam of our cylinder antenna is a long strip in the N–S direction, we place the digital beams in three parallel rows across the East–West (E–W) direction, and on each row 32 beams are lined along the N–S direction, with constant E–W and N–S spacings, as shown in Figure 13.

Following Shaw et al. (2015) and Yu et al. (2023, 2024), we model the voltage response as a product of the N–S distribution and the E–W distribution,

$$A(\hat{n}) = A_{\text{NS}}(\sin^{-1}(\hat{n} \cdot \hat{x}); \theta_{\text{NS}}) \times A_{\text{EW}}(\sin^{-1}(\hat{n} \cdot \hat{y}); \theta_{\text{EW}}) \quad (15)$$

where \hat{x} and \hat{y} are the unit vectors pointing to East and North, respectively, and the N–S and E–W functions are

$$A_{\text{NS}} = \exp \left[-4 \ln 2 \left(\frac{\theta}{\theta_{\text{NS}}} \right)^2 \right], \quad (16)$$

where $\theta_{\text{NS}}(\nu) = \alpha_{\text{NS}} \frac{\lambda}{D_{\text{NS}}}$, with $D_{\text{NS}} = 0.3$ m;

$$A_{\text{EW}} \propto \int_{-\frac{W}{2}}^{\frac{W}{2}} A_D \left(2 \tan^{-1} \left(\frac{x}{2FW} \right); \theta_{\text{EW}} \right) e^{-i \frac{2\pi}{\lambda} x \sin \theta} dx, \quad (17)$$

where

$$A_D(\theta; \theta_{\text{FWHM}}) = \exp \left[-\frac{\ln 2}{2} \frac{\tan^2 \theta}{\tan^2(\theta_{\text{FWHM}}/2)} \right], \quad (18)$$

$W = 15$ m, and fitting the parameters α_{NS} , F_R and θ_{EW} to the results of electromagnetic field simulation (Sun et al. 2022) yields $\alpha_{\text{NS}} = 1.04$, $F_R = 0.2$ and $\theta_{\text{EW}} = 2.74$.

The effective digital beam will be a product of the primary beam and the array form factor. For the cylinder array, both the primary beam and the digitally formed beam are elongated along the N–S direction, as demonstrated in Figure 14.

Each digital beam has a $\sim 0.5^\circ$ width in the E–W direction, so the three rows of the digital beam could cover about 1.5° if they do not overlap each other, while the primary beam has a width of $\sim 2^\circ$. As confirmed in Figure 14, the digital beams have multiple sidelobes along the E–W direction. This is unavoidable as the baselines between feeds on different cylinders are longer than the half-wavelength. These lobes are however

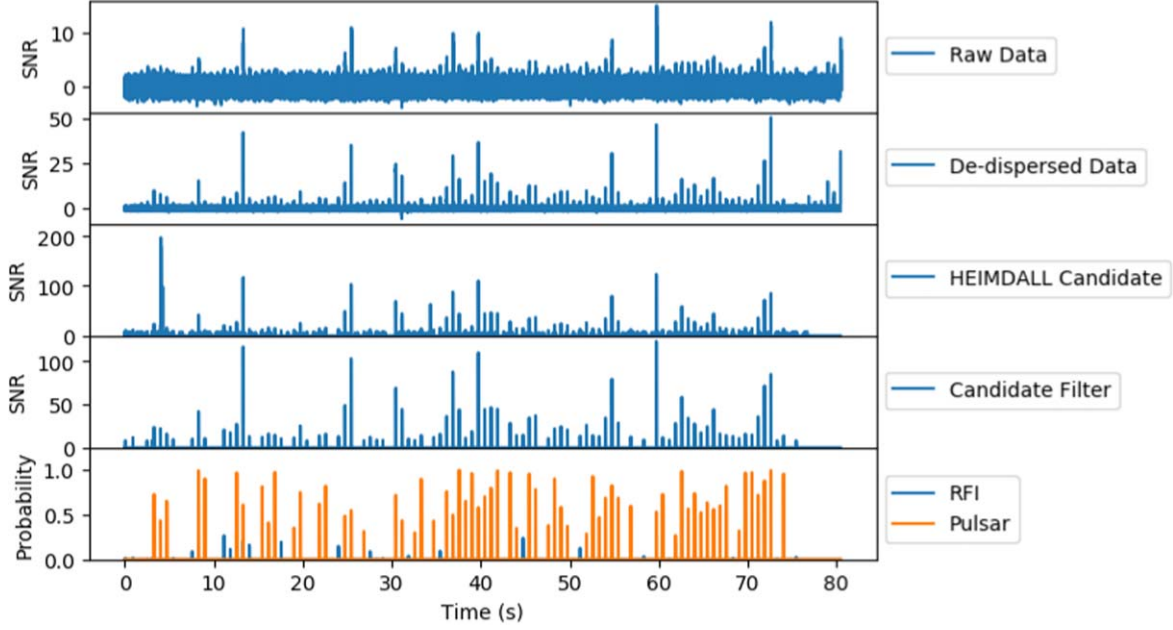


Figure 11. Pipeline tested with PSR B0329+54. Top panel: the raw time series without de-dispersion. The second panel: the time series data de-dispersed with DM 27.31 pc cm^{-3} . The third panel: the candidates reported by HEIMDALL. The fourth panel: the candidates remaining after filtering. Bottom panel: the mean probability of each candidate being an astronomical signal provided by the three FETCH models.

