Publications of the Astronomical Society of Australia (2021), **38**, e017, 11 pages
doi:10.1017/pasa.2021.12

CAMBRIDGE
UNIVERSITY PRESS

Research Paper

pinta: The uGMRT data processing pipeline for the Indian Pulsar Timing Array

Abhimanyu Susobhanan[1], Yogesh Maan[2], Bhal Chandra Joshi[3], T. Prabu[4], Shantanu Desai[5], K. Nobleson[6], Sai Chaitanya Susarla[7], Raghav Girgaonkar[5], Lankeswar Dey[1], Neelam Dhanda Batra[8], Yashwant Gupta[3], A. Gopakumar[1], Manjari Bagchi[9,10], Avishek Basu[3,11], Suryarao Bethapudi[12], Arpita Choudhary[9], Kishalay De [13], M. A. Krishnakumar[14], P. K. Manoharan[15], Arun Kumar Naidu[16], Dhruv Pathak[9,10], Jaikhomba Singha[17], Mayuresh P. Surnis[11]

[1]Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Dr. Homi Bhabha Road, Mumbai, Maharashtra 400005, India, [2]ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, Dwingeloo 7990 AA, The Netherlands, [3]National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Ganeshkhind, Pune, Maharashtra 411007, India, [4]Raman Research Institute, Bengaluru 560080, Karnataka, India, [5]Department of Physics, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India, [6]Department of Physics, BITS Pilani Hyderabad Campus, Hyderabad, Telangana 500078, India, [7]Indian Institute of Science Education and Research Thiruvananthapuram, Vithura, Kerala 695551, India, [8]Department of Physics, Indian Institute of Technology Delhi, New Delhi-110016, India, [9]The Institute of Mathematical Sciences, C. I. T. Campus, Tharamani, Chennai, Tamil Nadu 600113, India, [10]Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai, Maharashtra 400094, India, [11]Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester M13 9PL, UK, [12]Department of Physics and Astronomy, University of Texas, Rio Grande Valley, Brownsville, TX 78520, USA, [13]Cahill Center for Astrophysics, California Institute of Technology, 1200 E. California Blvd. Pasadena, CA 91125, USA, [14]Fakultät für Physik, Universität Bielefeld, Postfach 100131, Bielefeld D-33501, Germany, [15]Arecibo Observatory, University of Central Florida, Arecibo, PR 00612, USA, [16]McGill Space Institute, McGill University, 3550 University Street, Montréal, QC H3A 2A7, Canada and [17]Department of Physics, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India

Abstract

We introduce pinta, a pipeline for reducing the upgraded Giant Metre-wave Radio Telescope (uGMRT) raw pulsar timing data, developed for the Indian Pulsar Timing Array experiment. We provide a detailed description of the workflow and usage of pinta, as well as its computational performance and RFI mitigation characteristics. We also discuss a novel and independent determination of the relative time offsets between the different back-end modes of uGMRT and the interpretation of the uGMRT observation frequency settings and their agreement with results obtained from engineering tests. Further, we demonstrate the capability of pinta to generate data products which can produce high-precision TOAs using PSR J1909−3744 as an example. These results are crucial for performing precision pulsar timing with the uGMRT.

Keywords: astronomy data analysis – pulsars

(Received 8 July 2020; revised 10 February 2021; accepted 23 February 2021)

1. Introduction

Ubiquitous galaxy mergers are expected to force their resident supermassive black holes to merge (Berczik et al. 2006; Pearson et al. 2019). During such merger and the preceding inspiral phases, the black hole pairs are expected to emit gravitational waves (GWs) in the nanohertz frequency range (Burke-Spolaor et al. 2019; Susobhanan et al. 2020). Pulsar Timing Arrays (PTAs: Hobbs & Dai 2017) aim to detect such GWs by accurately timing the arrival of pulses from an ensemble of millisecond pulsars (MSPs) as these are very precise celestial clocks (Hobbs et al. 2020). The most promising PTA sources include isolated supermassive black hole binaries (SMBHBs) emitting continuous GWs and an astrophysical stochastic GW background formed from an ensemble of many unresolved SMBHBs (Burke-Spolaor et al. 2019). The rapidly maturing PTA efforts are soon expected to open an additional window to the GW astronomy landscape inaugurated by the LIGO-Virgo collaboration (Abbott et al. 2019).

At present, there exist three advanced PTA experiments, namely the Parkes Pulsar Timing Array (PPTA: Hobbs 2013; Kerr et al. 2020), the European Pulsar Timing Array (EPTA: Kramer & Champion 2013; Desvignes et al. 2016), and the North American Nanohertz Observatory for Gravitational Waves (NANOGrav: McLaughlin 2013; Alam et al. 2021a, 2021b). Additionally, PTA efforts are gaining momentum in India, China, and South Africa (Joshi et al. 2018; Lee 2016; Bailes et al. 2018), and these collaborations are referred to as the emerging PTAs. The International Pulsar Timing Array (IPTA) consortium combines data and resources from various PTA efforts to enable faster detection of nanohertz GWs (Hobbs et al. 2010; Perera et al. 2019).

The Indian Pulsar Timing Array (InPTA) experiment, operational since 2015 (Joshi et al. 2018), aims to use the unique

Author for correspondence: Abhimanyu Susobhanan, E-mail: s.abhimanyu@tifr.res.in

Cite this article: Susobhanan A, Maan Y, Joshi BC, Prabu T, Desai S, Nobleson K, Susarla SC, Girgaonkar R, Dey L, Batra ND, Gupta Y, Gopakumar A, Bagchi M, Basu A, Bethapudi S, Choudhary A, De K, Krishnakumar MA, Manoharan PK, Naidu AK, Pathak D, Singha J and Surnis MP. (2021) pinta: The uGMRT data processing pipeline for the Indian Pulsar Timing Array. *Publications of the Astronomical Society of Australia* **38**, e017, 1–11. https://doi.org/10.1017/pasa.2021.12

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press

strengths of the Giant Metrewave Radio Telescope (GMRT: Swarup et al. 1991)—especially after its recent upgrade (uGMRT: Gupta et al. 2017)—along with the Ooty Radio Telescope (ORT: Swarup et al. 1971; Naidu et al. 2015) to complement the other PTA experiments. The uGMRT, with its ability to observe below 1 GHz, is an ideal instrument to characterise interstellar medium effects such as dispersion measure (DM) variations of PTA pulsars, which is necessary to achieve the nanosecond timing precision required for the first detection of nanohertz GWs (Joshi et al. 2018).

The first step in using uGMRT and ORT data for InPTA science goals is to reduce it to an *archive* format (Hotan, van Straten, & Manchester 2004)—a pulsar data format widely used among other PTAs. Then, these data can be further processed using well-known software to derive various astrophysically relevant quantities including the pulse time of arrival (TOA) and the DM (van Straten, Demorest, & Osłowski 2012). This calls for homogeneity in data reduction practices to avoid non-uniformity in the data products used for PTA analysis, which can introduce systematic errors. In this paper, we describe a uGMRT pulsar data analysis pipeline named 'Pipeline for the Indian Pulsar Timing Array' (pinta[a]), developed for the InPTA experiment to address these concerns as well as to improve the efficiency, reliability, and user friendliness of the data reduction process and to ensure faster turnaround time from observations to PTA analysis. We have developed pinta with the intention to commission it as a standard pipeline at the GMRT observatory to be used by the wider pulsar community. This can help avoid the transfer of large data files by enabling data reduction at the observatory itself.

For the pipeline to be useful to a wider community, we also discuss how to interpret the uGMRT observation frequency settings. We also present the results of our astronomical experiments carried out to validate the definition of the observing frequency in the engineering specifications of the uGMRT backend hardware and software. Using the same experiment, we also ascertained the instrumental delays between various back-end modes used at uGMRT measured through engineering tests. These delays form a crucial piece of information, not only for combining data from multiple bands in the InPTA analysis but also for other simultaneous multi-frequency observations which use different back-end modes of uGMRT.

