



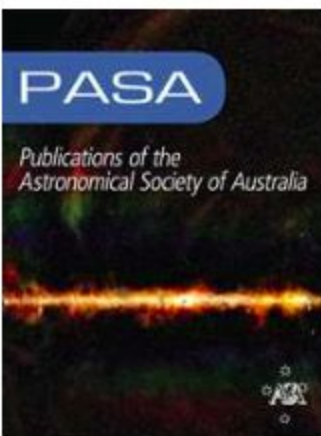
PINTA:

*The uGMRT Pipeline for Indian
Pulsar Timing Array*

**InPTA Student Week 2023:
Phase I**

Debabrata Deb, IMSc




























**Publications of the
Astronomical Society
of Australia**

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

Part of: Data Analysis Pipelines and Software

Published online by Cambridge University Press: **14 April 2021**

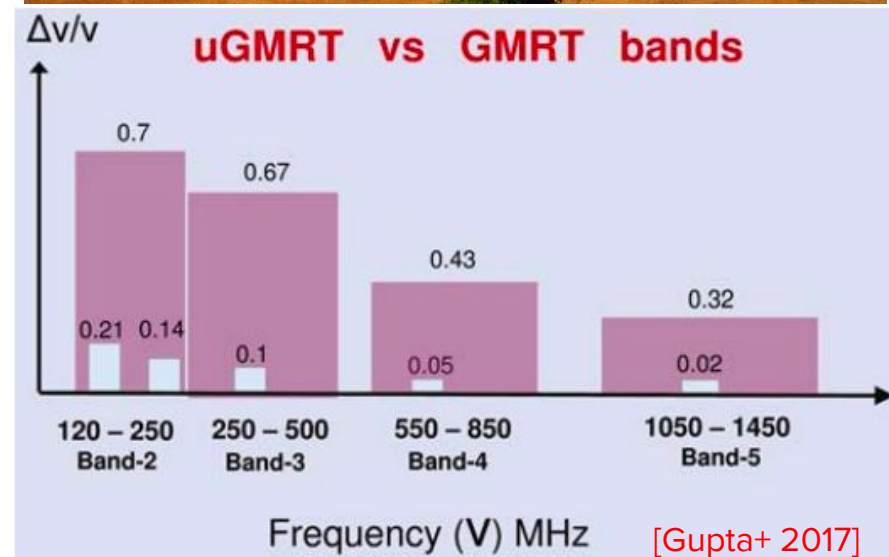
Abhimanyu Susobhanan , Yogesh Maan , Bhal Chandra Joshi , T. Prabu ,
Shantanu Desai , K. Nobleson , Sai Chaitanya Susarla , Raghav Girgaonkar ,
Lankeswar Dey , Neelam Dhanda Batra , Yashwant Gupta , A. Gopakumar ,
Manjari Bagchi , Avishek Basu , Suryarao Bethapudi , Arpita Choudhary ,
Kishalay De , M. A. Krishnakumar , P. K. Manoharan , Arun Kumar Naidu ,
Dhruv Pathak , Jaikhomba Singha  and Mayuresh P. Surnis 

Show author details 

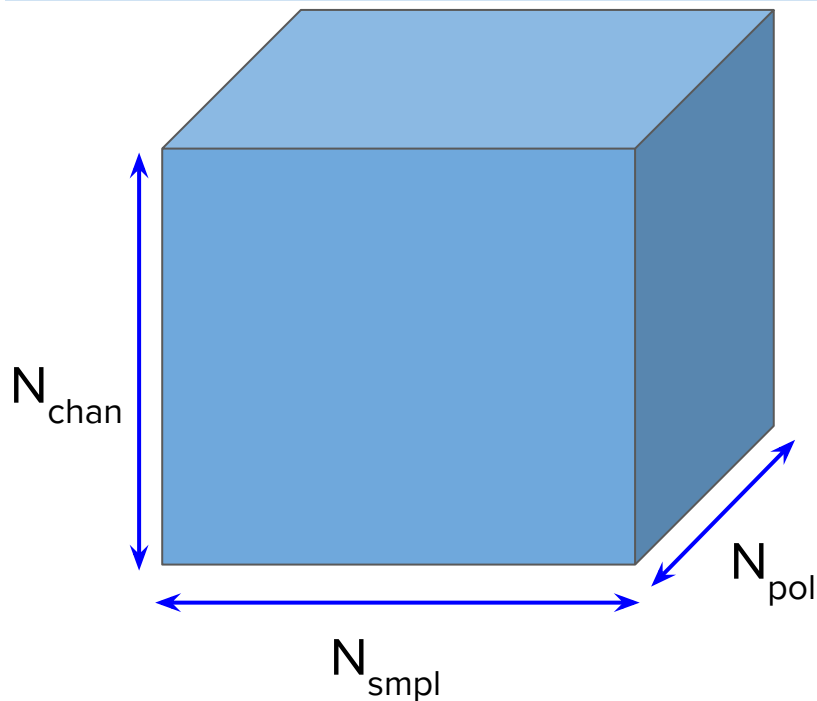
<https://doi.org/10.1017/pasa.2021.12>

The upgraded GMRT

- 30-element interferometer
- New wide-band feeds provide seamless frequency coverage between 150-1460 MHz
- 4 phased array beams – simultaneous multi-frequency observations. (IA/PA/CDPA)
- A new GPS-synchronized hydrogen maser for precision timing – 1-10 ns precision timing.



