Searching for Pulsars with PRESTO

By Scott Ransom NRAO / UVa

Getting PRESTO

Homepage:

http://www.cv.nrao.edu/~sransom/presto/

 PRESTO is freely available from github https://github.com/scottransom/presto

 You are highly encouraged to fork your own copy, study / modify the code, and make bugfixes, improvements, etc....

For this tutorial...

- You will need a fully working version of PRESTO (including the python extensions)
- If you have questions about a command, just try it out!
 Typing the command name alone usually gives usage info.
- You need at least 1GB of free disk space
 - Linux users: if you have more than that amount of RAM, I encourage you to do everything in a subdirectory under /dev/shm
- Commands will be > typewriter script
- The sample dataset that I'll use is here (25MB) http://www.cv.nrao.edu/~sransom/GBT_Lband_PSR.fil

Outline of a PRESTO Search

- 1) Examine data format (readfile)
- 2) Search for RFI (rfifind)
- 3) Make a topocentric, DM=0 time series (prepdata and exploredat)
- 4) FFT the time series (realfft)
- 5) Identify "birdies" to zap in searches (explorefft and accelsearch)
- 6) Make zaplist (makezaplist.py)
- 7) Make De-dispersion plan (DDplan.py)
- 8) De-disperse (prepsubband)
- 9) Search the data for periodic signals (accelsearch)
- 10) Search the data for single pulses (single_pulse_search.py)
- 11) Sift through the candidates (ACCEL_sift.py)
- 12) Fold the best candidates (prepfold)
- 13) Start timing the new pulsar (prepfold and get_TOAs.py)

Examine the raw data

> readfile GBT_Lband_PSR.fil

```
> readfile GBT Lband PSR.fil
Assuming the data is a SIGPROC filterbank file.
1: From the SIGPROC filterbank file 'GBT Lband PSR.fil':
                  Telescope = GBT
                Source Name = Mystery PSR
            Obs Date String = 2004-01-06T11:38:09
             MJD start time = 53010.48482638889254
                   RA J2000 = 16:43:38.1000
             RA J2000 (dea) = 250.90875
                  Dec J2000 = -12:24:58.7000
            Dec J2000 (deg) = -12.4163055555556
                  Tracking? = True
              Azimuth (deq) = 0
           Zenith Ang (deg) = 0
            Number of polns = 2 (summed)
           Sample time (us) = 72
         Central freq (MHz) = 1400
          Low channel (MHz) = 1352.5
         High channel (MHz) = 1447.5
        Channel width (MHz) = 1
         Number of channels = 96
      Total Bandwidth (MHz) = 96
                       Beam = 1 \text{ of } 1
            Beam FWHM (deg) = 0.147
         Spectra per subint = 2400
           Spectra per file = 531000
      Time per subint (sec) = 0.1728
        Time per file (sec) = 38.232
            bits per sample = 4
          bytes per spectra = 48
        samples per spectra = 96
           bytes per subint = 115200
         samples per subint = 230400
                zero offset = 0
           Invert the band? = False
       bytes in file header = 365
```

- readfile can automatically identify most of the datatypes that PRESTO can handle (in PRESTO v2, though, this is only SIGPROC filterbank and PSRFITs)
- It prints the meta-data about the observation

> rfifind -time 2.0 -o Lband GBT_Lband_PSR.fil

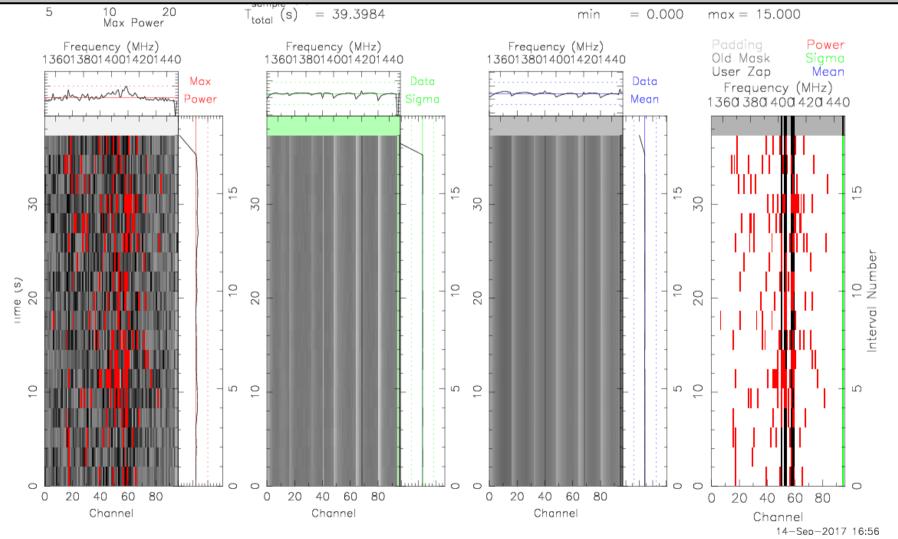
```
> rfifind -time 2.0 -o Lband GBT Lband PSR.fil
               Pulsar Data RFI Finder
                 by Scott M. Ransom
Assuming the data are SIGPROC filterbank format...
Reading SIGPROC filterbank data from 1 file:
  'GBT Lband PSR.fil'
    Number of files = 1
       Num of polns = 2 (summed)
 Center freq (MHz) = 1400
    Num of channels = 96
    Sample time (s) = 7.2e-05
    Spectra/subint = 2400
   Total points (N) = 531000
    Total time (s) = 38.232
    Clipping sigma = 6.000
   Invert the band? = False
         Byteswap? = False
     Remove zeroDM? = False
File Start Spec Samples Padding
                      531000
                                      0 53010.48482638889254
Analyzing data sections of length 28800 points (2.0736 sec).
 Prime factors are: 2 2 2 2 2 2 3 3 5 5
Writing mask data to 'Lband rfifind.mask'.
Writing RFI data to 'Lband rfifind.rfi'.
Writing statistics to 'Lband rfifind.stats'.
Massaging the data ...
Amount Complete = 37%^C
```