The outline of this paper is as follows. A detailed description of the uGMRT raw data as well as the workflow and usage of pinta is provided in Section 2. Details of the uGMRT observation frequency settings and the astronomical experiments which were used to validate these settings are presented in Section 3. The performance and RFI mitigation characteristics of pinta are reported in Section 4. The ability of pinta to generate data products from which high-precision TOAs can be derived is demonstrated in Section 5 using J1909−3744 as an example. A summary of the pinta pipeline discussed in this paper is given in Section 6, and our future plans for the development of InPTA-relevant codes including pinta are summarised in Section 7.

2. Description of the pipeline

pinta accepts uGMRT raw pulsar timing data as input, performs RFI mitigation and folding, and provides the partially folded pulse profile in the Timer archive format (van Straten & Bailes 2011) as its output. In what follows, we give a detailed description of the uGMRT raw data and the workflow of the pinta pipeline.

The thirty GMRT antennas are divided in groups to form multiple subarrays, and each subarray is phased to form voltage beams for two polarisations, and the gains of the two polarisations are equalised during phasing. These voltage beams are then digitised and Fourier transformed (no polyphase filter is employed) to form power spectra across a certain number of frequency channels (Reddy et al. 2017). For the phased array (PA) mode that we use in our InPTA timing observations, the spectral powers from the two polarisations are added to form the total intensity I without applying any calibration, and is integrated maintaining the required spectral and time resolution for the observation specified in terms of the number of channels N_{chan} and the sampling time T_{smpl}. Note that the two polarisation voltages can also be combined to compute the Stokes parameters (I, Q, U, V: Hamaker, Bregman, & Sault 1996). While the recording of the full Stokes data is possible at uGMRT, the implementation of its reduction in the pipeline described here is currently being developed and tested. In addition, a real-time coherent dedispersion observing mode is employed to process the voltages to form and record the coherently dedispersed phased array (CDPA) raw data stream (De & Gupta 2016). Lastly, an incoherent array (IA) data stream can be formed by incoherently adding the spectral powers from different antennas.

The PA and the CDPA total intensity modes are used for the InPTA observations discussed in this paper. The CDPA mode is primarily used at the lower frequency bands where the effect of interstellar dispersion is prominent. The raw data stream from either of these modes, namely a data cube of spectral intensities at N_{chan} frequency channels for each time sample, are stored as 16-bit integers in a binary raw data file, and the timestamp (in Indian Standard Time) at the start of the observation is saved as a separate ASCII file. An example timestamp file is shown below.

```
#Start time and date
IST Time: 19:59:57.633098240
Date: 25:08:2018
#Start ACQ SEQ NO = 17
```

pinta converts the timestamp given in the timestamp file to MJD using astropy (Price-Whelan et al. 2018). Note that the raw data files do not store any metadata required for downstream processing, and it must be provided to the pipeline through a separate file.

Reduction of PTA data involves processing a large number of such high-volume data sets (obtained from different MSPs at different epochs in separate bands) through complex processing steps.[b] In order to ensure that processing can be efficient for such batch processing jobs and to avoid premature run-time failures, a set of checks are done on all the relevant files and folders, and the processing is initiated only if all the checks pass.[c] If one of the checks fail, an informative error message is shown to enable easier troubleshooting.

The data processing workflow of pinta is illustrated in Figure 1. pinta uses two separate packages for Radio Frequency

[a] Available at https://github.com/abhisrkckl/pinta.

[b] InPTA currently observes six pulsars at biweekly cadence simultaneously in two bands. Each such observation creates of the order of 100–150 GB of raw data per band per pulsar.

[c] These checks include the existence and read/write permissions of the relevant files and folders.

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press

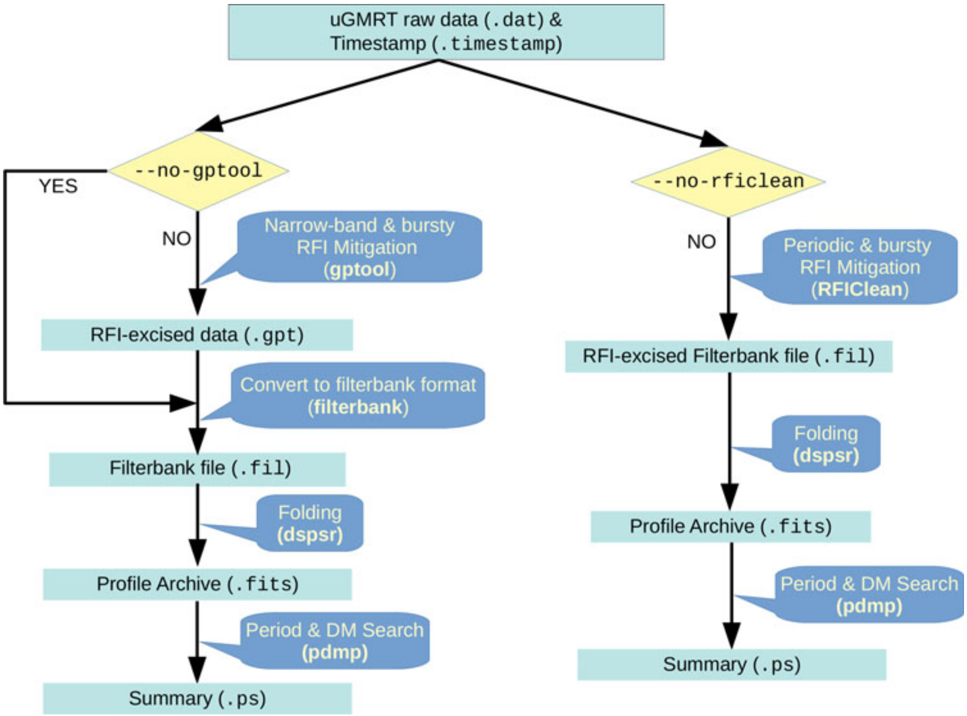

Figure 1. The workflow of `pinta`. The pinta pipeline uses uses two separate packages for the RFI mitigation, namely `gptool` and `RFIClean`. A typical data reduction workflow can optionally engage these RFI mitigation choices. Note that the profile archives generated by `pinta` are in the `Timer` format although their extension is '`.fits`'. They can be converted to `PSRFITS` format using `PSRCHIVE` (Hotan et al. 2004).

Interference (RFI) mitigation, namely `gptool`[d] (Chowdhury & Gupta, in preparation) and `RFIClean`[e] (Maan, van Leeuwen, & Vohl 2020). Brief descriptions of these packages are given below.

2.1. Details of `gptool`

`gptool` is both an RFI mitigation and a data reduction tool for the beamformer data from GMRT. It mitigates both narrow-band spectral line RFI and broadband bursty time-domain RFI. For narrow-band RFI, it offers a choice of two options for flagging RFI-affected frequency channels: (a) it derives a median band shape and flags channels for which the median absolute deviation (MAD) exceeds a defined threshold or (b) it checks for a drop in mean-to-RMS ratio for each channel below a specified threshold to identify channels corrupted by RFI. Our `pinta` pipeline employs both of these methods available in `gptool`. For identifying broadband bursty RFI, `gptool` once again offers two options for removal of outlier time samples, based on different ways of estimating central tendency and variability in the histogram of the frequency-collapsed time series. In the first method, a standard median and MAD-based scheme is employed to identify RFI-contaminated time samples. However, when strong RFI is present for a significant duration of the observation time block, the histogram may deviate from unimodality, affecting the robustness of median and MAD estimates. In such cases, the major mode and the full width at half maximum around the major mode provide robust estimates of the central tendency and variability of the underlying distribution, and a novel scheme for broadband RFI mitigation has been implemented in `gptool` based on these statistics. This novel scheme has been found to give superior results, and hence is used in our pipeline. For further handling of the channels and time samples that are flagged as RFI by `gptool`, it offers two options to the user: either to replace the existing values by zero or to replace the existing values by a local median. In our pipeline, we use the replace by the local median option as it is known to give better results. Both the RFI mitigated and unmitigated data can then be dedispersed and folded to the ephemeris of the observed pulsar. When `gptool` is run in the interactive mode, the time-series, folded profile, and the band-shape are displayed as the tool processes the raw data. `pinta` uses the non-interactive mode of `gptool`, where the RFI mitigated data, in the same format as the raw input data, is written to an output file along with estimated statistics in auxiliary files without performing dedispersion or folding. `gptool` provides an option for the removal of a baseline computed by dedispersing the data to zero DM, useful for broadband RFI mitigation, and an option for flattening the variations of the band shape across the observing bandwidth by renormalising the output of each frequency channel to the same mean value. The parameters for RFI removal and the selected modes are specified with a configuration file, named `gptool.in`. `gptool` has also been extensively used for RFI mitigation in the uGMRT for many other pulsar projects since the beginning of the wide-band observations with the uGMRT (Pleunis et al. 2020).