uGMRT Pulsar Data Format



Data Volume = $N_{\text{chan}} * N_{\text{smpl}} * N_{\text{pol}} * 2$
bytes

- The binary raw data file stores **Npol** polarization products in **Nchan** channels for every time sample as **16-bit integers**.
 - Npol = 1 : Total intensity
 - Npol = 4 : Stokes I,Q,U,V
- The timestamp at the start of observation is stored as an ASCII file.

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

pinta Overview

- pinta is a python script which calls various pulsar data processing codes to reduce the GMRT pulsar raw data to a folded Timer archive.
 - Timer format is compatible with packages used for downstream processing.
- Performs RFI mitigation using two different packages.
 - **RFIClean**
 - **gptool**
- Metadata required for processing is provided as an ASCII input file.

<https://github.com/inpta/pinta>

Summary

- Inputs

- uGMRT raw data (.dat)
- Timestamp (.timestamp)
- Pulsar ephemeris (.par)
- Config file (pipeline.in)

- Output

- Folded profile
 - With RFIClean
 - With gptool
 - Without RFI removal

Summary

- Inputs

- **uGMRT raw data (.dat)**

A binary file that contains
 $N_{\text{pol}} * N_{\text{chan}} * N_{\text{sample}}$ values.

- Timestamp (.timestamp)
- Pulsar ephemeris (.par)
- Config file (pipeline.in)

- Output

- Folded profile
 - With RFIClean
 - With gptool
 - Without RFI removal

Summary

- Inputs

- uGMRT raw data (.dat)
- **Timestamp (.timestamp)**
- Pulsar ephemeris (.par)
- Config file (pipeline.in)

ASCII file containing the timestamp of start of observation

- Output

- Folded profile
 - With RFIClean
 - With gptool
 - Without RFI removal

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


Summary

- Inputs

- uGMRT raw data (.dat)
- Timestamp (.timestamp)
- **Pulsar ephemeris (.par)**
- Config file (pipeline.in)

Pulsar ephemeris in
TEMPO2 format.

- Output

- Folded profile
 - With RFIClean
 - With gptool
 - Without RFI removal

PSRJ	J1857+0943	
ELONG	286.86348828	1.000e-08
ELAT	32.32148622	2.000e-08
DM	13.3140	2.500e-03
PEPOCH	55367.00000	
F0	186.494081249931	3.000e-12
F1	-6.2046E-16	3.000e-20
POSEPOCH	55367.00	
DMEPOCH	56106	
BINARY	ELL1	
PB	12.32717119157	1.800e-10
A1	9.2307802	3.000e-07
TASC	55360.513155155	1.900e-08
EPS1	-2.150E-5	3.000e-08
EPS2	2.440E-6	1.800e-08
CLK	TT(BIPM2015)	
EPHEM	DE436	
RM	22.2	9.000e-01
PX	0.6	2.000e-01
DM1	0.0017	2.000e-04
PMELONG	-3.27	1.000e-02
PMELAT	-5.06	2.000e-02
STIG	0.966	5.000e-03
H3	1.07E-6	4.000e-08
UNITS	TDB	

Summary

- Inputs
 - uGMRT raw data (.dat)
 - Timestamp (.timestamp)
 - Pulsar ephemeris (.par)
 - **Config file (pipeline.in)**

#JName	RawData	Timestamp	Freq	Nbin	NChan	BW	TSmpl	SB	NPol	TSubint	Cohded
J0437-4715	J0437-4715_bm3_pa_1460_200_8_18mar2023.raw	J0437-4715_bm3_pa_1460_200_8_18mar2023.raw.hdr	1460	-1	1024	200	0.00004096	LSB	1	10	0
J0751+1807	J0751+1807_bm3_pa_1460_200_8_18mar2023.raw	J0751+1807_bm3_pa_1460_200_8_18mar2023.raw.hdr	1460	-1	1024	200	0.00004096	LSB	1	10	0
J0613-0200	J0613-0200_bm3_pa_1460_200_8_18mar2023.raw	J0613-0200_bm3_pa_1460_200_8_18mar2023.raw.hdr	1460	-1	1024	200	0.00004096	LSB	1	10	0
J0740+6620	J0740+6620_bm3_pa_1460_200_8_18mar2023.raw	J0740+6620_bm3_pa_1460_200_8_18mar2023.raw.hdr	1460	-1	1024	200	0.00004096	LSB	1	10	0
J0900-3144	J0900-3144_bm3_pa_1460_200_8_18mar2023.raw	J0900-3144_bm3_pa_1460_200_8_18mar2023.raw.hdr	1460	-1	1024	200	0.00004096	LSB	1	10	0
J1022+1001	J1022+1001_bm3_pa_1460_200_8_18mar2023.raw	J1022+1001_bm3_pa_1460_200_8_18mar2023.raw.hdr	1460	-1	1024	200	0.00004096	LSB	1	10	0