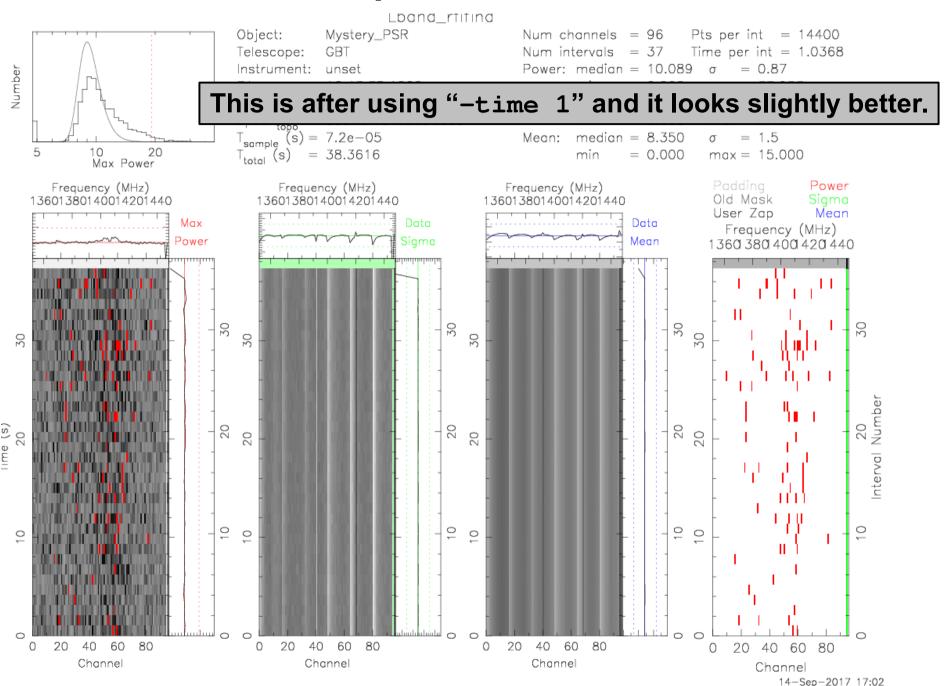
- rfifind identifies strong narrow-band and/or short duration broadband RFI
- Creates a "mask" (basename determined by "-o") where RFI is replaced by median values
- PRESTO programs automatically clip strong, transient, DM=0 signals (turn off using -noclip) Usually a good thing!
- Typical integration times (-time) should be a few seconds
- Modify the resulting mask using "-nocompute -mask ..." and the other rfifind options

```
Writing mask data to 'Lband rfifind.mask'.
Writing RFI data to 'Lband rfifind.rfi'.
Writing statistics to 'Lband rfifind.stats'.
Massaging the data ...
Amount Complete = 100%
There are 31 RFI instances.
Total number of intervals in the data:
  Number of padded intervals:
                                          (5.263%)
  Number of good intervals:
                                   1487
                                         (81.524%)
  Number of bad
                    intervals:
                                    241 (13.213%)
  Ten most significant birdies:
   Sigma
             Period(ms)
                              Freq(Hz)
                                              Number
   6.83
                              86.5644
             11.5521
                                              147
   6.71
             11.6494
                              85.841
                                              170
                              86.0822
   6.68
             11.6168
                                              146
   6.57
             8.76787
                              114.053
   6.53
             11.5844
                              86.3233
                                              145
   6.10
             11.52
                              86.8055
                                              135
   5.96
             11.4881
                              87.0467
                                              107
   5.89
             11.7153
                              85.3588
                                              21
                                              23
   5.88
             11.6823
                              85.5999
10 5.65
             11.7484
                              85.1177
                                              24
  Ten most numerous birdies:
   Number
             Period(ms)
                              Freq(Hz)
                                              Sigma
   493
             34.56
                              28.9352
                                              4.82
   351
             34.8504
                              28.6941
                                              4.75
   280
             17.28
                              57.8704
                                              4.85
   271
             17.3523
                              57.6292
                                              4.80
   180
             17.4252
                              57.3881
                                              4.68
   179
             17.4987
                              57.147
                                              4.67
   170
             11.6494
                              85.841
                                              6.71
   147
             11.5521
                              86.5644
                                              6.83
   146
             11.6168
                              86.0822
                                              6.68
10 145
                              86.3233
             11.5844
                                              6.53
Done.
```