2.2. Details of `RFIClean`

`RFIClean` excises periodic RFI in the Fourier domain and then mitigates narrow-band spectral line RFI and broadband bursty

[d] Stands for GMRT Pulsar Tool.

[e] Available at https://github.com/ymaan4/rficlean.

time-domain RFI using robust statistics. The periodic RFI could severely limit the efficacy of conventional RFI mitigation techniques. There are many terrestrial sources of periodic interference, the most infamous being the household 50/60 Hz power lines. `RFIClean` identifies and mitigates periodic interference in the time series of individual frequency channels using Fourier domain analysis. After the excision of periodic interference, `RFIClean` uses the more conventional threshold-based techniques to identify the time samples as well as frequency channels, respectively, contaminated by broadband bursts and narrow-band RFI. The identified time samples and frequency channels are replaced by mean values, computed robustly in the local regions around the affected samples. `RFIClean` has been extensively and successfully tested against any artefacts which might get incorporated in the data during the periodic RFI excision and might be relevant to the PTA analysis. The details of these tests can be found in Maan et al. (2020). Before inclusion in `pinta`, `RFIClean` was also independently tested as a stand-alone programme using InPTA data and was found to significantly enhance the quality of the reduced data and the timing analysis. For some pulsars with their spin frequency or any of its harmonics unfavourably close to 50 Hz, detection of the pulsar signal at several epochs was possible only after `RFIClean`'s mitigation of the periodic and other RFI. `RFIClean` has also been used in several other completed and ongoing projects (e.g., Maan et al. 2019; Oostrum et al. 2020), including in timing experiments and searches for fast radio bursts (Sosa Fiscella et al. 2021; Pastor-Marazuela et al. 2020).

We note here an important difference between `gptool` and `RFIClean`: `gptool` performs band shape normalisation on the raw data while `RFIClean` retains the original band shape. Thus, noticeable difference in shape of the *band-averaged* profiles can occur between the two branches of the pipeline, especially in wide-band observations of pulsars exhibiting significant profile evolution with frequency and interstellar scintillation. Therefore, we advocate the use of separate templates for generating TOAs from profiles obtained through `gptool` and `RFIClean`, especially for high precision pulsar timing applications such as PTAs. In addition, the use of frequency-dependent two-dimensional templates may also help mitigate this issue (Pennucci 2019).

`gptool` accepts uGMRT raw data as input and writes the output in the same format. The conversion to the filterbank format is carried out by a version of the `filterbank` command provided by the `sigproc` package (Lorimer 2011), customised for uGMRT and distributed along with `pinta`. On the other hand, `RFIClean` accepts input either in uGMRT raw data format or in the `sigproc`-filterbank format and outputs a `sigproc`-filterbank file.

It may be illuminating to compare and contrast the RFI mitigation methods available in `pinta` with that available in the `CoastGuard` data analysis package[f] (Lazarus et al. 2016) developed for the PSRIX backend of the Effelsberg 100-m Radio Telescope. `CoastGuard` provides four algorithms to find and mask or replace channels, sub-integrations, and phase bins in the folded profile contaminated with RFI. The major difference between the RFI mitigation algorithms available in `pinta` and `CoastGuard` is that the former act on raw data, whereas the latter acts on folded profile archives. The mitigation of periodic RFI such as the RFI generated by power distribution lines implemented in `RFIClean` is not possible in the folded profiles. In addition, the time domain bursty RFI removed by `gptool` and `RFIClean` typically occur at GMRT at timescales much shorter than our sub-integration interval of 10 s. These are our main reasons for opting for RFI removal in the raw data rather than folded profiles in our analysis.

While both the RFI mitigation packages have been well tested, the possibility of discovering new artefacts in the future cannot be ruled out. Hence, to avoid the need of reanalysing all the data in such an unlikely future situation, we have designed `pinta` such that it allows the user to process the data in two separate branches, one for each RFI mitigation package, and produces two separate outputs. Availability of data reduced by two independent parts of the pipeline facilitates detailed comparisons and the choice of the optimal RFI mitigation method. The RFI-mitigated filterbank files are folded using `dspsr` (van Straten & Bailes 2011) and saved in the `Timer` format, significantly reducing the data volume. Finally, a period and DM search is performed on the resulting profile archive using the `pdmp` command provided by `psrchive`, producing a summary document in the postscript format. This file is used as a visual check to ensure that the pulsar has been detected and that the analysis has finished successfully.

2.3. *Usage*

The `pinta` pipeline can be invoked from the command line with the following syntax.

```
$ pinta [-help] [-test] [-no-gptool]
[-no-rficlean] [-nodel] [-retain-aux]
[-log-to-file] [-gptdir <...>]
[-pardir <...>] [-rficconf <...>]
<input_dir> <working_dir>
```

`pinta` requires specifying two mandatory parameters and a few other optional parameters as inputs as listed below.

1. **Input directory** (`input_dir`)—The directory where the raw data files and the corresponding timestamp files are stored.
2. **Working directory** (`working_dir`)—The output files, as well as all the intermediate products, will be written to this directory. This directory must contain a file named `pipeline.in` as specified in Section 2.5, and the user must have 'read' and 'write' permissions for this directory. The working directory can be the same as the input directory.
3. `gptool` **configuration directory** (`gpt_dir`)—This directory should contain the configuration files required to run `gptool`, named `gptool.in.xxx` where 'xxx' represents the local oscillator frequency of the uGMRT band.
4. **Pulsar ephemeris directory** (`par_dir`)—This directory should contain the pulsar ephemeris (`.par`) files in the `tempo2` format, required for folding the data. Each ephemeris file should be named `JNAME.par` where 'JNAME' is the name of the pulsar in the J2000 epoch.
5. `RFIClean` **configuration file** (`rficconf`)—This file contains the settings and flags required to run `RFIClean` for `pinta`.

In addition, we shall refer to the directory from which `pinta` is invoked and the directory where the `pinta` script is stored as the *current directory* (`current_dir`) and *script directory* (`script_dir`), respectively.

[f] Available at https://github.com/plazar/coast_guard.

Table 1. Command line options available in `pinta`

Argument	Description	Mandatory/Optional
Positional arguments		
`<input_dir>`	The input directory	Mandatory
`<working_dir>`	The working directory	Mandatory
Options		
`--help`	Output a help message	Optional
`--test`	Do not execute data processing commands. All checks are performed on the input files and the commands are printed on the screen. This option is present for troubleshooting	Optional
`--no-gptool`	Do not run `gptool`. Produces an output file without RFI mitigation	Optional
`--no-rficlean`	Do not run `RFIClean`	Optional
`--nodel`	The pipeline deletes all intermediate output files by default to conserve disk space. This option preserves the intermediate outputs	Optional
`--retain-aux`	Components of the pipeline produce various side products in addition to the primary data products, which are removed by `pinta` by default. This option preserves these files by moving them to a folder named `aux` inside the `working_dir`	Optional
`--log-to-file`	This option redirects the standard output generated from `pinta` to a log file in the `current_dir`	Optional
`--gptdir <...>`	Specifies the directory where the gptool configuration files are stored. By default, this is specified in the configuration file (See Section 2.4)	Optional
`--pardir <...>`	Specifies the directory where the pulsar ephemeris files are stored. By default, this is specified in the configuration file (See Section 2.4)	Optional
`--rficonf <...>`	Specifies the `RFIClean` configuration file. By default, this is specified in the configuration file (See Section 2.4)	Optional

Note that both `working_dir` and the `current_dir` require write access. The `input_dir` and `working_dir` are mandatory positional arguments to be passed to `pinta`, while `gpt_dir`, `par_dir`, and `rficconf` are by default read from a configuration file, detailed in the next subsection. `gpt_dir`, `par_dir`, and `rficconf` can be explicitly specified in the command line through the `-gptdir`, `-pardir`, and `-rficonf` options, respectively. The various options and command line arguments are summarised in Table 1.