- With gptool
- Without RFI removal

Summary

- Inputs

- uGMRT raw data (.dat)
- Timestamp (.timestamp)
- Pulsar ephemeris (.par)
- Config file (pipeline.in)

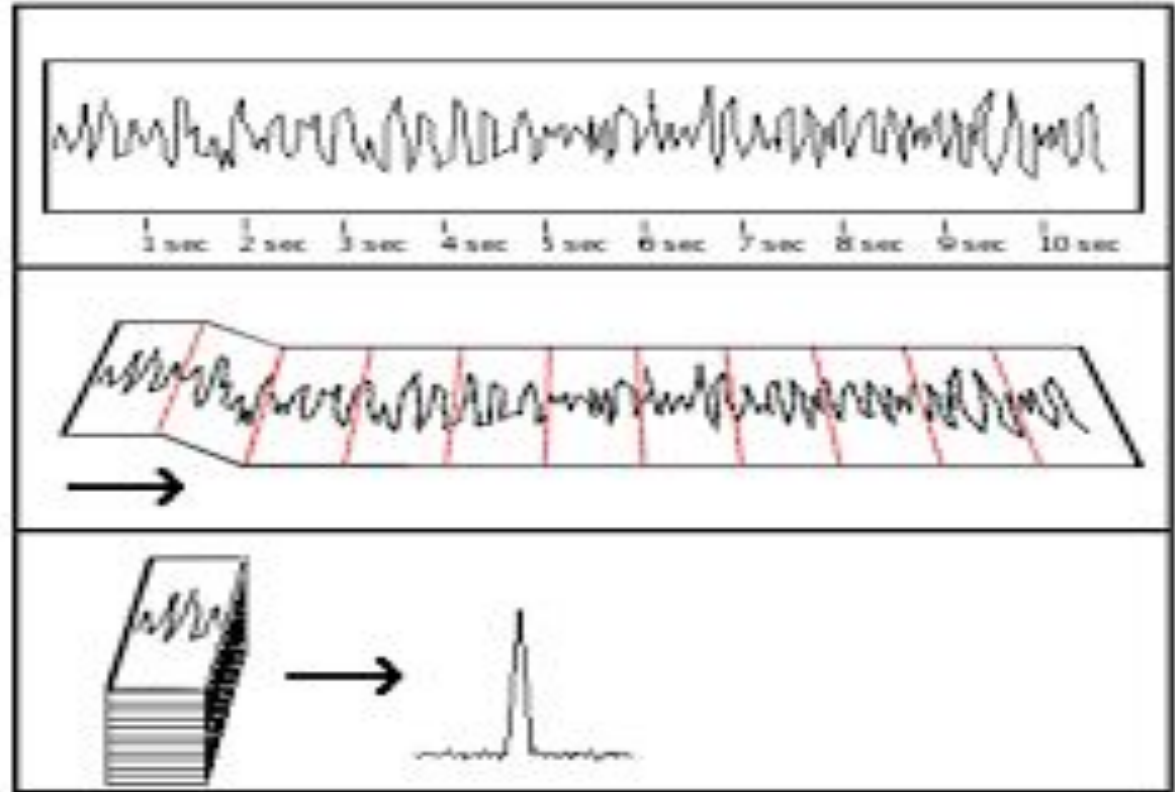
- Output

- **Folded profile (.fits)**

- With RFIClean
- With gptool
- Without RFI removal

- FITS files containing $N_{\text{bin}} * N_{\text{subint}} * N_{\text{chan}} * N_{\text{pol}}$ entries.
- Can be manipulated using psrchive.

Folding using dsp



Folding - Add large number of pulses in phase to improve signal to noise ratio.

What is RFI?