- Check the number of bad intervals. Usually should be less than ~20%
- Most significant and most numbers birdies are listed (to see all, use -rfixwin)
- Makes a bunch of output files including "...rfifind.ps" where colors are bad (red is periodic RFI, blue/green are timedomain statistical issues)
- Re-run with "-time 1" or recompute with "-nocompute" in this case

This is not so great... too much color, and randomly arranged!
Usually we see bad channels or bad time intervals.
Random red color probably means we are masking a bit too much data.





Look for persistent low-level RFI

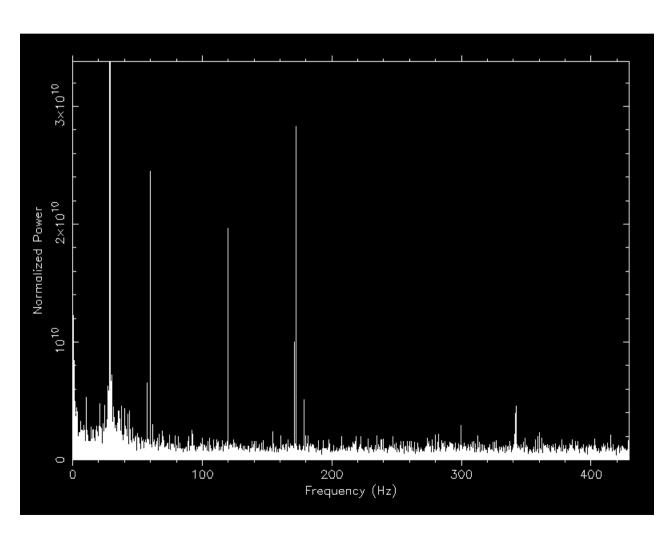
```
> prepdata -nobary -o Lband_topo_DM0.00 \
    -dm 0.0 -mask Lband_rfifind.mask \
    -numout 530000 GBT_Lband_PSR.fil
```

- prepdata de-disperses a single time-series. The "-nobary" flag tells PRESTO not to barycenter the time series.
- If you need to de-disperse multiple time-series, use prepsubband
- Since we will search these data (and FFT them), make sure that the resulting time-series has a "good" (i.e. easily factorable) number of points (-numout)

```
Pulsar Data Preparation Routine
    Type conversion, de-dispersion, barycentering.
                 by Scott M. Ransom
Assuming the data are SIGPROC filterbank format...
Reading SIGPROC filterbank data from 1 file:
  'GBT Lband PSR.fil'
    Number of files = 1
      Num of polns = 2 (summed)
  Center freq (MHz) = 1400
    Num of channels = 96
    Sample time (s) = 7.2e-05
     Spectra/subint = 2400
  Total points (N) = 531000
     Total time (s) = 38.232
     Clipping sigma = 6.000
  Invert the band? = False
          Byteswap? = False
     Remove zeroDM? = False
File Start Spec
                                       0 53010.48482638889254
Read mask information from 'Lband rfifind.mask'
Attempting to read the data statistics from 'Lband_rfifind.stats'...
...succeded. Set the padding values equal to the mid-80% channel averages.
Writing output data to 'Lband topo DM0.00.dat'.
Writing information to 'Lband topo DM0.00.inf'.
Massaging the data ...
Amount Complete = 100%
Done.
Simple statistics of the output data:
             Data points written: 530000
           Maximum value of data: 909.05
          Minimum value of data:
              Data average value: 785.54
         Data standard deviation: 23.12
```

Explore and FFT the time-series

- > exploredat Lband_topo_DM0.00.dat
- > realfft Lband_topo_DM0.00.dat
- > explorefft Lband_topo_DM0.00.fft



- exploredat and explorefft allow you to interactively view a time-series or its power spectrum (for finding RFI)
- changing the power normalization (key 'n') in explorefft is often very helpful
- realfft requires that the time-series is easily factorable (and at least has 1 factor of '2').
 Check using "factor".