2.4. The configuration file

The `pinta` configuration file stores the default settings required to run the pipeline, such as the `gpt_dir`, `par_dir`, and `rficconf` in `YAML` format.[g] This file should be named `pinta.yaml` and stored in the `script_dir`.

A sample configuration file is shown below.

```
pinta:
 pardir: /path/to/pulsar/ephemeris/dir/
 gptdir: /path/to/gptool/config/dir/
 rficonf: /path/to/rfiClean/config/file/
```

2.5. The `pipeline.in` file

Since the raw input data files do not contain any metadata required for downstream processing, such as the number of channels and the bandwidth, it must be provided separately. `pinta` accepts this information through a space-separated ASCII file named `pipeline.in` stored in the `working_dir`. Each row in `pipeline.in` corresponds to one raw data file and the various columns are described in Table 2. Rows starting with '#' are treated as comments and ignored. `pinta` processes rows in the `pipeline.in` files serially until all rows are processed successfully or a validation criterion is not met.

An example `pipeline.in` file is shown in Figure 2.

2.6. Storage requirements

The uGMRT raw data file generated by an hour-long observation is typically of the order of a hundred Gigabytes. A uGMRT raw data file contains, for each time sample, $N_{\rm pol}$ polarisation intensities/correlations in $N_{\rm chan}$ frequency channels represented as 16-bit integers. In general, the file size of the raw data file for an observation duration $T_{\rm obs}$ and sampling time $T_{\rm smpl}$ is given by

$$S_{\rm raw} = N_{\rm pol} N_{\rm chan} \frac{T_{\rm obs}}{T_{\rm smpl}} \times 2 \text{ Bytes.} \tag{1}$$

The intermediate products generated by the pipeline, namely, `.gpt` and `.fil` files, will have roughly the same size as the input file along with a small header which stores observation metadata. The output archive files are typically smaller, of the order of hundreds of Megabytes in size, since we fold the raw data over longer sub-integrations. The size of the output archive, excluding the header, is approximately given by

$$S_{\rm arch} \sim \frac{T_{\rm smpl}}{T_{\rm subint}} N_{\rm bin} S_{\rm raw}, \tag{2}$$

where $T_{\rm subint}$ is the duration of a sub-integration and $N_{\rm bin}$ is the number of phase bins in the profile. In our analysis, we typically use $T_{\rm subint} = 10$ s. In general, the maximum amount of disk space required by `pinta` is less than four times the total size of the raw

[g] https://yaml.org/.

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press

Table 2. Description of various columns in the `pipeline.in` file

Column	Parameter	Description	Data type	Unit
1	JName	The name of the pulsar in J2000 epoch	String	
2	RawDataFile	Raw data file name. Only the file name is required and not the full path	String	
3	TimestampFile	Timestamp file name. Only the file name is required and not the full path	String	
4	Frequency (F_{LO})	Local oscillator frequency of the observing band	Float	MHz
5	NBins (N_{bin})	Number of phase bins for the folded profile	Integer	
6	NChans (N_{chan})	Number of frequency channels	Integer	
7	BandWidth (ΔF)	Bandwidth of the observing band	Float	MHz
8	TSample (T_{smpl})	The sampling time used for observation	Float	s
9	SideBand	The side-band. This should be either LSB (lower side-band) or USB (upper side-band)	String	
10	NPol (N_{pol})	Number of polarisations (1:=(I), 4:=(I,Q,U,V))	Integer	
11	TSubInt (T_{subint})	The duration of individual sub-integrations within which the data will be folded over the pulsar period	Float	s
12	Cohded	Whether the data has been coherently dedispersed (De & Gupta 2016). 1 represents Yes and 0 represents No	Boolean	

```
#JName       RawData                         Timestamp                             Freq   Nbin   NChan   BandWidth     TSmpl        SB    NPol   TSubint   Cohded
J1939+2134   J1939+2134.25032019.B3.cdp.dat  J1939+2134.25032019.B3.cdp.timestamp  500    128    1024    100           0.00008192   LSB   1      10.0      1
J1939+2134   J1939+2134.25032019.B4.pa.raw   J1939+2134.25032019.B4.pa.hdr         750    128    1024    100           0.00008192   LSB   1      10.0      0
J1939+2134   J1939+2134.25032019.B5.cdp.dat  J1939+2134.25032019.B5.cdp.timestamp  1460   128    1024    100           0.00008192   LSB   1      10.0      1
```

Figure 2. An example `pipeline.in` file.

data files, while preserving all intermediate files (i.e., using the `--nodel` option). If the `--nodel` option is not used, the maximum amount of disk space required is approximately the size of the largest raw data file.

3. Interpretation of observatory frequency settings

The GMRT Wide-band Back-end (GWB; Reddy et al. 2017) provides three different observation modes, namely IA, PA, or CDPA, as described in Section 2. The settings used during a pulsar observation depend on the band of observation and the mode of the observatory back-end. These settings are required for data reduction using `pinta` and are communicated to the pipeline through a `pipeline.in` file as mentioned in Section 2.5. As the frequency labelling of the pulsar data cube varies with the back-end mode used, these need to be determined and encoded in `pinta` in a manner which simplifies the specification of observation settings for the user.

The times of arrivals (TOAs) of a pulsar pulse recorded simultaneously in two bands A and B, using backend modes P and Q, respectively, are related by

$$t_{AP} - t_{BQ} = \Delta_{PQ} + \mathcal{D} \times \mathrm{DM}\left(F_{1A}^{-2} - F_{1B}^{-2}\right) , \qquad (3)$$

where t_{AP} and t_{BQ} are the TOAs, Δ_{PQ} is the relative instrumental offset between modes P and Q, $\mathcal{D}$ is the dispersion measure constant, DM is the dispersion measure of the pulsar at the epoch of observation, and F_{1A} and F_{1B} are the frequency labels of the channels to which the signals in bands A and B are dedispersed. Both the offsets Δ_{PQ} and the frequency labels F_{1X} (where X represents the band of observation) are crucial for performing precision pulsar timing using uGMRT. These are defined as part of the engineering specifications of the GWB hardware and software (Reddy et al. 2017; De & Gupta 2016). Engineering tests with standard inputs to the hardware were carried out to verify these definitions and revealed that there is no offset between time series in IA and PA mode, whereas a 1 buffer (256 Mbytes) offset exists between IA/PA and CDPA modes. This offset is 0.67108864 s for 200 and 400 MHz bandwidths and 1.34217728 s for 100 MHz bandwidth, and this was verified up to 5 ns precision in engineering tests. Likewise, the frequency definitions were worked out from engineering considerations and tested in an engineering sense with fixed frequency tones. While the precision of astronomical tests is not likely to be high due to system noise and coarser sampling, nevertheless such tests with wide-band radio emission are also needed to gain confidence, particularly for coherently dedispersed data. In this section, we describe the astronomical tests carried out to validate the frequency labelling F_{1X} to be encoded in `pinta` and to determine the offsets Δ_{PQ}.