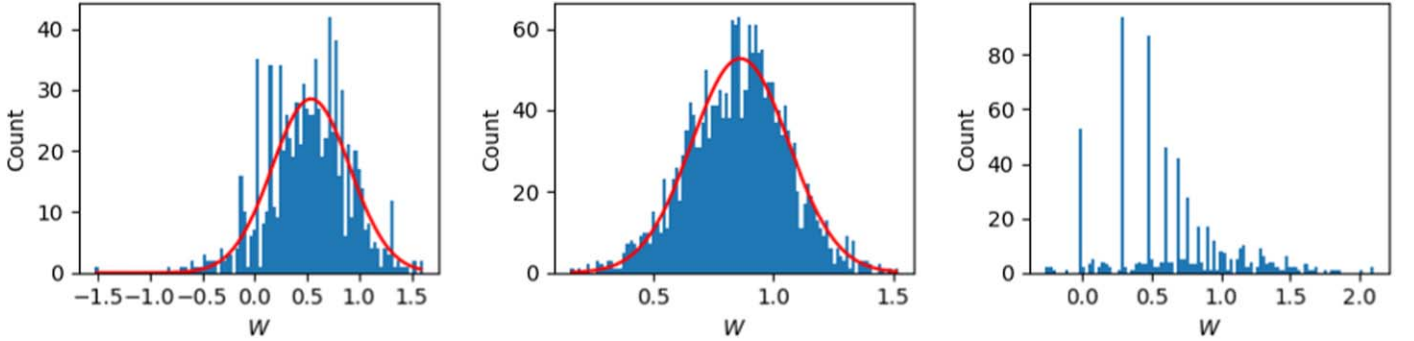


Figure 12. Left: Histogram of the adopted FRB width in the Blinkverse FRB catalog with normal distribution fitting traced in red line. Middle: Histogram of the FAST sample with normal distribution fitting marked in red line. The FAST sample includes mainly thousands of bursts from only a few active repeaters, which may not be representative of the overall distribution. The FAST sample has larger $\mu_w = 0.8626$ and smaller $\sigma_w = 0.2020$. Right: Histogram of excluded CHIME samples. CHIME samples are excluded as the majority of them are only several values.

modulated by the primary beam, so that they are attenuated as we move off the center of the primary beam. The presence of these lobes could cause some ambiguity in the localization of a burst, but this ambiguity could be partly broken by jointly using the signal strength of multiple beams. To reduce the gap between beams, we shift the beams in the center row by half of the N–S spacing toward the South, so that the tip of the beams in the two adjacent rows fits nicely into the spacing between the beams in the central row, forming a closely packed structure.

The detection rate function is constructed as follows. First we grid the sky evenly in a topocentric Cartesian coordinate system. The N–S component of the unit vector, from -0.9 to

0.9 , is gridded into 501 pixels, corresponding to altitude angle greater than 25.8° . The E–W component of the unit vector, from -0.15 to 0.15 , is gridded into 401 pixels, corresponding to the E–W offset within $\pm 8.6^\circ$. We obtain the beam response $B_{ij,k}$ of the k th beam at pixel (i,j) by multiplying the synthesized beam profile centered at its given direction with the primary beam profile. The fluence threshold is estimated as

$$F_{\text{th}} = N_{\text{th}} S_n \Delta t = N_{\text{th}} S_{\text{cal}} \sqrt{\tau \Delta t} \frac{B_{\text{cal}}}{B_n} \frac{\sigma_T}{\Delta_{\text{cal}}}. \quad (19)$$

For a given point in the sky, one of the digitally formed beams will have the most sensitive response. This digital beam

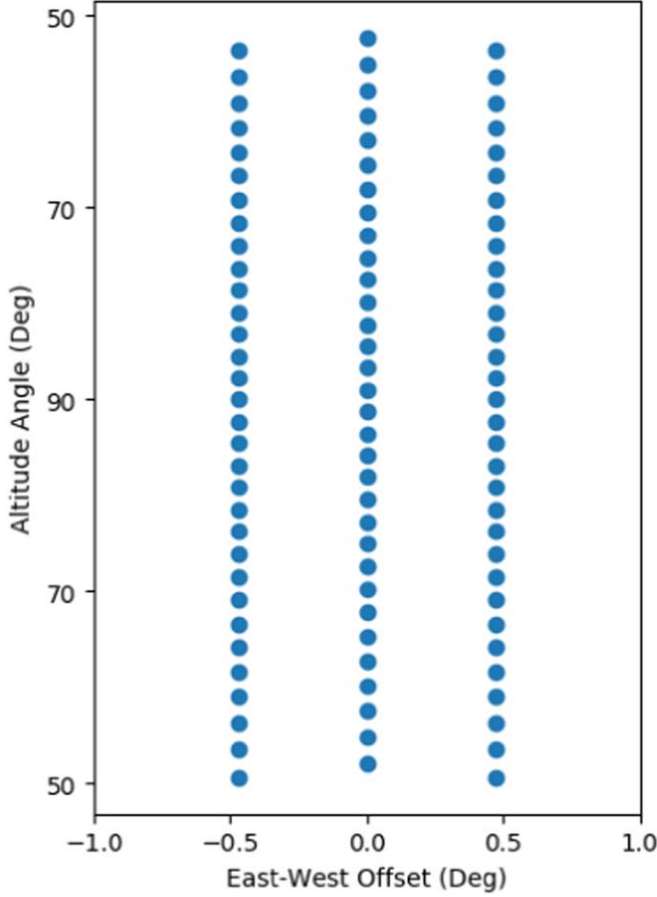


Figure 13. The adopted beam arrangement, each point marks the center of a digital beam. The plot here shows the optimal configuration, with a spacing of 2.27° in N-S and 0.47° in E-W.