Find the periodic interference

> accelsearch -numharm 4 -zmax 0 \
 Lband_topo_DM0.00.dat

Cand		Summed	C-1									
Cand		Sullilleu	Coherent	Num	Period	Frequency	FFT 'r'	Freq De	eriv	FFT 'z'	Accel	
	Sigma	Power	Power	Harm	(ms)	(Hz)	(bin)	(Hz/s	5)	(bins)	(m/s^2)	Notes
1	40.44	840.68	5691.97	2	34.770(8)	28.760(7)	1097.50(25)	0.0000	9(7)	0.0(1.0)	0.0(7.2)x10^3	
2	18.25	195.52	671.15	4	16.6692(9)	59.991(3)	2289.25(13)	0.0000	9(3)	0.00(50)	0.0(1.7)x10 ³	
3	8.24	58.52	54.88	4	5.8484(1)	170.987(3)	6524.88(13)	0.000	9(3)	0.00(50)	0.0(6.0)x10^2	
4	5.28	36.74	64.46	4	5.6024(1)	178.495(3)	6811.38(13)	0.0000	9(3)	0.00(50)	0.0(5.8)x10^2	
			Power	,	Raw	FFT 'r'	Pred 'r'	FFT 'z'	Pred 'z'	Phase	Centroid	Purity
Cand	Harm	Sigma	Loc Po		Power	(bin)	(bin)	(bins)	(bins)	(rad)	(0-1)	< = 1
1	1	100.4	5.05(10)x	10^3	855	1097.4198(54)	1097.50	-0.022(42)	0.00	2.487(10	0.5010(29	1.0085(44
	2	5.84	19.8(6.		21.7	2194.840(89)	2195.00	-0.04(69)	0.00	5.15(16)	0.484(46)	0.990(72)
2	1	12.02	76(12		47.6	2289.254(43)	2289.25	0.25(33)	0.00	5.409(8)		1.030(35)
	2	21.43	234(22		146	4578.509(25)	4578.50	0.50(20)	0.00	5.400(46		1.005(21)
	3	4.09	10.7(4.	6)	11.9	6867.76(12)	6867.75	0.75(87)	0.00	3.73(22)	0.510(62)	1.031(94)
	4	3.23	7.4(3.		6.35	9157.02(15)	9157.00	1.0(1.2)	0.00	2.04(26)	0.419(75)	0.95(12)
3	1	6.98	27.3(7.	4)	26.7	6524.715(77)	6524.88	-0.93(61)	0.00	1.94(14)	0.489(39)	0.974(62)
	2	3.67	9.0(4.	2)	10.7	13049.43(16)	13049.75	-1.9(1.6)	0.00	4.09(24)	0.344(68)	0.80(13)
	3	2.57	5.3(3.	3)	5.98	19574.14(35)	19574.62	-2.8(5.5)	0.00	4.66(31)	0.297(89)	0.49(28)
	4	2.84	6.1(3.	5)	12.4	26098.86(21)	26099.50	-3.7(2.2)	0.00	5.13(29)	0.345(83)	0.74(17)
4	1	6.47	23.7(6.	9)	17.7	6811.388(86)	6811.38	-0.34(72)	0.00	4.98(15)	0.410(42)	0.929(70)
	2	2.86	6.2(3.	5)	4.7	13622.78(16)	13622.75	-0.7(1.3)	0.00	3.87(28)	0.412(82)	0.97(13)
	3	2.58	5.3(3.	3)	4.44	20434.16(23)	20434.12	-1.0(2.5)	0.00	4.48(31)	0.523(89)	0.72(19)
	4	2.94	6.4(3.	6)	7.12	27245.55(15)	27245.50	-1.4(1.2)	0.00	0.32(28)	0.410(81)	1.00(12)

- We "trick" accelsearch into finding periodic interference (it found 4 candidates, with several harmonics in each)
- That information will be used to create a "birds" file
- ".inf" file is human readable ASCII (it is also found in the ACCEL file).