3.1. Calibration experiment

The required frequency labelling and the instrumental offsets were validated using observations of the Crab pulsar (PSR J0534+2200) and PSR J0332+5434. The former is a bright pulsar with 33.7 ms period and a relatively high DM (56.7 pc cm^{-3} : Lyne et al. 2014). The DM of the Crab pulsar varies from epoch to epoch, and this pulsar exhibits sporadic intense pulses, called giant pulses (GPs; Lundgren et al. 1995; Hankins et al. 2003), typically once every four minutes at uGMRT frequencies at uGMRT sensitivity. The GPs provide a time marker, which is a strong function of frequency due to interstellar dispersion. Moreover, the arrival times of this marker across different frequencies vary with epoch due to DM variations. Thus, GPs provide a sensitive probe to validate the assumed frequency labels for the spectral data. PSR J0332+5434,

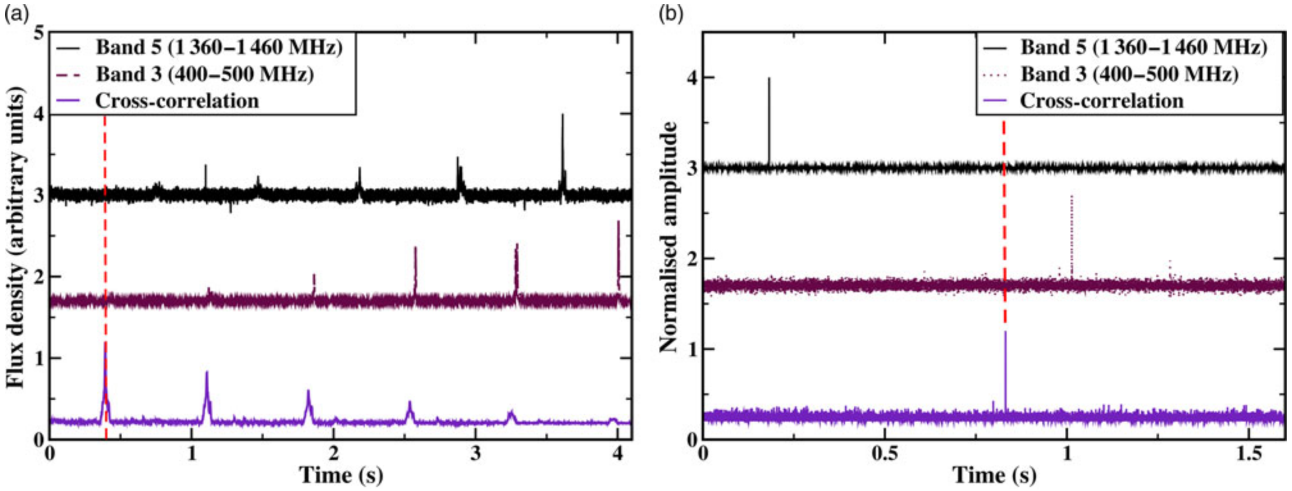

Figure 3. Time series observed using Band 5 (1 360–1 460 MHz : top plot in each panel) and Band 3 (400–500 MHz : middle plot in each panel) was used to determine the delay between the two bands using pulsars PSRs J0332+5434 and J0534+2200. The delay is obtained from the lag measured using the cross-correlation (shown in the bottom plot of each panel) of the two time series. The delay in each case was compared with that expected (labelled with vertical red dashed lines in the plot) due to dispersion in ionised interstellar medium to determine both the frequency definition as well as relative pipeline delays : (a) Observations of single pulses of the bright pulsar J0332+5434 showing a delayed single pulse pattern in Band 3 compared to Band 4, (b) Observations of a Giant pulse of PSR J0534+2200 where the delay between Band 5 (top plot) and Band 3 (middle plot) was found consistent with that expected due to dispersion, assuming the correct frequency definitions (Equations (4a) and (4b)) and zero relative fixed pipeline delay. (a) PSR J0332+5434 (b) PSR J0534+2200.

with a flux density of ∼1 500 mJy at 408 MHz, is the brightest pulsar in the northern hemisphere at metre-centimetre wavelengths with a period of 714 ms and a DM of 26.76 pc cm^{-3} (Lorimer et al. 1995; Hassall et al. 2012). Bright single pulses with pulse-to-pulse intensity variations interspersed with pulse nulls are seen in this pulsar (see Figure 3a).

The GWB can simultaneously be used in its different modes of operation in different bands using any combination of the four beams provided (Gupta et al. 2017; Reddy et al. 2017). This capability was exploited to record data on GPs from the Crab pulsar and single pulses from PSR J0332+5434 in IA, PA, and CDPA modes of GWB using different frequency bands available with the uGMRT. For the Crab pulsar, first the GPs were identified in IA, PA, and CDPA mode data at both Band 3 and Band 5. We investigated the cross-correlation in the recorded time series around the identified GPs from different modes and frequency bands to determine the lag in the arrival times of the GPs. This lag, recorded for example with PA in Band 5 and CDPA in Band 3, depends on the DM of the pulsar (specified up to a precision of 0.001 pc cm^{-3}) and the frequency labeling used for the two bands, as given by Equation (3). As the DM time series of this pulsar is known to the required precision from independent measurements (Lyne, Pritchard, & Graham Smith 1993; Lyne et al. 2014) made public by the Jodrell Bank Observatory,[h] the expected lag in the arrival times of identified GPs was calculated from the DM nearest to the epoch of observations. Hence, any difference between the expected and measured lags is due to either (a) incorrect frequency labelling or (b) relative time offset between the two modes. As the DM of this pulsar varies over a timescale of one month, two observations separated by one month will yield different delays due to frequency labelling, whereas the relative instrumental delay is expected to be constant. Thus, both the frequency labelling and relative offsets can be simultaneously determined by two such observations. We check these results for consistency using similar analysis with PSR J0332+5434.

3.2. Calibration observations and results

Calibration observations were carried out on 2019 December 16 (MJD 58832), 2020 January 24 (MJD 58871), and 2020 May 22 (MJD 58991). The estimated lags for one combination of modes on 2020 January 24 are shown in Figure 3a and b. The relative offsets and frequency labelling were then determined by matching the measured and expected lags, given by Equation (3), and the estimated relative offsets for different modes are tabulated in Table 3. While the uncertainty on measurements of these relative pipeline delays ranges from 10 to 80 μs due to coarser sampling and system noise, these measurements are consistent with the engineering measurements. The relative pipeline delays measured as a result of tests conducted in the first two epochs were corrected in the software by the GMRT engineering team in 2020 April. This was verified in the tests conducted on 2020 May 22, as can be seen from Table 3.

The frequency labelling F_{1X} for the different modes are expressed in terms of the value of the highest frequency channel in the following expressions:
For IA and PA,

$$F_{1X} = \begin{cases} F_{\rm LO} & \text{for LSB} \\ F_{\rm LO} + \Delta F & \text{for USB} \end{cases}, \tag{4a}$$

and for CDPA,

$$F_{1X} = \begin{cases} F_{\rm LO} - \frac{\Delta F}{N_{\rm chan}} & \text{for LSB} \\ F_{\rm LO} + \Delta F\left(1 - \frac{1}{N_{\rm chan}}\right) & \text{for USB} \end{cases}. \tag{4b}$$

[h] http://www.jb.man.ac.uk/pulsar/crab/crab2.txt.

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press

Table 3. Results of time delay measurements simultaneously at two different frequency using PSR J0534+2200 for validating frequency definitions and relative pipeline delays (Δ_{PQ}) for different modes of pulsar observations. The epoch of observations is given in the first column along-with Dispersion measure at that epoch in second column followed by sampling time used, expected and observed delay in samples for different combination of modes at the two frequencies in fourth, fifth, sixth, seventh and third column respectively. The last column presents the relative pipeline delays (Δ_{PQ}). The abbreviations B5CDPA, B3CDPA, B5PA and B3PA indicate data acquisition using Band 5 in CDPA mode, using Band 3 in CDPA mode, Band 5 in PA mode, and Band 3 in PA mode respectively

Epoch	DM		Sampling time	Expected	Observed	Δ_{PQ}
(MJD)	(pc cm-3)	Bands and modes	(μs)	delay (s)	delay (s)	(s)
58 832	56.7528	B5CDPA–B3CDPA	81.92	0.83172	0.83165(8)	0.0
		B3CDPA–B5PA	81.92	0.83173	0.83173(8)	1.34218
		B5CDPA–B5PA	81.92	0.00001	0.00008(8)	1.34226
		B5PA–B3PA	81.92	0.83137	0.83141(8)	0.0
58 871	56.7401	B5CDPA–B3CDPA	20.48	0.83259	0.83259(2)	0.0
		B3CDPA–B5PA	81.92	2.1748	2.1746(2)	1.342177
		B5PA–B3PA	81.92	0.8312	0.8312(1)	0.0
58 991	56.7781	B5CDPA–B3CDPA	5.12	0.8374	0.8375(1)	0.0
		B5CDPA–B4PA	40.96	0.39062	0.39051(8)	0.0
		B3CDPA–B4PA	40.96	0.4468	0.4470(2)	0.0
		B4PA–B5PA	40.96	0.4470	0.4472(2)	0.0
		B5PA–B5IA	40.96	0.0	0.0	0.0

Here, F_{LO} refers to the Local Oscillator (LO) frequency (MHz) used for the observations, ΔF is the acquisition bandwidth (typically 100 or 200 MHz), and N_{chan} denotes the number of channels or sub-bands across the band. The expression is different for each side band denoted by USB or LSB. When F_{LO} is chosen at the lowest edge of the band being used, this is called upper side band (USB) where frequencies are ordered from lowest to highest frequency. The reverse order of frequencies are used in lower side band (LSB) with F_{LO} chosen at the highest edge of the band. Equations (4a)–(4b) are in agreement with what is expected from the implementation of the IA, PA, and CDPA pipelines in GWB (Reddy et al. 2017; De & Gupta 2016).