can be identified by finding the maximum of $B_k(\mathbf{n})$, though on some exceptional points (boundary) more than one beam may have equal sensitivity. In more sophisticated algorithms, the data from more than one beam may be combined to enhance the sensitivity, but this will increase the complexity of the algorithm, and in most cases the actual enhancement in sensitivity is slight, though the localization of the burst can be improved by the joint fit of multiple beams. Below for the sensitivity estimate we will use the one best digital beam and its corresponding threshold, with SNR threshold N_{th} set to 10 to estimate our detection rate. We then have

$$F_{th} = 12.58 N_{th} \sqrt{\tau \Delta t} \left(\frac{B_{cal}}{\max_k (B_{ij,k})} \right) \text{ Jy ms.} \quad (20)$$

The FRB detection rate of a pixel is then estimated by

$$R_{ij} = \frac{A_{ij}}{(360^2/\pi)} \sum_{F, \Delta t}^{F > F_{th}(\Delta t)} \Delta D_{Flu}(F) D_w(\Delta t), \quad (21)$$

where i, j denotes the pixel coordinate, A_{ij} is the size of a pixel in the unit of square degrees, and the whole sky area is $360^2/\pi \text{ deg}^2$. R_{ij} is plotted in Figure 15. The overall FRB detection rate is obtained by summing over the contribution from all pixels.

In Figure 15, the left panel shows an arrangement with the center of the digital beams at the zenith, which best utilized the most sensitive part of the primary beam. In the right panel, we display any arrangement where the center is shifted 10° northward, so that the north celestial pole (NCP) which has an altitude angle of 44° is better covered. However, due to the limited number of beams, the southern part of the sky is then less well covered, and because for this cylinder array the sky area with lower altitude generally has lower sensitivities, the overall detection rate will be reduced.

To optimize our FRB detection rate, we adjust the pointing for the digitally formed beams, trying different separations between the rows of beams, and separations between beam centers between each column. The optimization is achieved by maximizing a function that takes these parameters as input and returns the estimated FRB detection rate. We search for the maximum of the detection rate function by varying the beam separation in the range of (0, 14) and (0, 15) pixels for the beam separation between and within columns, respectively, with a starting value of 2 for both parameters.

In Figure 16, we show how the detection rate varies with altitude of the center beam, the N-S beam separation, and the E-W direction beam separation. These parameters are varied one at a time, with the other two parameters fixed to the optimal value. As we can see from the left panel of Figure 16, the highest detection rate is achieved when the center of the array of the digitally formed beams is aimed at the zenith, which is the peak of the primary beam of the cylinder, thus having the highest sensitivity. If we shift the center away from the zenith, the total detection rate would be lowered.

Varying the separation of the digitally formed beams in either the N-S or E-W directions has two effects: on the one hand, increasing the beam separation reduces the beam overlap, and allows the digitally formed beams to cover a larger total sky area, which increases the rate of FRB detection; on the other hand, this also means that some of the digitally formed beams are placed off-center, leaving holes in the main lobe of the primary beam where the sensitivity is the highest. The optimal detection rate is achieved at a point where these two trends are balanced, and most of the crest of the primary beam is covered. We find that for the optimal arrangement, the beam separation is 0.47° in the E-W direction and 2.27° in the N-S direction near the center of FoV. At this optimal configuration, the detection rate is 0.571 FRBs per month for a 10σ burst.

In Figure 17, we show the distribution of the fluence and width of the detected bursts. There is a clear lower limit in the fluence of the detected bursts, which increases with increasing width of the burst, as we already noted in the discussion on the

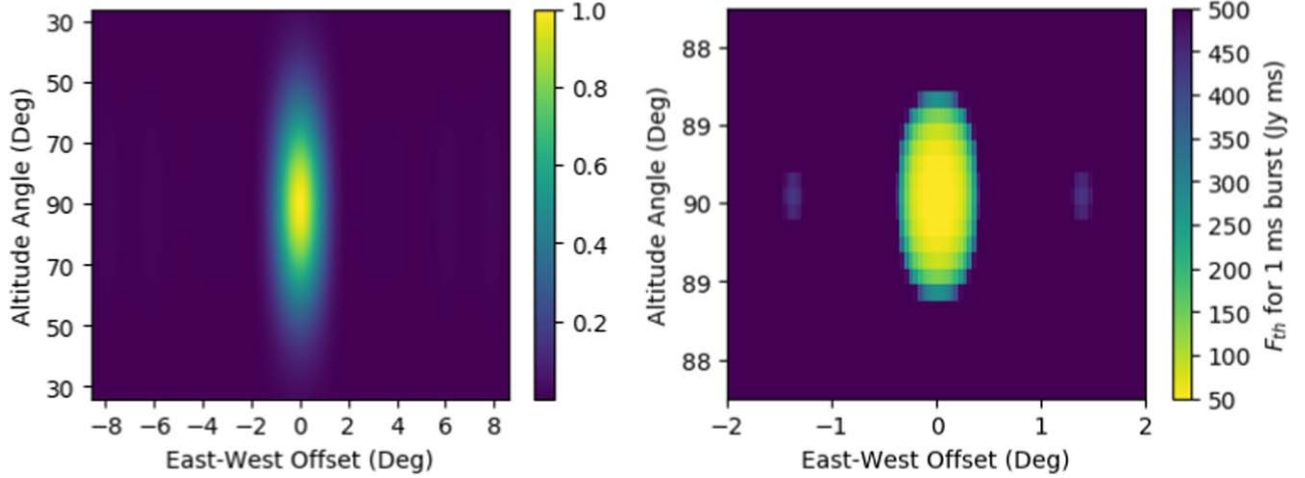


Figure 14. Left: The adopted primary beam profile for the cylinder array. Right: The profile of an effective digital beam.