Make a "birds" file

- What the heck is a "birds" file?
 - "birds" are pulsar astronomer jargon for periodic interference that shows up in our power spectra. We usually "zap" them by zeroing them out before we search the power spectrum.
- In PRESTO, a .birds file is a simple ASCII text file with 5 columns
 - The fundamental frequency of the periodic interference in Hz
 - The width of the interference in Hz (power lines RFI at 50 or 60 Hz is often quite wide, but some interference is only a single FFT bin wide)
 - The number of harmonics of the fundamental to zap, and then 0/1
 (no/yes) for whether the width of the harmonics should grow with
 harmonic number and whether the freqs are barycentric or not (e.g. the
 ATNF database freq for a strong pulsar in the data is barycentric)
 - A row starting with a "#" is a comment
 - Here is an example .birds file:

> cat	Lband.bir	ds		
#Freq	Width	#harm	grow?	bary?
1.2	0.02	5	Õ	0
25.0 60.0	0.01	20	0	0
60.0	0.1	5	1	0
100.0	0.02	24	0	0
>				

Make a "birds" file

- Use explorefft and the *ACCEL_0 files to identify the main periodic signals. Since these are DM=0, they are *almost* certainly RFI.
- Edit the .birds file with a text editor
- Given the results of our earlier accelsearch run, here is an example (where I examined the signals with explorefft to check their widths):

```
> cat Lband.birds
#Freq Width #harm grow? bary?
28.760 0.1 2 0 0
60.0 0.05 2 1 0
```

Notes:

- Don't stress out too much over getting a perfect .birds file (especially about high frequency not-too-strong signals they will be smeared out at high DMs). You mainly want to get the really strong stuff, with Fourier powers more than 50 or so.
- Usually I make a .birds file only for a certain type of data (like once for a whole project where the data are all the same) or for really important single pointings.

Convert the "birds" file to a zaplist

- Make an associated ".inf" file for the ".birds" file
 - > cp Lband_rfifind.inf Lband.inf
- Now convert all of the "birds" and harmonics into individual freqs/widths
 - > makezaplist.py Lband.birds
- The resulting "Lband.zaplist" is ASCII and can be edited by hand
- It can also be loaded into explorefft so you can see if you are zapping everything you need (see the explorefft help screen)
- Apply the zaplist using "zapbirds":
- Zapping barycentric time-series requires "-baryv" to convert topocentric RFI freqs to barycentric. Get that by running prepdata or prepfold on raw data (you can ctrl-c to stop them). As an example:
 - > prepdata -o tmp GBT_Lband_PSR.fil | grep Average Average topocentric velocity (c) = -5.697334e-05

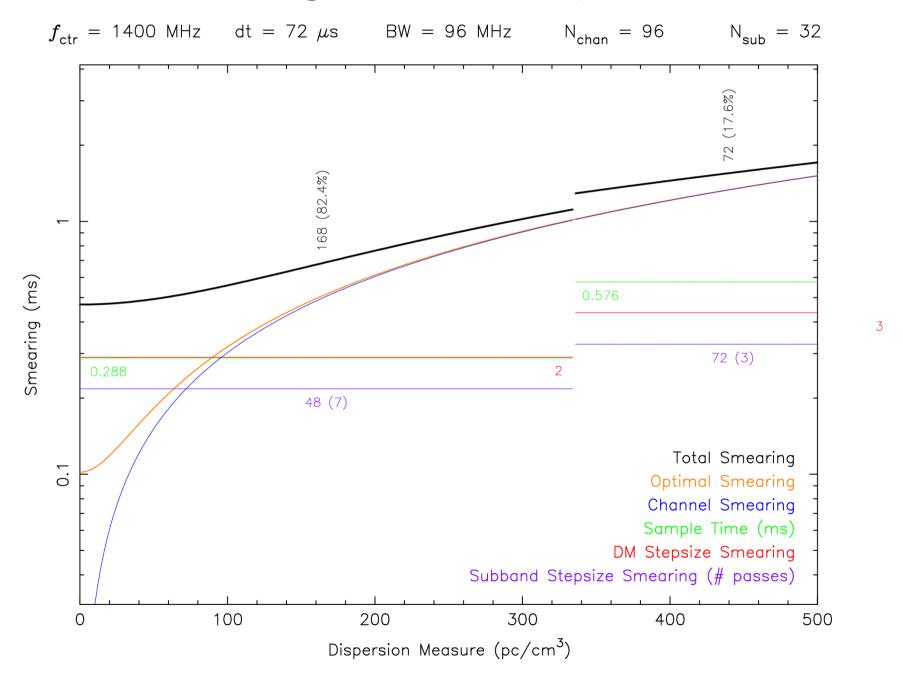
Determining a De-Dispersion Plan

```
> DDplan.py -d 500.0 -n 96 -b 96 -t 0.000072 \
-f 1400.0 -s 32 -r 0.5
```

```
> DDplan.pv -d 500.0 -n 96 -b 96 -t 0.000072 -f 1400.0 -s 32 -r 0.5
Minimum total smearing : 0.102 ms
Minimum channel smearing : 1.51e-05 ms
                                                        "-r" reduces the effective time
Minimum smearing across BW : 0.00145 ms
Minimum sample time
                    : 0.072 ms
                                                                resolution to speed up search
Setting the new 'best' resolution to : 0.5 ms
  Note: ok smearing > dt (i.e. data is higher resolution than needed)
        New dt is 4 \times 0.072 \text{ ms} = 0.288 \text{ ms}
Best guess for optimal initial dDM is 1.984
           High DM
  Low DM
                      dDM DownSamp dsubDM
                                             #DMs DMs/call calls WorkFract
    0.000
         336.000
                      2.00
                                     48.00
                                              168
                                                                    0.8235
  336.000
            552.000
                       3.00
                                    72.00
                                              72
                                                       24
                                                                    0.1765
```