These equations are implemented in `pinta` to make it simpler for the user to use our data reduction pipeline. The user specifies the LO frequency, the side band, the acquisition bandwidth, and the number of sub-bands/channels in the `pipeline.in` file using the same values as specified for the backend observation setup. The relative offsets determined in these experiments are not coded in `pinta`, but are included as jumps while performing any timing analysis of the uGMRT data.

4. Performance

To validate the pipeline and investigate its performance, we performed a series of tests using a variety of uGMRT data sets with varying data volume and observation frequencies.

To gauge the computational performance of `pinta`, we sliced the raw data files from ten different observations (the details of these data sets are given in Table 4) into file sizes of 1 GiB,[i] 2 GiB, 4 GiB, 8 GiB, 16 GiB, and 32 GiB, processed each slice separately in `pinta`, and in each case recorded the execution time of each component of `pinta` as well as the total execution time. The result of this exercise is shown in Figure 4 where the ratio of the execution time to the observation duration (observe-to-reduce time ratio) is plotted against the observation duration. Each point in Figure 4 represents the median of ten test cases, and the error bar represents the corresponding median absolute deviation. This plot shows the observe-to-reduce ratio to be approximately between 1.5 and 3 and that it is not strongly dependent on the data volume. This behaviour is desirable and the observe-to-reduce ratio can indeed be improved to be better than real-time by optimising and parallelising the pipeline, which we plan to do in the future. Such improvements can in principle allow `pinta` to be deployed as a real-time observatory pipeline for pulsar data reduction. We also note that the observe-to-reduce ratio while using only one of the two branches is close to or better than real-time.

Table 4. The details of the datasets used for characterizing the performance and RFI mitigation efficacy of `pinta`. Bands 3, 4 and 5 represent 400–500 MHz, 650–750 MHz, and 1 360–1 460 MHz respectively for our observations

Dataset	Pulsar	Date	Band	Coherent dedispersion
1	J1857+0943	25/08/2018	5	Yes
2	J2145−0750	22/05/2018	4	No
3	J2145−0750	25/08/2018	3	Yes
4	J2145−0750	10/09/2018	3	Yes
5	J1939+2134	21/05/2018	3	Yes
6	J1939+2134	28/09/2018	4	No
7	J1713+0747	07/06/2018	5	Yes
8	J2124−3358	10/09/2018	3	Yes
9	J1643−1224	08/07/2018	3	Yes
10	J1643−1224	25/08/2018	5	Yes

[i] 1 GiB = 2^{30} Bytes.

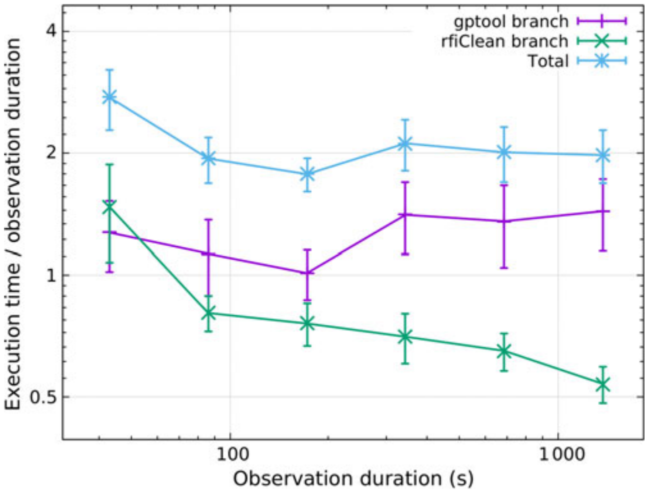

Figure 4. Ratio of execution time by observation duration (observe-to-reduce ratio) plotted versus the observation duration. The observe-to-reduce ratio for each of the two branches of `pinta` as well as the same for the entire pipeline is plotted. Each data point represents the median of 10 tests, and the error bars represent the corresponding median absolute deviation.

To ensure the reliability of the pipeline, these tests were repeated by multiple users on the same data sets mentioned above using different command line options, and the results were compared with each other as well as with results obtained by running the various data reduction codes used in pinta directly to ensure that the results are reproducible.

4.1. *RFI mitigation*

RFI mitigation is one of the most important processing steps in the `pinta` pipeline. In order to illustrate the RFI mitigation in the pipeline, we present here a study on ten different data sets (see Table 4), each having varying levels of RFI. Data segments were selected from the uGMRT observation bands 3, 4, and 5, MJD 58260-58389 with a total length for the segments 11 544 s. The data quality of each segment prior to and after the `pinta` RFI mitigation was studied. The `rfifind` command of `PRESTO` (Ransom 2011) was used to report the percentage of good intervals in the data. The percentage of good intervals that is gained after the RFI mitigation is shown (in red) in Figure 5. This study provides a feel for the typical RFI mitigation available in the pipeline, and we see from Figure 5 that the degree of improvement after applying RFI mitigation varies greatly from data set to data set, which is expected since the RFI environment itself is highly variable. Data set 3 is of specific interest as the percentage of good intervals more than doubles after applying RFI mitigation, and the pulsar was detected in this data set only after applying RFI mitigation.

To further illustrate the efficacy of the RFI mitigation available in `pinta`, we show in Figure 6 pulse profiles generated using `gptool`, `RFIClean` and without performing any RFI mitigation for two observations. The profiles without any RFI mitigation are produced by running `pinta` with `-no-gptool -no-rficlean` options. The signal to noise ratios (SNRs) quoted in Figure 6 are computed using the `pdmp`[j] command of `PSRCHIVE`. In light of the caveat regarding band shape normalisation discussed in Section 2, we have chosen two observations without significant interstellar scintillation in order to show a fair comparison between `gptool` and `RFIClean`.

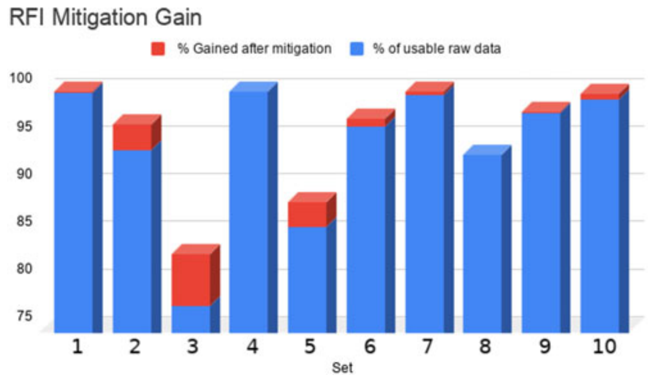

Figure 5. Effectiveness of RFI Mitigation. Each bar represents one data set. The details of each data set are given in Table 4.

Figure 6 shows the gain in profile SNR for both data sets while using RFI mitigation. Nevertheless, it should be noted that the SNRs for J2124−3358 reported by `pdmp` may be inaccurate due to its large duty cycle. This does not affect our comparison between the RFI mitigated and non-RFI mitigated data sets as it is clear from the bottom panel of Figure 6b that the RFI mitigated profiles agree with each other better than with the non-RFI mitigated profile, indicating a reduction in the noise level.