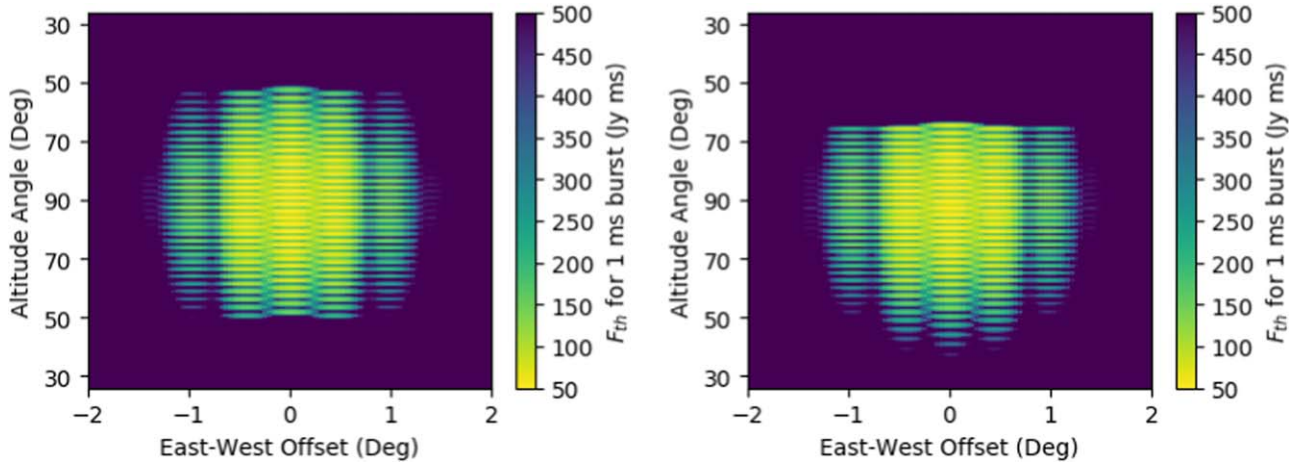


Figure 15. The overall beam profile R_{ij} , given in terms of the detection threshold F_{th} . Left: the case where the center is at the zenith; Right: the case where the center is shifted 10° to North has more coverage on the NCP.

sensitivity in Section 3.2. A high concentration of the detected bursts will have a short burst width of a few ms, with fluence less than 100 Jy ms, but of course there will also be many detectable bursts. This result is somewhat dependent on the assumed model of FRB fluence and duration distribution, but it shows what kind of FRBs we expect to detect.

5. An FRB Detection

During test observation, we detected a bright FRB, designated as FRB 20220414A, at UT 17:26:40.368 on 2022 April 14. In this observation, we did not use the optimal configuration of beam arrangement discussed above, instead the 96 beams were arranged in a single row along the meridian, with $\text{Alt} \geq 60^\circ$. The separation between beams was $\approx 0.6316^\circ$. The FRB was detected in three neighboring beams (#40, #41, #42) with SNR = 11.9, 17.7 and 10.9, respectively. In Figure 18, we display the raw FT waterfall plot of beam 41. The band at 770–780 MHz is

contaminated by an RFI (probably from the newly allocated 5G mobile signal) and is removed. In Figure 19 we plot the de-dispersion result for the three beams where it is detected, as well as two neighboring beams where it is undetected. The detection was reported on the Astronomer’s Telegram (Yu et al. 2022a), and the Transient Name Server (TNS).¹⁹

For a burst in the direction \mathbf{q} with fluence F , the observed data of synthesized beam α are

$$F_\alpha = FB_\alpha(\mathbf{q}) + \eta_\alpha,$$

where η_α is the noise, $FB_\alpha(\mathbf{q})$ is the expected signal, and the response of the synthesized beam $B_\alpha(\mathbf{q})$ is given by

$$B_\alpha(\mathbf{q}) = \left| A(\mathbf{q}) \sum_a \exp(-2\pi j(\mathbf{q} - \mathbf{k}_a) \cdot \mathbf{u}_a) \right|^2, \quad (22)$$