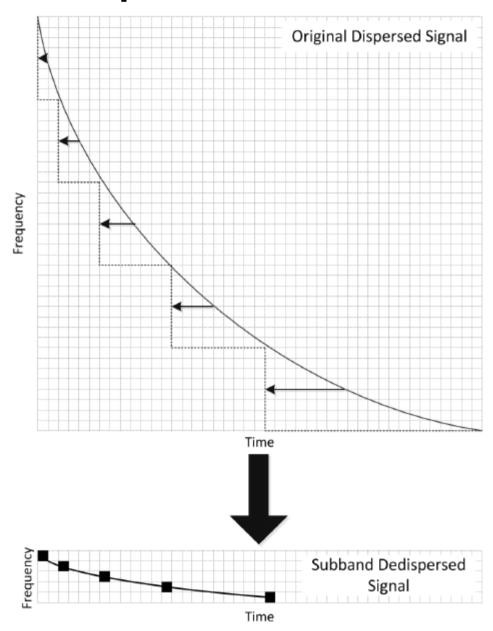
- DDplan.py determines near-optimal ways to de-disperse your data to maintain sensitivity to fast pulsars yet save CPU and I/O time
- Assumes using prepsubband to do multiple-passes through the data using "subband" de-dispersion
- Specify command line information from readfile

Determining a De-Dispersion Plan



Subband De-Dispersion 1

- Incoherent de-dispersion requires you to shift the arrival times of each input channel for a particular DM
- This can be made much quicker by partially shifting groups of channels (subbands) to some nominal DM
- The resulting subband dataset can then be de-dispersed around neighboring DMs with many fewer calculations
- In PRESTO, we do this subband de-dispersion with prepsubband and mpiprepsubband



From Magro and Zarb Adami, MNRAS in press

Subband De-Dispersion 2

```
> prepsubband -nsub 32 -lodm 0.0 -dmstep 2.0 -numdms 24 -numout 132500 -downsamp 4 -mask Lband_rfifind.mask -o Lband GBT_Lband_PSR.fil
```

That command comes from the first call of the first plan line:

```
Low DM High DM dDM DownSamp dsubDM #DMs DMs/call calls WorkFract 0.000 336.000 2.00 4 48.00 168 24 7 0.8235 336.000 552.000 3.00 8 72.00 72 24 3 0.1765
```

- Run prepsubband as many times as there are "calls" in the plan
- Accepted file formats to run prepsubband on are SIGPROC filterbank (".fil") and PSRFITS (".sf" or ".fits")
- If you have a parallel computer (and very long observations), you
 can use the fully parallel program mpiprepsubband to have one
 machine read the data, broadcast it to many other CPUs and then
 each CPU effectively makes a "call"
- The dedisp.py script in \$PRESTO/python can help you automate this process (and generates subbands as well, which can be used to fold candidates and see the DM-curve much faster than by folding raw data). When the file has been edited, do: python dedisp.py

Prepare for Searching the Data

- First we'll clean up this directory but putting the subband files in their own directory and getting rid of the temporary topocentric files
 - > mkdir subbands
 > mv *.sub* subbands/
 > rm -f Lband*topo*
- Use xargs (awesome Unix command) to fft and zap the *.dat files

```
> ls *.dat | xargs -n 1 realfft
> ls *.fft | xargs -n 1 zapbirds -zap \
-zapfile Lband.zaplist -baryv -5.69733e-05
```

- Remember that we can get the barycentric value (i.e. average topocentric velocity) by running a fake prepdata or prepfold command on the raw data:
 - > prepdata -o tmp GBT_Lband_PSR.fil | grep Average
- Now we are ready to run accelsearch on the *.fft files
- If your time series are short (like these), you can use accelsearch to do its own FFTing and zapping by calling it on the ".dat" file. See the -zaplist and -baryv options for accelsearch.