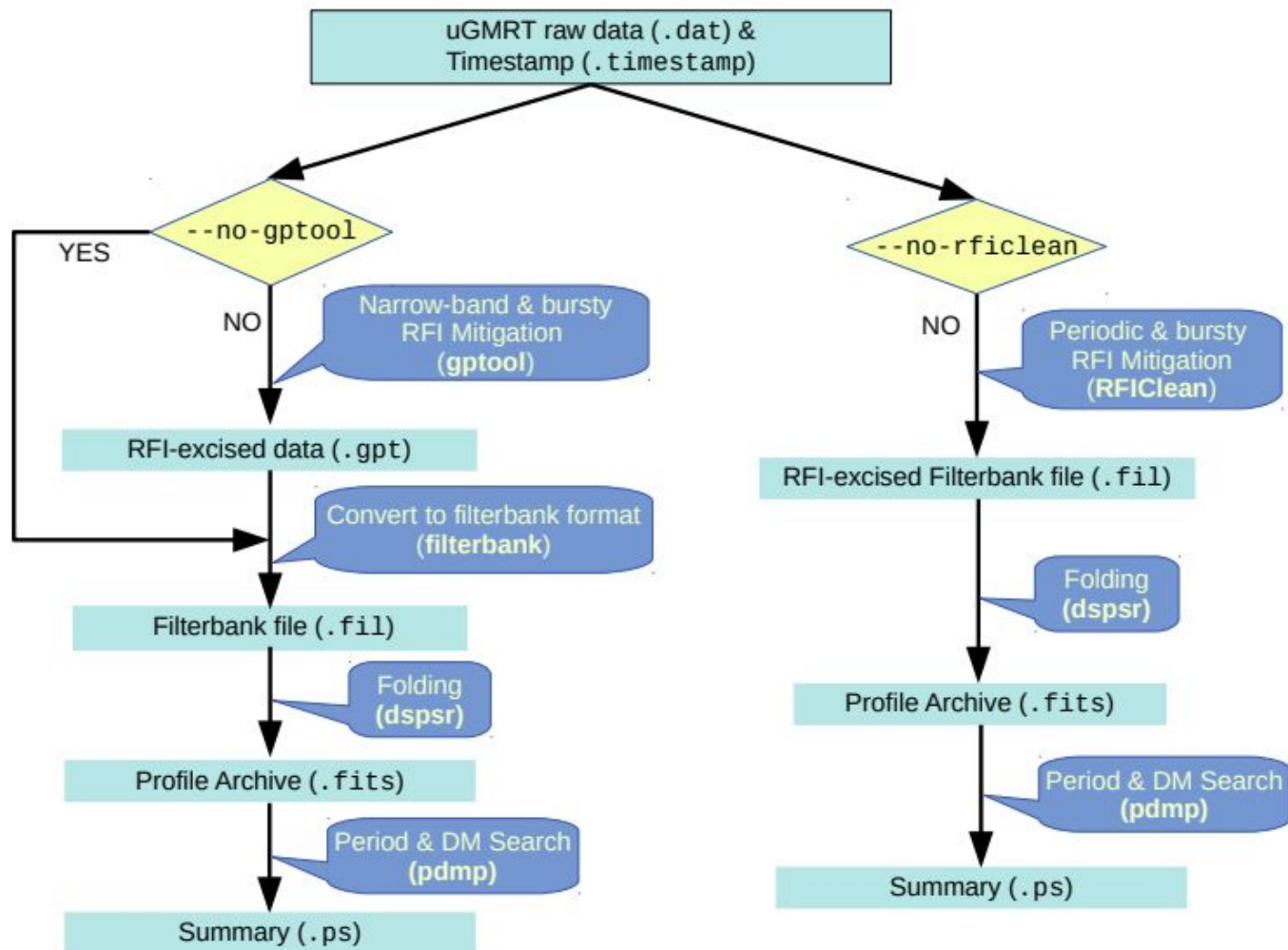
- Unwanted signals picked up by the telescope - mostly human-made
- Strong RFI completely swamps the astrophysical signal.
- Also presence of Weak RFI reduces S/N ratio.
- Broadly three categories of RFI seen at GMRT
 - Periodic RFI
 - Eg: 50 Hz Power lines
 - Spectral line RFI
 - Eg: Satellites, TV signals, mobile phone signals, ...
 - Bursty RFI
 - Eg: Opening your microwave oven before it is done cooking (*google **perytones at Parkes***), turning on your water pump, car spark plug, ...

gptool:

- narrow-band spectral line RFI and broadband bursty time-domain RFI

rfiClean:

- periodic RFI in the Fourier domain,
- and then mitigates narrow-band spectral line RFI and broadband bursty time-domain RFI



Usage

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

- `gptdir` = Folder containing gptool configuration files
- `pardir` = Folder containing pulsar ephemeris files
- `input_dir` = Folder containing raw data and timestamp files.
- `working_dir` = Folder containing pipeline.in file. The output will be written to this location.

pipeline.in

#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

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 polarizations (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	

Creating pipeline.in

1. File name tells us

- a. The pulsar name
- b. GWB mode : ia/pa/cdp
Cohded = 1 for cdp
Cohded = 0 for pa / ia
- c. Observation date

```
-rw-rw-r--. 1 visitor1 svisitor 49G Dec 8 16:00 J2124-3358_300_200_bm2_08Dec2023.raw0
-rw-rw-r--. 1 visitor1 svisitor 1.2K Dec 8 15:42 J2124-3358_300_200_bm2_08Dec2023.raw0.ahdr
-rw-r--r--. 1 visitor1 svisitor 96K Dec 8 15:42 J2124-3358_300_200_bm2_08Dec2023.raw0.bhdr
-rw-rw-r--. 1 visitor1 svisitor 90 Dec 8 15:42 J2124-3358_300_200_bm2_08Dec2023.raw0.hdr
```

Creating pipeline.in

- The observation settings we need are Frequency, Bandwidth, NChan, SideBand, NPol, TSmpl, and Cohded.
- Frequency & Cohded can be found from the file name
- Find the setup file and command file for Bandwidth, NChan, SideBand, NPol, TSmpl.
 - For CDP, TSmpl is not directly given in the setup file.