¹⁹ <https://www.wis-tns.org/object/20220414a>

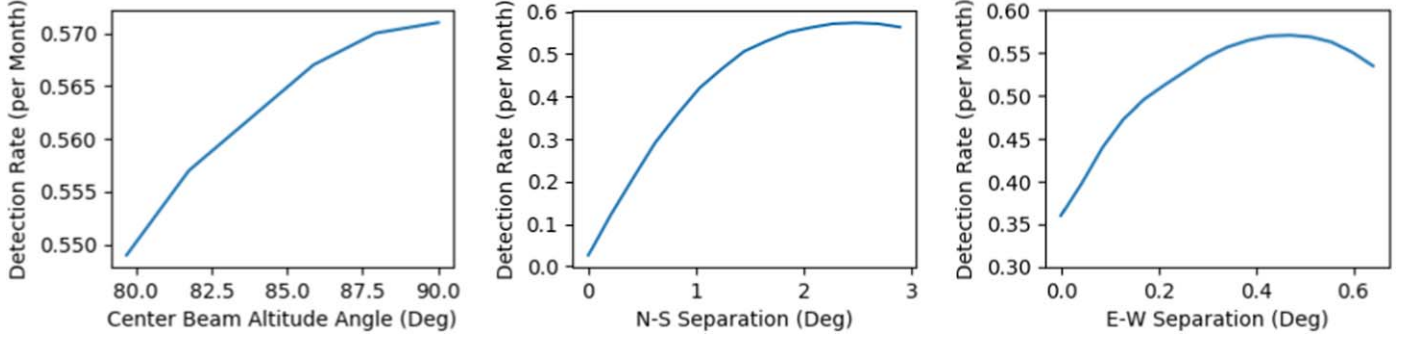


Figure 16. Detection rate with respect to the center beam's altitude angle (left) and beam separation in the N-S (middle) and E-W (right) directions. The optimal values are 90° , 2.27° and 0.47° , respectively.

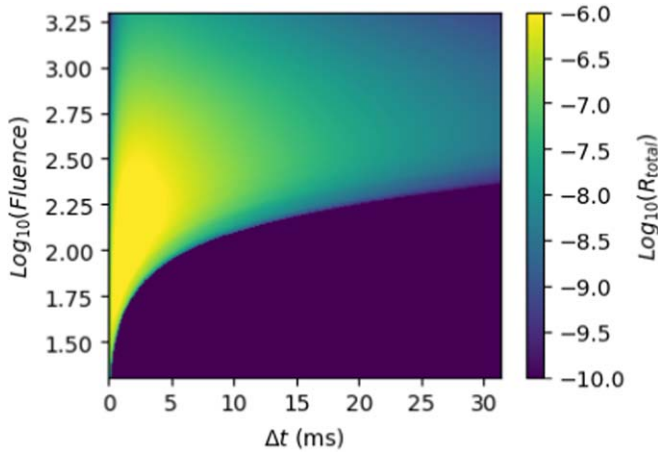


Figure 17. Estimation of the fluence distribution of observed FRBs.

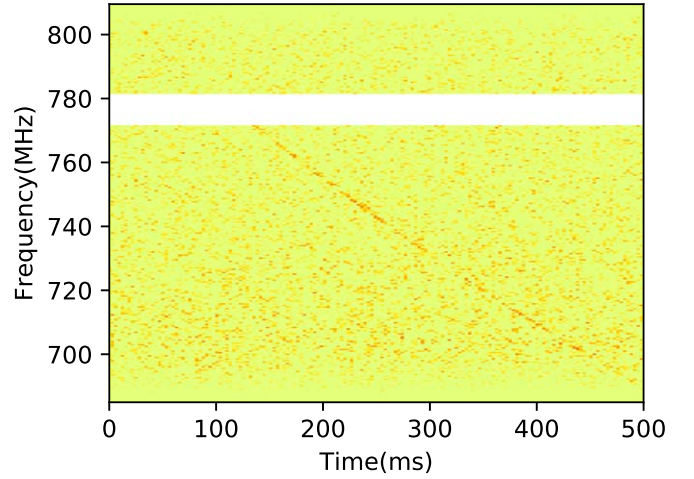


Figure 18. The raw waterfall plot of beam 41 for FRB 20220414A.

where \mathbf{k}_α is the unit vector for the direction of the synthesized beam α , and $A(\mathbf{q})$ is the voltage response of the primary beam.

The burst can be located by minimizing the likelihood. We adopt a reduced log likelihood,

$$\begin{aligned} \ln \mathcal{L}(\mathbf{q}, F_s) &= -\sum_{\alpha} \frac{(F_s B_{\alpha}(\mathbf{q}) - F_{\alpha})^2}{2\sigma_T^2 \Delta t_s / \tau}, \\ &= -\frac{1}{2} \sum_{\alpha} \left(\frac{F_s B_{\alpha}(\mathbf{q})}{\sigma_T \sqrt{\Delta t_s / \tau}} - \text{SNR}_{\alpha} \right)^2, \end{aligned} \quad (23)$$

where F_{α} is the flux detected by the beam α , F_s is the model fluence of the FRB, and Δt_s is the observed burst width. The second line is written in terms of the SNR, SNR_{α} for beam α . Only beams with significant SNR would contribute significantly to this likelihood. In principle, Equation (22) is frequency dependent. However, the main lobe does not vary significantly with frequency, therefore we average $B_{\alpha}(\mathbf{q})$ over 700–800 MHz and use the averaged synthesized beam for likelihood.

As mentioned in Section 4, there are multiple sidelobes in the E–W direction. In the present case, the beams are aligned

only in the N–S direction, so the location is even more uncertain in the E–W direction. However, due to the primary beam of the antenna in the E–W direction, if the FRB occurs in a sidelobe, it would be very bright in order to be detected, and in that case it might also be seen in other beams. The fact that it is not detected by the other beams itself places some constraint on its location in the E–W direction.