Searching for Periodic Signals

- > accelsearch -zmax 0 Lband_DM0.00.fft
- Accelsearch conducts Fourier-domain acceleration (or not, if zmax=0) searches for periodic signals using Fourier interpolation and harmonic summing of 1, 2, 4, 8 and/or 16.
- "zmax" is the max number of Fourier bins the highest harmonic for a particular search (i.e. fundamental or 1st harm. for a 1 harm. search, 8th harm. for a 8 harm. search) can linearly drift in the power spectrum (i.e. due to orbital motion). Sub-harmonics drift proportionally less (i.e. if 2nd harmonic drifts 10 bins, the fundamental will drift 5).
- The time that the searches take doubles for each additional level of harmonic summing, and is linearly proportional to zmax.
- For MSPs, 8 harmonics is almost always enough. And zmax < 200 or so (beyond that non-linear acceleration start to creep in).
- You can use xargs: ls *.fft | xargs -n 1 accelsearch ...
- For this tutorial data, which is short, you might want to use "-flo 8" so that the rednoise at the very lowest freq bins aren't detected

Sifting the periodic candidates

- > python ACCEL_sift.py > cands.txt
 - ACCEL_sift.py is in \$PRESTO/python and can be edited and tweaked on an observation specific basis
 - It uses several heuristics to reject bad candidates that are unlikely to be pulsars. And it combines multiple detections of the same candidate signals over various DMs (and harmonics as well).
 - The output is a human-readable ranked list of the best candidates
 - ASCII "plots" in the cands.txt file allow you to see rough signal-to-noise versus DM (if there is a peak at DM != 0, that is good)
 - The format for the "candidate" is the candfile:candnum (as you would use them with prepfold)
 - You can also look through the ACCEL files themselves. The ones ending in numbers are human readable (use less -s). Summaries of the candidates are at top and details of their harmonics at bottom.
 - For large single ACCEL files, you can use quick_prune_cands.py

Folding Pulsar Candidates

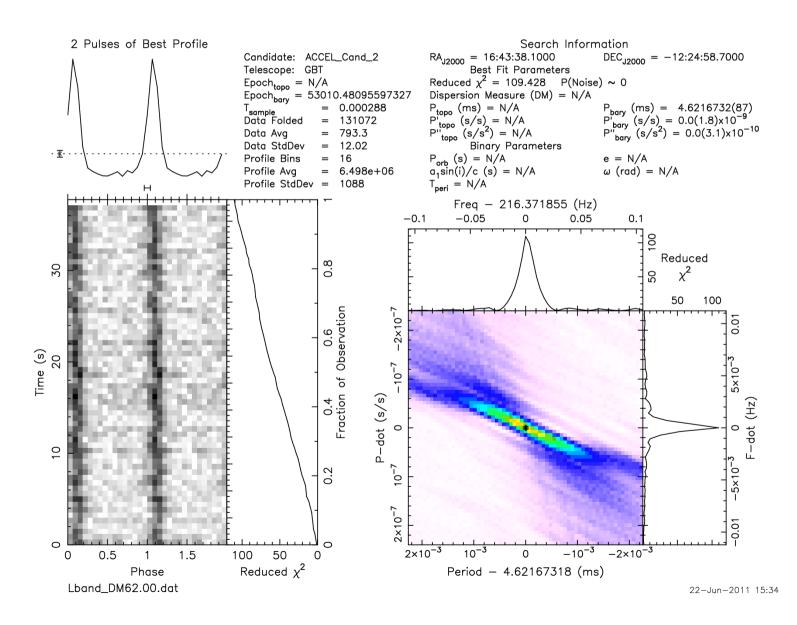
```
> prepfold -accelcand 2 -accelfile \
Lband_DM62.00_ACCEL_0.cand Lband_DM62.00.dat
```

- prepfold can fold time-series (*.dat files), subbands (*.sub?? files),
 or rawdata files. Many ways to specify period (-p) / freq (-f) etc.
- Folding time-series is very fast and is useful to decide which candidates to fold the raw data
- When you fold subbands and/or the raw data, make sure that you specify the DM (and choose the set of subbands with closest DM).
- For modern raw data, using 64 or more subbands (-nsub) is a good idea for folding (to see narrow band RFI and scintillation better)
- If RFI is bad, can zap it using show_pfd or re-fold using -mask

```
> prepfold -dm 62.0 -accelcand 2 -accelfile \
Lband_DM62.00_ACCEL_0.cand \
subbands/Lband_DM72.00.sub??
> prepfold -n 64 -nsub 96 -p 0.004621638 -dm 62.0 \
GBT_Lband_PSR.fil
```