Step-by-step

- Copy all the raw data and timestamp files to a directory (This will be the `input_dir`).
 - Make sure that you have read permission to all input files.
- Create your working directory
 - Make sure that this directory has write permissions and has sufficient disk quota.
- Create pipeline.in file
 - Either from a template or from scratch using observation files.
- Run the pipeline.
 - This may take a long time.
 - Use of screen/nohup command is recommended.

Step-by-step

- Copy all PINTA generated output files at Kaveri
 - Copy PINTA generated pdmp summary output files and fits files at the required directory in Kaveri within the respective pulsar directory BAND wise.
- Copy pipeline.in and pinta_summary.txt file at Kaveri
 - Rename pipeline.in and pinta_summary.txt file based on epoch, BAND and observation date and copy to the required directory at Kaveri

```
-rw-rw----. 1 visitor1 pulsarg 7.6M Oct 29 11:49 J1939+2134_60117.939563_1400.norfix.fits
-rw-rw----. 1 visitor1 pulsarg 65K Oct 29 11:49 J1939+2134_60117.939563_1400.norfix.summary.pdf
-rw-rw----. 1 visitor1 pulsarg 7.6M Oct 29 12:01 J1939+2134_60117.939563_1400.rfiClean.fits
-rw-rw----. 1 visitor1 pulsarg 65K Oct 29 12:01 J1939+2134_60117.939563_1400.rfiClean.summary.pdf
drwxrwx---. 2 visitor1 pulsarg 4.0K Oct 29 11:54 RFIClean_ps
drwxrwx---. 3 visitor1 pulsarg 50 Oct 29 11:43 log
-rw-rw----. 1 visitor1 pulsarg 12K Oct 29 12:02 output_B5.log
-rw-rw----. 1 visitor1 pulsarg 263 Oct 29 12:01 pinta_summary.txt
-rw-rw----. 1 visitor1 pulsarg 226 Oct 29 11:43 pipeline.in
```

Step-by-step

- Copy all PINTA generated output files at Kaveri
 - Copy PINTA generated pdmp summary output files and fits files at the required directory in Kaveri within the respective pulsar directory BAND wise.
- Copy pipeline.in and pinta_summary.txt file at Kaveri
 - Rename pipeline.in and pinta_summary.txt file based on epoch, BAND and observation date and copy to the required directory at Kaveri

```
-rwxrwx--- 1 prabu          ugmrtpsr 843 Dec  5 00:31 pipeline.in.07Nov2023.60254.BAND3
-rwxrwx--- 1 prabu          ugmrtpsr 921 Dec  5 00:21 pipeline.in.07Nov2023.60254.BAND5
```

- Enter snr values achieved from gptool/norfix and rfclean to the SNR.log file based on the pulsars and BANDs

```
-rwxrwx--- 1 prabu          ugmrtpsr 1.6K Dec  5 00:31 pinta_summary.txt.07Nov2023.60254.BAND3
-rwxrwx--- 1 prabu          ugmrtpsr 1.6K Dec  5 00:21 pinta_summary.txt.07Nov2023.60254.BAND5
```

- Rename setup file and observation log based on epoch and observation date and copy to the required directory at Kaveri

Step-by-step

```
#=====J0613-0200/BAND3/CYCL45=====
#MJD          gptool          norfix          rfiClean
#=====
60267.982743   -              342.88          870.50
60274.820178   -              23.82           85.72
60244.951594   --             16.08           53.30
60285.815473   --             52.76          233.55
```

- Enter data to SNR.log files present at Kaveri
 - Enter snr values achieved from gptool/norfix and rficlean to the SNR.log file based on the pulsars and BANDs
- Copy setup file and observation log at Kaveri
 - Rename setup file and observation log based on epoch and observation date and copy to the required directory at Kaveri