5. Timing of PSR J1909–3744

In this section, we demonstrate the capability of `pinta` to generate profiles from which high-precision TOAs can be derived. We use PSR J1909−3744 as an example for this purpose.

The data presented in this section were obtained as part of the InPTA campaign from 2020 April to 2020 October with a cadence of $\sim$15 d. The observations were carried out by splitting the 30 uGMRT antennas into two phased subarrays, where the innermost 8 antennas were used in Band 3 (300–500 MHz) and 16 of the outer antennas were used in Band 5 (1 260–1 460MHz). The pulsar was observed simultaneously in both bands in each epoch, with 200 MHz bandwidth and 1 024 frequency channels in each band. The Band 3 data were coherently dedispersed to the known DM of the pulsar and were recorded at 20.48 μs sampling time, whereas Band 5 data were obtained using the PA mode with a sampling time of 40.96 μs. The data were processed using `pinta`, and the TOAs were extracted from the resulting `Timer` archives using `PSRCHIVE` after time and frequency collapsing the folded profiles. The resulting TOAs were fit using `TEMPO2` (Hobbs, Edwards, & Manchester 2006) using the pulsar ephemeris available in the NANOGrav 12.5 yr data set (Alam et al. 2021a), as our data span is too short to provide a reliable timing solution. Post-fit residuals after fitting for pulsar rotational parameters (F0, F1), and DM are plotted in Figure 7. We do not use any time offsets between the two bands as such offsets were corrected in GWB software since 2020 April based on results mentioned in Section 3.1. The corresponding pre-fit and post-fit parameters, along with the RMS timing residual values, are listed in Table 5. A more thorough timing solution of this data using frequency-resolved TOAs, DM corrections, and rigorous noise analysis will be published elsewhere.

[j] http://psrchive.sourceforge.net/manuals/psrstat/algorithms/snr/.

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press

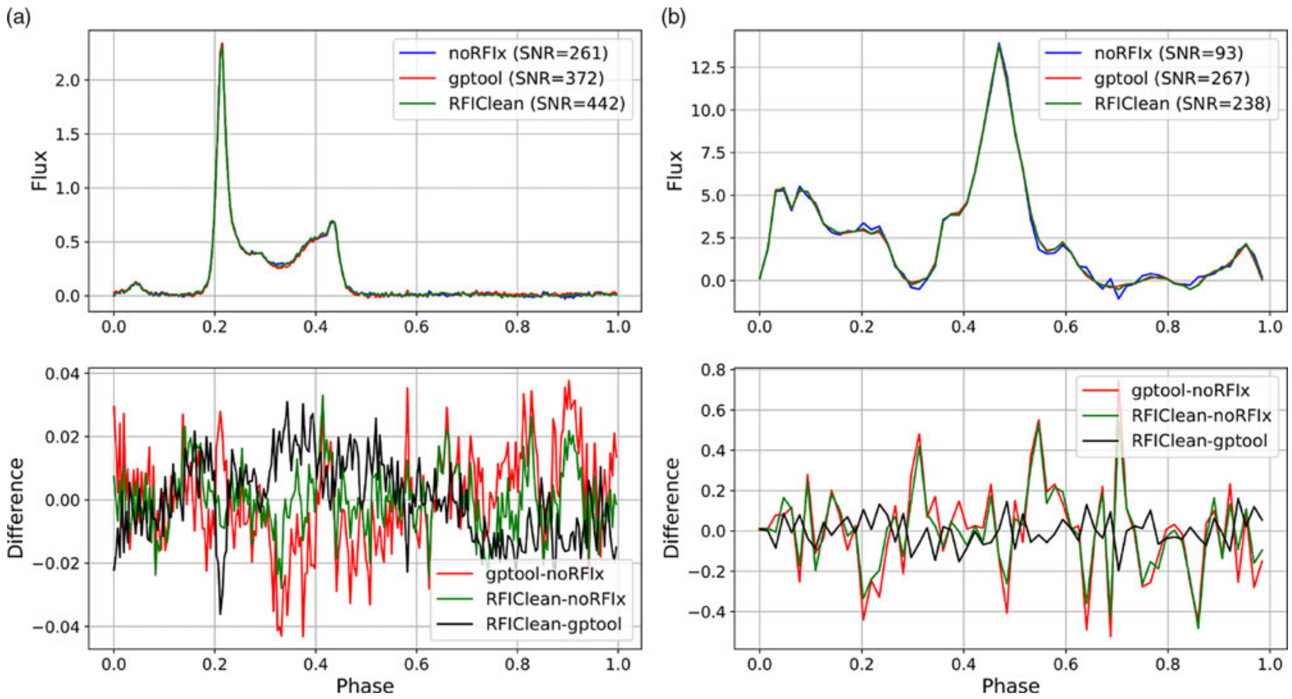

Figure 6. Comparison of frequency collapsed profiles obtained using `gptool`, `RFIClean`, and without any RFI mitigation (`noRFIx`). The `noRFIx` profiles are generated using the `-no-gptool -no-rficlean` options. The fluxes are uncalibrated and are in arbitrary units. The SNRs reported in the plots are obtained using the `pdmp` command. Both epochs show significant improvement in SNR while using RFI mitigation. (a) PSR J2145−0750 observed on 2020 June 16 in Band 5 (1 260–1 460 MHz) with 40.96 μs sampling time and no coherent dedispersion. The total integration time is 55 min. (b) PSR J2124-3358 observed on 2018 August 25 in Band 3 (400–500 MHz) with 81.92 μs sampling time with coherent dedispersion. The total integration time is 24 min.

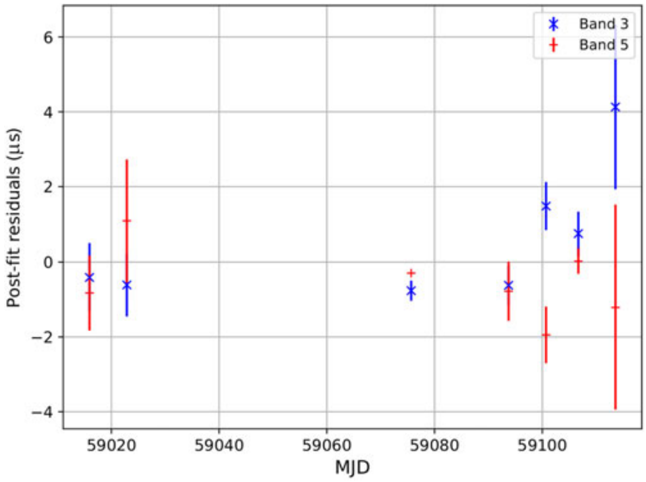

Figure 7. The timing residuals for PSR J1909−3744 generated using uGMRT observations processed with `pinta`. Band 3 is 300–500 MHz and Band 5 is 1 260–1 460 MHz. We used the ephemeris available in the NANOGrav 12.5 yr data set, and after changing the PEPOCH and DMEPOCH to MJD 59050, we fitted for F0, F1, and DM. The fit parameters are listed in Table 5. The timing residuals have an RMS of 1.46 μs.

Table 5. The fit parameters for PSR J1909−3744 generated using uGMRT observations processed with `pinta`. We used the ephemeris available in NANOGrav 12.5 dataset, and after changing the PEPOCH and DMEPOCH to MJD 59050, and fitted for F0, F1 and DM. The timing residuals are plotted in Figure 7

Parameter	Post-fit value	Post-fit uncertainty
F0 (Hz)	339.315691914442	1.2e-12
F1 (s^{-2})	−1.52e-15	2.5e-17
DM (pc/cm^3) . .	10.39090	0.00001
RMS residuals (μs)	1.46	

From Table 5, we note that the uGMRT observations processed using `pinta` are able to produce an RMS post-fit timing residuals of 1.46 μs. This demonstrates that the data products produced using `pinta` can indeed be used for high-precision timing applications such as PTAs. We expect to further reduce the RMS timing residuals after applying DM corrections, which are discussed elsewhere (Krishnakumar et al. 2021).