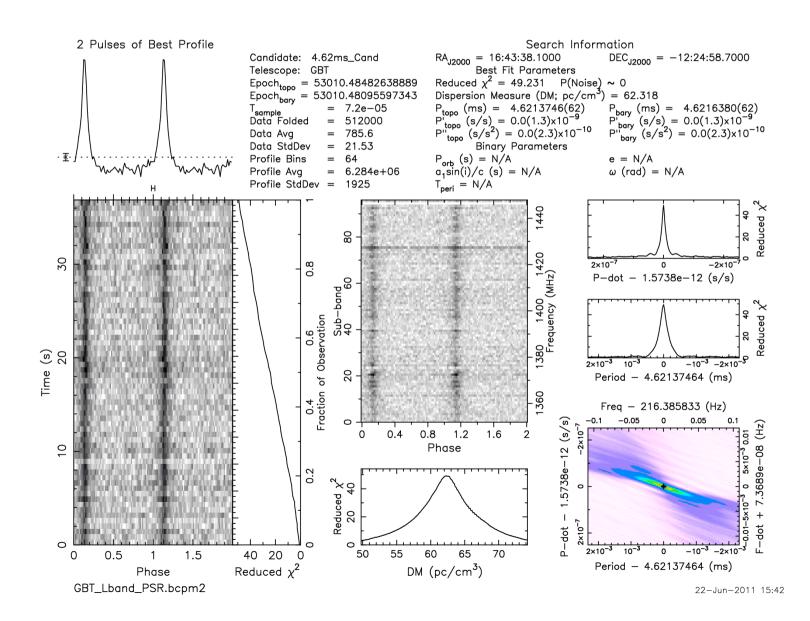
Pulsar! (timeseries)

> prepfold -accelcand 2 -accelfile \
Lband_DM62.00_ACCEL_0.cand Lband_DM62.00.dat



Pulsar! (raw data)

> prepfold -n 64 -nsub 96 -p 0.004621638 -dm 62.0 \ GBT_Lband_PSR.fil



Searching for Transient Bursts

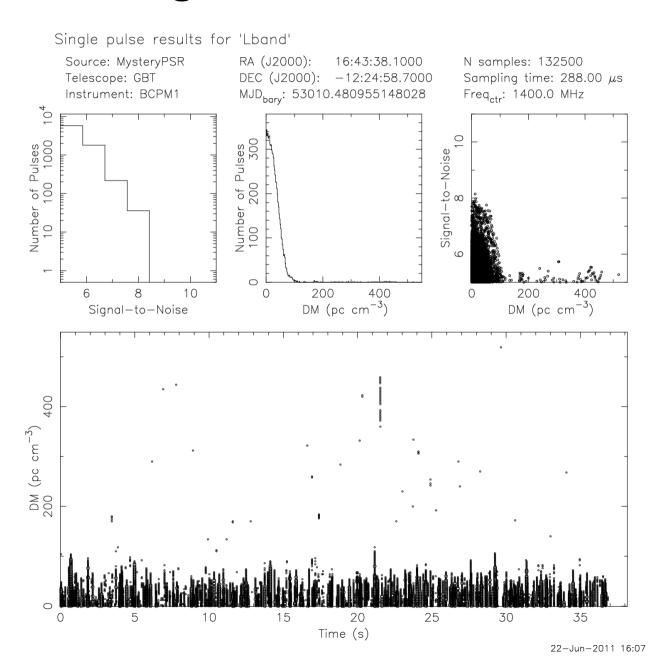
```
> single_pulse_search.py *.dat
```

- single_pulse_search.py conducts matched-filtering single-pulse searches using "boxcar" templates.
- --fast can make things about a factor of 2 faster, but only use it if the data are well-behaved (relatively constant power levels)
- If there are very strong pulses in your data, they can look like RFI. For those cases, turn off bad-block finding (--nobadblocks)
- Generates *.singlepulse files that are ASCII and a single-pulse plot
- Can regenerate a plot using (for instance)

```
> single_pulse_search.py *DM1??.??*.singlepulse
```

• Can choose start and end times as well (--start and --end)

Searching for Transient Bursts



Making TOAs from the discovery obs

- get_TOAs.py needs to be run on a prepfold file of either a topocentric time series or a fold of raw data. The fold must have been made either using a parfile (use -timing) or with the (-nosearch) option.
- The must be either a single gaussian (-g FWHM), an ASCII profile (i.e. a bestprof file from prepfold) or a multi-gaussian-template (derived using pygaussfit.py: "-g template.gaussian")
- -n is the number of TOAs (and must factor the number of parts (-npart) from the prepfold file
- ¬s is the number of subband TOAs to generate (1 is default)
 - > get_TOAs.py -g 0.1 -n 20 newpulsar.pfd

Now try it from scratch...

• There is another sample data set (with mystery pulsar) here:

http://www.cv.nrao.edu/~sransom/Parkes_70cm_PSR.fits

Command history for this tutorial can be found here:

http://www.cv.nrao.edu/~sransom/GBT_Lband_PSR_cmd_history.txt

Let me know if you have any problems or suggestions!

Scott Ransom <sransom@nrao.edu>