The localization may be enhanced if we adopt a prior on the fluence F_s using the fluence distribution in Equation (13),

$$\ln L_{\text{Flu}}(F_s) = \ln \Delta D_{\text{Flu}}(F_s), \quad (24)$$

which helps to improve our localization in the E–W direction. The probability of the location of the burst is depicted in Figure 20, where the red, orange and blue circles show the -3 dB range from the peaks of beams #40, #41, and #42 respectively. The pixels with probability less than the minimum of the color bar, i.e., 10^{-6} , are displayed in white. We can see that the most likely positions are at the overlapping regions of the three beams, though there are also sidelobe areas to the East and West of the main lobe region. A zoom-in plot of the main lobe is shown in Figure 21, where the ellipse represents the

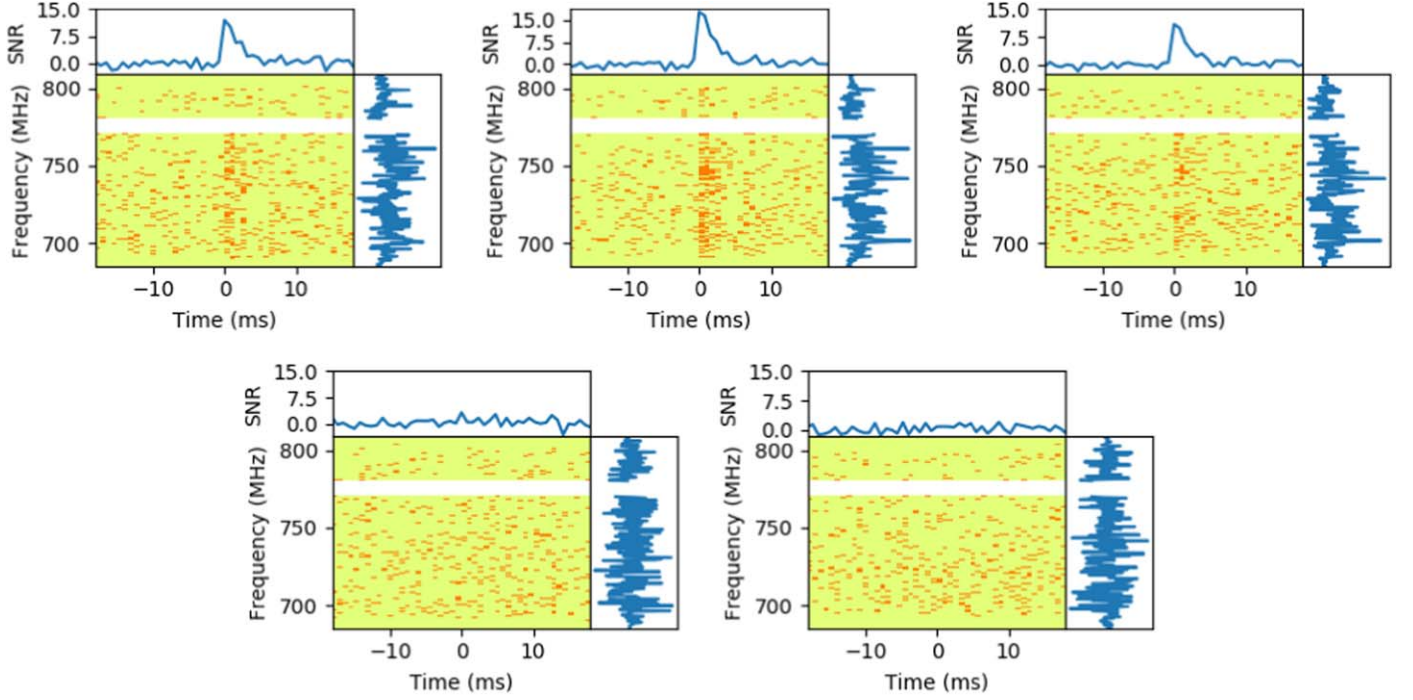


Figure 19. The de-dispersion plot of FRB 2220414A. Top: the three beams 40 (top left) 41 (top middle), and 42 (top right); Bottom: the two neighboring beams where it is not detected: 39 (bottom left) and 43 (bottom right).

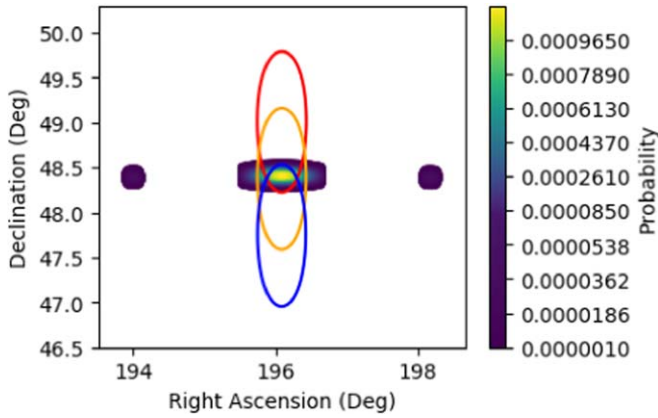


Figure 20. The FRB 2220414A localization probability. The red, orange and blue circles label the -3 dB gain range of beam 40, 41 and 42, respectively, and the region with probability less than the minimum of the color bar is shown as white.