6. Summary and discussion

We have developed a pipeline to reduce uGMRT pulsar timing raw data for the InPTA experiment, named `pinta`, which reduces the raw data input to RFI-mitigated folded profile archives. Since the uGMRT raw data input does not contain any metadata such as the observation settings, they are provided to the pipeline via an ASCII input file named `pipeline.in`, whose contents are summarised in Table 2. `pinta` performs RFI mitigation using two different packages, namely `gptool` and `RFIClean`, running them in two different branches which produce two different output archives. `pinta` provides various command line options to control how these two branches are run, and these are summarised in Table 1.

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press

It is crucial to use the correct interpretation of the observatory frequency settings while performing the data reduction. We performed validation and calibration experiments using GPs from the Crab pulsar and single pulses from the bright pulsar J0332+5434 to ensure that our interpretation of the observation frequency for IA, PA, and CDPA pipelines of uGMRT matches what is given in Equations (4a) and (4b). This experiment also allowed us to measure the instrumental delays between IA, PA, and CDPA pipelines of uGMRT, which are consistent with the instrumental delays expected from engineering considerations.

To characterise the computational performance of `pinta`, we conducted a number of tests using different data sets. These tests showed that the net observe-to-reduce time ratio of `pinta` is approximately 2, while the observe-to-time ratio of individual branches is less than 1.5. These results lead us to strive to achieve real-time observe-to-time ratio by employing parallelisation techniques to the pipeline. We also conducted tests to investigate the RFI mitigation efficacy of `pinta` on the same data sets, the results of which are shown in Figure 5. We observe that the RFI mitigation gains seen in different data sets, having different RFI characteristics, vary significantly as expected, with some data sets yielding up to $\sim$ 10% gain after RFI mitigation. We also demonstrate improvements in the significance of pulse profiles by using the different RFI mitigation paths in `pinta`, which further advocates their importance in the pipeline. These results substantiate the addition of RFI mitigation tools in `pinta`. To demonstrate the ability of `pinta` to generate data products from which high-precision TOAs can be derived, we showed the timing of uGMRT observations of PSR J1909−3744, and we are able to produce timing residuals with RMS of the order of 1 μs.

7. Future scope

Our plans for the future development of `pinta` include the improvement of its computational efficiency to achieve better than real-time performance. This may be achieved by (a) running the two branches of the pipeline parallelly instead of serially, (b) modifying the `filterbank` program to use GPUs, and (c) utilising the GPU processing option in `dspsr`.

Similar pipelines for reducing the data obtained using the legacy GMRT and the ORT are also under development, ensuring a high level of compatibility with `pinta`. In addition, we plan on developing 'InPTA Data Management System', a database for tracking metadata associated with the observations and data analysis of the InPTA experiment, which will be tightly integrated with `pinta` as well as the legacy GMRT and ORT pipelines.

Acknowledgements. We are grateful to the anonymous referee for a detailed perusal and constructive feedback on the manuscript. We thank the staff of the GMRT who made our observations possible. GMRT is run by the National Centre for Radio Astrophysics of the Tata Institute of Fundamental Research. BCJ, YG, and AB acknowledge the support of the Department of Atomic Energy, Government of India, under project # 12-R&D-TFR-5.02-0700. AS, AG, and LD acknowledge the support of the Department of Atomic Energy, Government of India, under project # 12-R&D-TFR-5.02-0200. MPS acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 694745). AC acknowledges the funding received from Department of Science and Technology, Government of India, WOS-A scheme, file no. SR/WOS-A/PM-26/2018.

References

Abbott, B. P., et al. 2019, PhRvX, 9, 031040
Alam, M. F., et al. 2021a, ApJ, 252, 4
Alam, M. F., et al. 2021b, ApJ, 252, 5
Bailes, M., et al. 2018, in Proceedings of MeerKAT Science: On the Pathway to the SKA—PoS(MeerKAT2016), 11, 10.22323/1.277.0011
Berczik, P., Merritt, D., Spurzem, R., & Bischof, H.-P. 2006, ApJ, 642, L21
Burke-Spolaor, S., et al. 2019, AAR, 27, 5
Chowdhury, A., & Gupta, Y. 2021, In preparation
De, K., & Gupta, Y. 2016, ExA, 41, 67
Desvignes, G., et al. 2016, MNRAS, 458, 3341
Gupta, Y., et al. 2017, CSci, 113, 707
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, AAS, 117, 137
Hankins, T. H., Kern, J. S., Weatherall, J. C., & Eilek, J. A. 2003, Natur, 422, 141
Hassall, T. E., et al. 2012, A&A, 543, A66
Hobbs, G. 2013, CQG, 30, 224007
Hobbs, G., & Dai, S. 2017, NSR, 4, 707
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655
Hobbs, G., et al. 2010, CQG, 27, 084013
Hobbs, G., et al. 2020, MNRAS, 491, 5951
Hotan, A. W., van Straten, W., & Manchester, R. N. 2004, PASA, 21, 302
Joshi, B. C., et al. 2018, JAA, 39, 51
Kerr, M., et al. 2020, PASA, 37, e020
Kramer, M., & Champion, D. J. 2013, CQG, 30, 224009
Krishnakumar, M. A., et al. 2021, arXiv e-prints, p. arXiv:2101.05334
Lazarus, P., Karuppusamy, R., Graikou, E., Caballero, R. N., Champion, D. J., Lee, K. J., Verbiest, J. P. W., & Kramer, M. 2016, MNRAS, 458, 868
Lee, K. J. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain, & D. Li (Astronomical Society of the Pacific), 19, http://www.aspbooks.org/a/volumes/article_details/?paper_id=37688
Lorimer, D. R. 2011, SIGPROC: Pulsar Signal Processing Programs (ascl:1107.016)
Lorimer, D. R., Yates, J. A., Lyne, A. G., & Gould, D. M. 1995, MNRAS, 273, 411
Lundgren, S. C., Cordes, J. M., Ulmer, M., Matz, S. M., Lomatch, S., Foster, R. S., & Hankins, T. 1995, ApJ, 453, 433
Lyne, A. G., Jordan, C. A., Graham-Smith, F., Espinoza, C. M., Stappers, B. W., & Weltevrede, P. 2014, MNRAS, 446, 857
Lyne, A. G., Pritchard, R. S., & Graham Smith, F. 1993, MNRAS, 265, 1003
Maan, Y., Joshi, B. C., Surnis, M. P., Bagchi, M., & Manoharan, P. K. 2019, AsJL, 882, L9
Maan, Y., van Leeuwen, J., & Vohl, D. 2020, arXiv e-prints, p. 2012.11630
McLaughlin, M. A. 2013, CQG, 30, 224008
Naidu, A., Joshi, B. C., Manoharan, P. K., & Krishnakumar, M. A. 2015, ExA, 39, 319
Oostrum, L. C., van Leeuwen, J., Maan, Y., Coenen, T., & Ishwara-Chandra, C. H. 2020, MNRAS, 492, 4825
Pastor-Marazuela, I., et al. 2020, arXiv e-prints, p. arXiv:2012.08348
Pearson, W. J., et al. 2019, A&A, 631, A51
Pennucci, T. T. 2019, ApJ, 871, 34
Perera, B. B. P., et al. 2019, MNRAS, 490, 4666
Pleunis, Z., et al. 2020, arXiv e-prints, p. arXiv:2012.08372
Price-Whelan, A. M., et al. 2018, ApJ, 156, 123
Ransom, S. 2011, PRESTO: PulsaR Exploration and Search TOolkit (ascl:1107.017)
Reddy, S. H., et al. 2017, JAI, 06, 1641011
Sosa Fiscella, V., et al. 2021, ApJ, 908, 158
Susobhanan, A., Gopakumar, A., Hobbs, G., & Taylor, S. R. 2020, PhRvD, 101, 043022
Swarup, G., et al. 1971, NPS, 230, 185
Swarup, G., Ananthakrishnan, S., Kapahi, V. K., Rao, A. P., Subrahmanya, C. R., & Kulkarni, V. K. 1991, CS, 60, 95
van Straten, W., & Bailes, M. 2011, PASA, 28, 1
van Straten, W., Demorest, P., & Osłowski, S. 2012, ART, 9, 237

https://doi.org/10.1017/pasa.2021.12 Published online by Cambridge University Press