Step-by-step

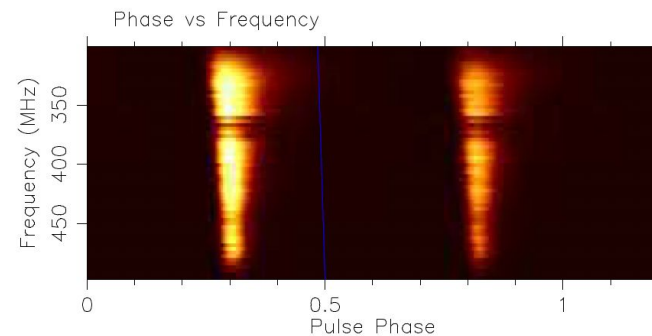
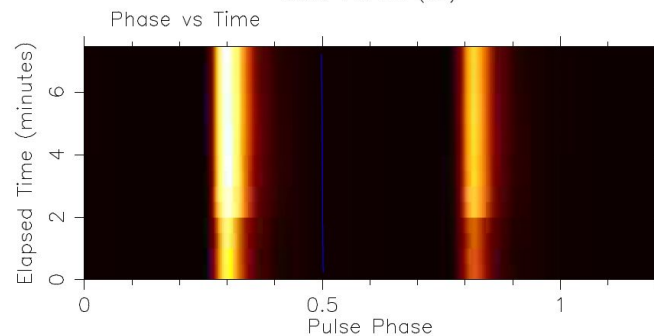
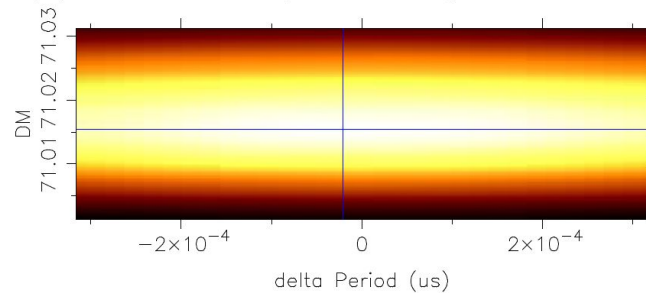
- Copy all PINTA generated output files at Kaveri
 - Copy PINTA generated pdmp summary output files and fits files at the required directory in Kaveri within the respective pulsar directory BAND wise.
- Copy pipeline.in and pinta_summary.txt file at Kaveri
 - Rename pipeline.in and pinta_summary.txt file based on epoch, BAND and observation date and copy to the required directory at Kaveri

```
-rwxrwx--- 1 aman.srivastava ugmrtpsr 4.8K Nov  8 10:07 45_006_31Oct2023.obslog.txt
-rwxrwx--- 1 aman.srivastava ugmrtpsr 9.2K Nov  8 10:09 gtac_45_006_31Oct2023_1400.txt
-rwxrwx--- 1 ptarafdar      ugmrtpsr 7.0K Nov 11 18:55 45_006_11Nov2023.obslog.txt
-rwxrwx--- 1 ptarafdar      ugmrtpsr 7.9K Nov 11 18:56 gtac_45_006_11Nov2023_1000.txt
```

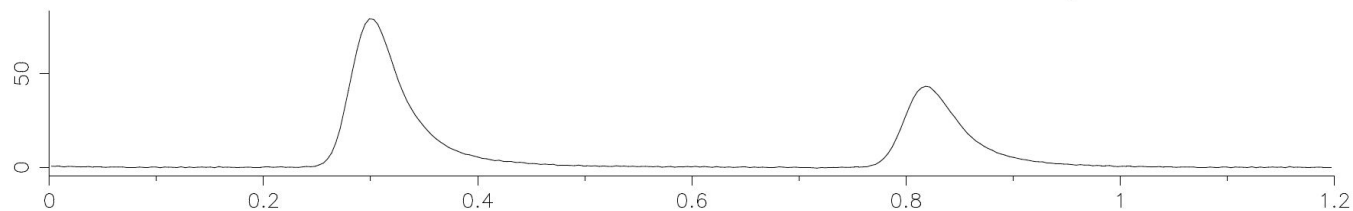
- Copy setup file and observation log at Kaveri
 - Rename setup file and observation log based on epoch and observation date and copy to the required directory at Kaveri

J1939+2134: ./J1939+2134_60259.515851_500.rfiClean.fits

BC P(ms)= 1.557806583 TC P(ms)= 1.557917538 DM= 71.016 RAJ= 19:39:38.56 DecJ= 21:34:59.1
BC MJD = 60259.517962 Centre freq(MHz) = 399.219 Bandwidth(MHz) = -200 l = 57.509 b = -0.290
NBin = 256 NChan = 64 NSub = 15 TBin(ms) = 0.006 TSub(s) = 30.000 TSpan(s) = 446.775
P(us): offset = 0.00000, step = 0.00002, range = 0.00032 DM: offset = 0.000, step = 0.000, range = 0.015



BC prd (ms):	1.557806562	TC prd (ms):	1.557917516	DM:	71.016	BC freq (Hz):	641.928224356
Corrn (ms):	-0.000000021	Corrn (ms):	-0.000000021	Corrn:	-0.001	Freq err. (Hz):	0.000015024
Error (ms):	0.000000036	Error (ms):	0.000000036	Error:	0.000	Width (ms):	0.116
						Best S/N:	528.68