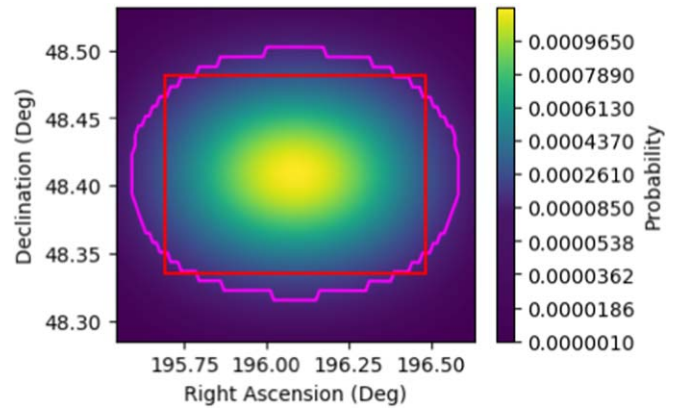


Figure 21. The main lobe of Figure 20. The magenta ellipsoid indicates the 90% cumulative probability region. The red rectangle indicates the 90% cumulative probability region obtained from the 1-d cumulative probability of R.A. and decl., in the respective directions.

region of 90% probability, while the rectangle marks the 90% cumulative probability region obtained from the 1-d R.A. and decl. probability. With the information from multiple beam detections, the burst is localized with a precision better than the synthesized beam size.

The properties of this FRB are listed in Table 6. The best fit total DM is 207.6 pc cm^{-3} , slightly lower than we originally reported in Astronomer's Telegram. We can infer the redshift

of the burst using the DM-redshift relation (Zhang 2018),

$$z_{\text{max}} \sim \frac{\text{DM}_E}{855 \text{ pc cm}^{-3}} (z_{\text{max}} < 3). \quad (25)$$

The total DM is given by

$$\text{DM}_{\text{total}} = \text{DM}_{\text{MW}} + \text{DM}_{\text{halo}} + \text{DM}_{\text{IGM}} + \frac{\text{DM}_{\text{host}}}{1+z}, \quad (26)$$

Table 6
Properties of FRB 20220414A

R.A.	13 ^h 04 ^m 21 ^s ($\pm 1^m 35^s$)
Decl.	+48°24′32″ ($\pm 4′ 25″$)
DM	207.6 pc cm ⁻³
Galactic DM (YMW16)	21.1 pc cm ⁻³
Galactic DM (NE2001)	27.5 pc cm ⁻³
DM-inferred redshift	≤ 0.24
Peak flux density	128.4 Jy
Duration	2.2 ms
Fluence	204.1 Jy ms
Estimated redshift	<0.16
Luminosity	$\leq 5.76 \times 10^{43}$ erg s ⁻¹

Table 7Optical Galaxies in SDSS DR17 Spectroscopic Catalog with R.A. and Decl.
Inside the 90% Confidence Interval and Redshift Lower than 0.24

Object	R.A.	Decl.	Redshift
SDSS J130556.97+482013.6	13 ^h 05 ^m 56 ^s .97	+48°20′13″.6	0.141
SDSS J130524.90+482110.2	13 ^h 05 ^m 24 ^s .90	+48°21′10″.2	0.181
SDSS J130535.68+482000.9	13 ^h 05 ^m 35 ^s .68	+48°20′00″.9	0.154
SDSS J130417.23+482651.7	13 ^h 04 ^m 17 ^s .23	+48°26′51″.7	0.182
SDSS J130449.78+482740.1	13 ^h 04 ^m 49 ^s .78	+48°27′40″.1	0.181
SDSS J130327.90+482302.0	13 ^h 03 ^m 27 ^s .90	+48°23′02″.0	0.135
SDSS J130342.72+482207.2	13 ^h 03 ^m 42 ^s .72	+48°22′07″.2	0.152
SDSS J130321.31+481950.9	13 ^h 03 ^m 21 ^s .31	+48°19′50″.9	0.008

where DM_{MW} is the galactic contribution, which can be estimated using the NE2001 and YMW16 models (Cordes & Lazio 2002; Yao et al. 2017; Price et al. 2021).

Assuming $DM_{MW} = 21.1$ pc cm⁻³, and neglecting the unknown contribution from the halo and the host, we derive an upper limit for the DM-inferred redshift $z_{\max} \leq 0.24$. On the other hand, if we adopt a galactic halo contribution (Dolag et al. 2015; Prochaska & Zheng 2019), $DM_{\text{halo}} \sim 50$ pc cm⁻³, the extragalactic DM would then be estimated as

$$DM_E = DM_{IGM} + \frac{DM_{\text{host}}}{1+z} = 136.5 \text{ pc cm}^{-3}, \quad (27)$$

leading to an upper value for the redshift $z_{\max} = 0.16$. Adopting the Planck Collaboration et al. (2016) model, the distance is then

$$D_L \leq \frac{cz_{\max}}{H_0} = 708 \text{ Mpc},$$

with luminosity

$$L_p \leq 4\pi D_L^2 S_{\text{peak}} \nu_c = 5.76 \times 10^{43} \text{ erg s}^{-1}.$$

Despite the large fluence and peak flux recorded, its luminosity is estimated to be at an average level for one-off FRBs due to its relatively small DM. A 2 hr follow-up observation was conducted by FAST on April 22, but no additional burst was detected.

We also searched for a possible host galaxy in the SDSS DR17 spectroscopic catalog.²⁰ The R.A. and decl. are constrained to be within the 90% confidence interval given in Table 6 and the redshift is lower than 0.24. We found eight optical galaxies and the information is listed in Table 7.