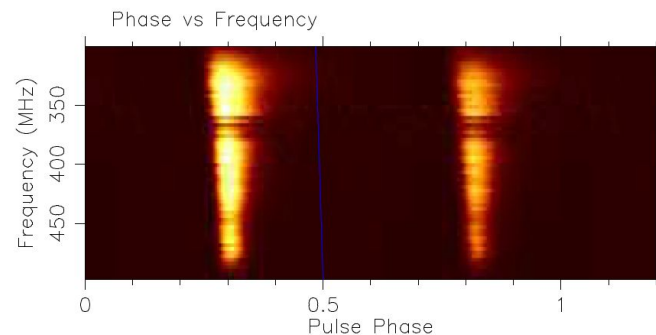
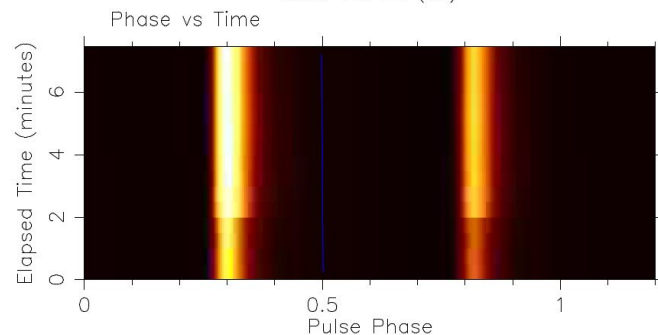
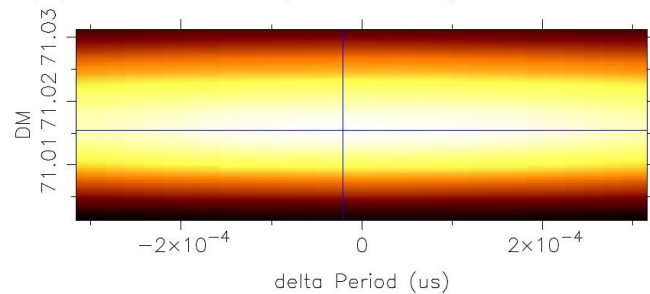
J1939+2134: ./J1939+2134_60259.515851_500.norfix.fits

BC P(ms)= 1.557806583 TC P(ms)= 1.557917538 DM= 71.016 RAJ= 19:39:38.56 DecJ= 21:34:59.1

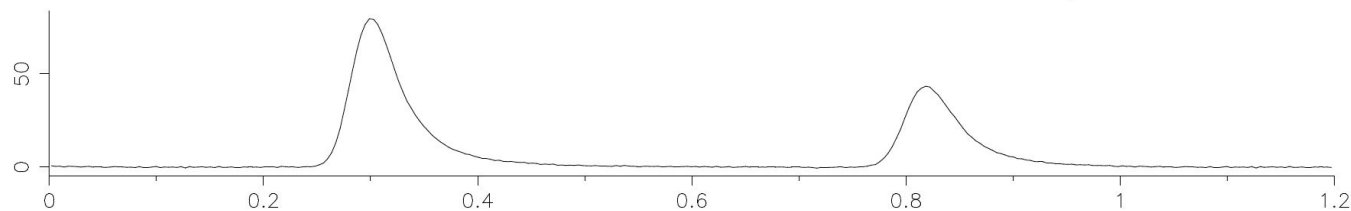
BC MJD = 60259.517962 Centre freq(MHz) = 399.219 Bandwidth(MHz) = -200 l = 57.509 b = -0.290

NBin = 256 NChan = 64 NSub = 15 TBin(ms) = 0.006 TSub(s) = 30.000 TSpan(s) = 446.775

P(us): offset = 0.00000, step = 0.00002, range = 0.00032 DM: offset = 0.000, step = 0.000, range = 0.015



BC prd (ms):	1.557806562	TC prd (ms):	1.557917516	DM:	71.016	BC freq (Hz):	641.928224356
Corrn (ms):	-0.000000021	Corrn (ms):	-0.000000021	Corrn:	-0.001	Freq err. (Hz):	0.000015024
Error (ms):	0.000000036	Error (ms):	0.000000036	Error:	0.000	Width (ms):	0.116
						Best S/N:	514.54



```

[visitor1@fs4 BAND5]$ psredit J1939+2134_60117.939563_1400.rfiClean.fits
file           Name of the file                      J1939+2134_60117.939563_1400.rfiClean.fits
nbin           Number of pulse phase bins           32
nchan          Number of frequency channels        1024
npol           Number of polarizations              1
nsubint        Number of sub-integrations           91
type           Observation type                    Pulsar
site           Telescope name                       GMRT
name           Source name                          J1939+2134
coord          Source coordinates                  19:39:38.561+21:34:59.12
freq          Centre frequency (MHz)              1300.09765625
bw            Bandwidth (MHz)                   -200
dm            Dispersion measure (pc/cm^3)        71.0195007324219
rm            Rotation measure (rad/m^2)      0
dmc           Dispersion corrected                  0
rmc           Faraday Rotation corrected           0
polc          Polarization calibrated            0
scale         Data units                      FluxDensity
state         Data state                        Intensity
length        Observation duration (s)            901.776261120001
int*:@        int:help for attribute list
ext:obs_mode  Observation Mode                      PSR
ext:obsfreq   Centre frequency                    1300.09765625
ext:obsbw     Bandwidth                          -200
ext:obsnchan  Number of channels                    1024
ext:hdrver    Header Version                     6.2
ext:date      File Creation Date                 2023-10-29T06:31:52
ext:coord_md  Coordinate mode                       J2000
ext:equinox   Coordinate equinox                   2000
ext:trk_mode  Tracking mode                       UNSET
ext:bpa       Beam position angle              0
ext:bmaj      Beam major axis                0
ext:bmin      Beam minor axis                  0
ext:stt_date  Start UT date                         UNSETTUNSE
ext:stt_time  Start UT
ext:stt_imjd  Start MJD                            60117
ext:stt_smjd  Start second                          81179
ext:stt_offs  Start fractional second                0.111732318226132