6. Conclusion

This paper presents the design and implementation of the FRB-search backend for the Tianlai Cylinder Pathfinder Array, as well as its performance obtained from simulations and test observation campaign. We described the beam forming principle, the hardware setup, as well as the pipeline for incoherent de-dispersion and candidate pulse search and classification. The present backend is capable of generating a total of 96 digital beams, covering approximately 40 square degrees of the sky (3 dB below the peak). This is smaller than the area of the primary beam, and a larger area could be covered if the computing power is increased to process a larger number of digitally formed beams. We have estimated the fluence detection threshold which varies with the pulse width, and depends also on its position. The performance of the backend has been estimated through mock FRB injection and validated through observation of pulsars.

We have optimized the digital beam placement by maximizing the FRB detection rate, which has been estimated to be around ~ 0.5 FRB event per month with the current 96 beam setup. However, this rate estimation and the optimized configuration are subject to the primary beam shape and FRB parameter distribution, both of which are poorly known at present. Nevertheless, it provides some approximate estimate that is useful for planning our experiment.

After conducting a four month FRB search campaign we realized our first discovery, namely FRB 20220414A, which we were able to localize with a precision of $\sim 0.1^\circ$. We are currently working on adding new outrigger cylinders, which would improve significantly the array localization capability.


Acknowledgments

The Tianlai Cylinder Pathfinder Array Transient Digital Backend was built with funding from the Repair and Procurement Grant of the Chinese Academy of Science. For this research we acknowledge the support of the National SKA program of China (Nos. 2022SKA0110100 and 2022SKA0110101), and the National Natural Science Foundation of China (NSFC, Grant Nos. 1236114814, 12203061, 12273070, and 12303004).

ORCID iDs

Zijie Yu  <https://orcid.org/0000-0001-8730-8726>

Furen Deng  <https://orcid.org/0000-0001-8075-0909>

Wei-Yang Wang  <https://orcid.org/0000-0001-9036-8543>

Xuelei Chen  <https://orcid.org/0000-0001-6475-8863>

²⁰ <https://skyserver.sdss.org/dr17/SearchTools/SQS>

References

- Agarwal, D., Aggarwal, K., Burke-Spolaor, S., Lorimer, D. R., & Garver-Daniels, N. 2020, *MNRAS*, **497**, 1661
- Barsdell, B. R., Bailes, M., Barnes, D. G., & Fluke, C. J. 2012, *MNRAS*, **422**, 379
- Bassa, C. G., Tendulkar, S. P., Adams, E. A. K., et al. 2017, *ApJL*, **843**, L8
- Caleb, M., & Keane, E. 2021, *Univ*, **7**, 453
- Chatterjee, S., Law, C. J., Wharton, R. S., et al. 2017, *Natur*, **541**, 58
- Chen, J. L., Wen, Z. G., Duan, X. F., et al. 2023, *ApJ*, **946**, 2
- Chen, X. 2012, *IJMP*, **12**, 256
- CHIME/FRB Collaboration, Amiri, M., Andersen, B. C., et al. 2021, *ApJS*, **257**, 59
- Cordes, J. M., & Lazio, T. J. W. 2002, arXiv:astro-ph/0207156
- Dolag, K., Gaensler, B. M., Beck, A. M., & Beck, M. C. 2015, *MNRAS*, **451**, 4277
- Li, J.-X., Wu, F.-Q., Sun, S.-J., et al. 2021, *RAA*, **21**, 059
- Li, J., Shifan, Z., & Fengquan, W. 2020, *SCPMA*, **63**, 129862
- Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, *Sci*, **318**, 777
- Lorimer, D. R., Yates, J. A., Lyne, A. G., & Gould, D. M. 1995, *MNRAS*, **273**, 411
- Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, *AJ*, **129**, 1993
- Marcote, B., Paragi, Z., Hessels, J. W. T., et al. 2017, *ApJL*, **834**, L8
- Niu, C.-H., Wang, Q.-X., MacMahon, D., et al. 2019, *RAA*, **19**, 102
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Price, D. C., Flynn, C., & Deller, A. 2021, *PASA*, **38**, e038
- Prochaska, J. X., & Zheng, Y. 2019, *MNRAS*, **485**, 648
- Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U. L. 2015, *PhRvD*, **91**, 083514
- Sun, S., Li, J., Wu, F., et al. 2022, *RAA*, **22**, 065020
- Thornton, D., Stappers, B., Bailes, M., et al. 2013, *Sci*, **341**, 53
- Wu, F., Li, J., Zuo, S., et al. 2021, *MNRAS*, **506**, 3455
- Xu, J., Feng, Y., Li, D., et al. 2023, *Univ*, **9**, 330
- Xu, Y., Wang, X., & Chen, X. 2014, *ApJ*, **798**, 40
- Yao, J. M., Manchester, R. N., & Wang, N. 2017, *ApJ*, **835**, 29
- Yu, K., Wu, F., Zuo, S., et al. 2023, *RAA*, **23**, 105008
- Yu, K., Zuo, S., Wu, F., Wang, Y., & Chen, X. 2024, *RAA*, **24**, 025002
- Yu, Z., Deng, F., Niu, C., et al. 2022a, *ATel*, **15342**, 1
- Yu, Z., Deng, F., Sun, S., et al. 2022b, *RAA*, **22**, 125007
- Zhang, B. 2018, *ApJL*, **867**, L21
- Zhang, J., Zuo, S.-F., Ansari, R., et al. 2016, *RAA*, **16**, 158
- Zuo, S., & Chen, X. 2020, *MNRAS*, **494**, 1994
- Zuo, S., Li, J., Li, Y., et al. 2021, *A&C*, **34**, 100439
- Zuo, S., Pen, U.-L., Wu, F., et al. 2019, *AJ*, **157**, 34