```


ext:stt_1st	Start LST	0
ext:stt_crd1	Start coord 1	00:00:00.000
ext:stt_crd2	Start coord 2	+00:00:00.000
ext:stp_crd1	Stop coord 1	UNSET
ext:stp_crd2	Stop coord 2	UNSET
ext:ra	Right ascension	19:39:38.561
ext:dec	Declination	+21:34:59.121
obs:observer	Observer name(s)	
obs:projid	Project name	
rcvr:name	Receiver name	uGMRT_B5
rcvr:basis	Basis of receptors	lin
rcvr:hand	Hand of receptor basis	+1
rcvr:sa	Symmetry angle of receptor basis	45deg
rcvr:rph	Reference source phase	0deg
rcvr:fdc	Receptor basis corrected	0
rcvr:prc	Receptor projection corrected	0
rcvr:ta	Tracking angle of feed	0deg
be:name	Name of the backend instrument	GWB
be:phase	Phase convention of backend	+1
be:dcc	Downconversion conjugation corrected	0
be:phc	Phase convention corrected	0
be:delay	Backend propn delay from digi. input.	-0
be:config	Configuration filename	1400 200 1024 LSB 40.96 0
be:nrcvr	Number of receiver channels	0
be:tcycle	Correlator cycle time	0
hist:nrow	Number of rows in history	2
hist:nbin_prd	Nr of bins per period	32
hist:tbin	Time per bin or sample	4.8679660057411e-05
hist:chan_bw	Channel bandwidth	-0.1953125
hist:cal_file	Calibrator filename	NONE
aux:dm_model	Auxiliary dispersion model	NONE
aux:dmc	Auxiliary dispersion corrected	0
aux:rm_model	Auxiliary birefringence model	NONE
aux:rmc	Auxiliary birefringence corrected	0
sub:int_type	Time axis (TIME, BINPHSPERI, BINLNGASC, etc)	TIME
sub:int_unit	Unit of time axis (SEC, PHS (0-1), DEG)	SEC
sub:tsamp	[s] Sample interval for SEARCH-mode data	0
sub:nbits	Nr of bits/datum (SEARCH mode 'X' data, else 1)	-1
sub:nch_strt	Start channel/sub-band number (0 to NCHAN-1)	-1
sub:nsblk	Samples/row (SEARCH mode, else 1)	-1
sub:nrows	Nr of rows in subint table (search mode)	91
sub:zero_off	Zero offset for SEARCH-mode data	0
sub:signint	1 for signed ints in SEARCH-mode data, else 0	0

DATA REDUCTION USING AUTOMATED PINTA IN KAVERI

1. Login to kaveri using ssh
2. Type the following commands:

```
newgrp ugmrtpsr  
umask  
bash
```

0007

3. Create two directories input directory and working directory.
4. Navigate to BAND3/BAND5 within your working directory and run the following command:

For BAND3 data:

```
nohup /Data/debabrata/heilpinta/heilpinta_B3.sh 'input_dir' 'output_dir' 'Obs_date'  
> output_B3.log
```

For BAND5 data:

```
nohup /Data/debabrata/heilpinta/heilpinta_B5.sh 'input_dir' 'output_dir' 'Obs_date'  
> output_B5.log
```

where

`input_dir` is the full path of the input directory,

`output_dir` is the full path of the output/working directory,

`Obs_date` is the date of the observation in the format: 01Jan2024

DATA REDUCTION USING AUTOMATED PINTA IN FS4

1. Login to FS4 using ssh.
2. Type the following commands:

```
newgrp pulsarg
umask 0007
Bash
```

3. Create two directories input directory and working directory.
4. Navigate to BAND3/BAND5 within your working directory and run the following command:

For BAND3 data:

```
nohup /Data/debabrata/heilpinta/heilpinta_B3.sh 'input_dir' 'output_dir' 'Obs_date'
'hostname' 'username'> output_B3.log
```

For BAND5 data:

```
nohup /Data/debabrata/heilpinta/heilpinta_B5.sh 'input_dir' 'output_dir' 'Obs_date'
'hostname' 'username' > output_B5.log
```

where

`input_dir` is the full path of the input directory,

`output_dir` is the full path of the output/working directory,

`Obs_date` is the date of the observation in the format: 01Jan2024

`hostname` is the hostname of the Kaveri server. Usually, it's kaveri.ncra.tifr.res.in

`username` is the username of your Kaveri account

Thank you !

