RFI Mitigation

(what it is, where it comes from, how to fix it)

or,

"Finding an astrophysical needle in an artificial haystack."

MPIfR, 22/01/2020

What is RFI?

RFI (Radio Frequency Interference) in a nutshell:

- **All the annoying signals that get in the way of you seeing your** `[insert astronomical source]`
  - Pulsar
  - RRAT
  - FRB
  - Etc…

But what is it really?

To answer completely, start with *"What are we looking at when we make an observation?"*

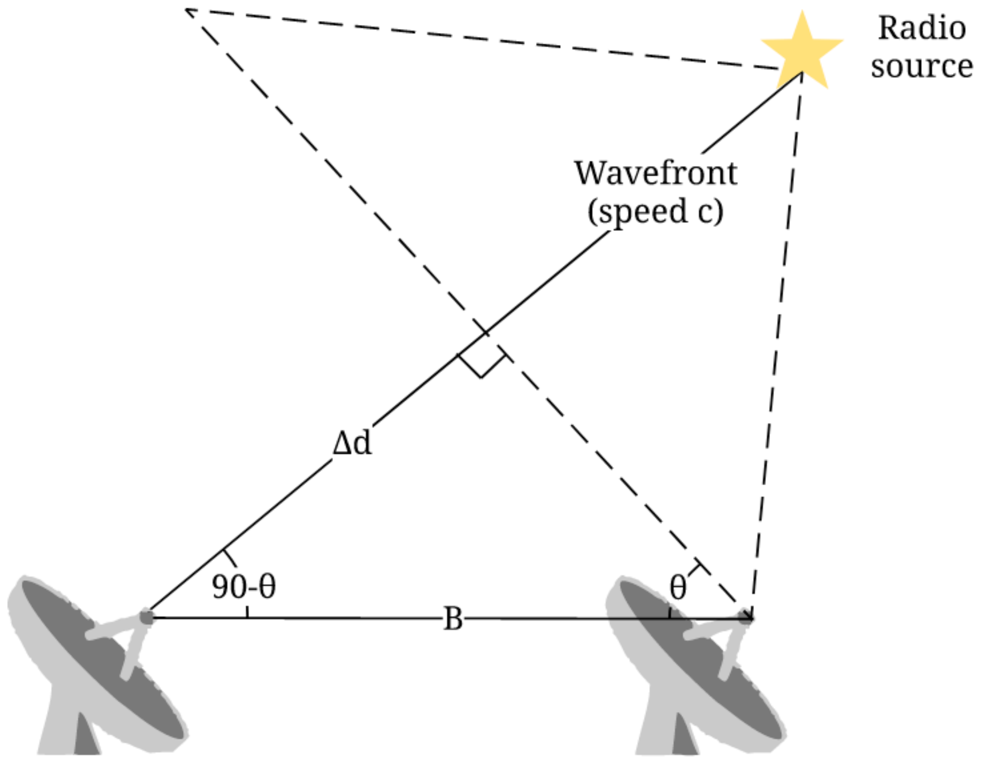
Two answers…

Romantic answer:

- The stars. The galaxies. The Universe!

Technical answer:

- **Voltage fluctuations** associated with **electromagnetic radiation** (radio waves) which induces
  currents in a conductor (our antenna)



And what do we want to see?

(in our *"Voltage fluctuations associated with EM blah blah blah… ?"*)

Again, two answers…

Romantic answer:

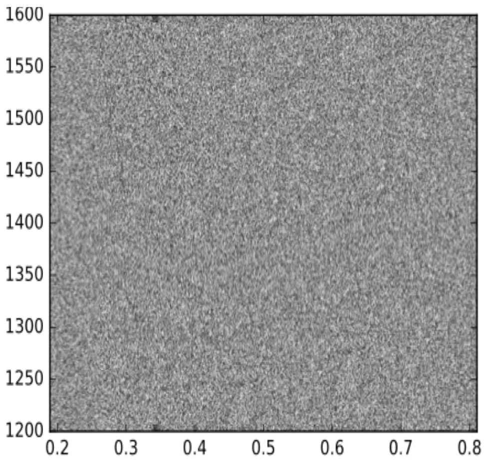
- The stars! Probably, the neutron stars.

Technical answer:

- **Signals** from our astrophysical radio **sources** against a background of (preferably low) **Gaussian noise**

Gaussian bit is important!

The radiometer equation defines the sensitivity of a receiver to radio signals

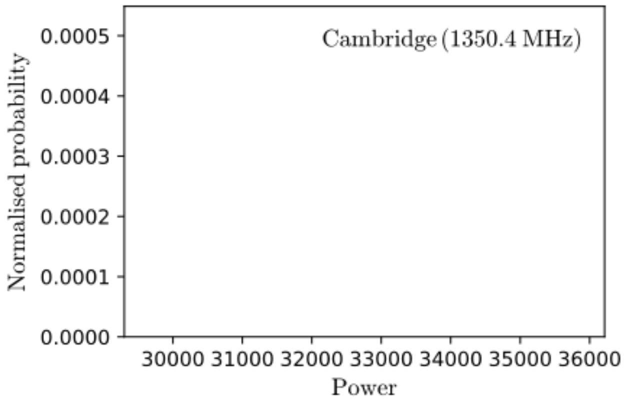


Central limit theorem states...

- *"When independent random variables are added, their properly normalised sum tends towards a gaussian distribution"*

In radio astronomy...

- Our independent random variables are a set of random data in our receiver from many weak radio signals and noise voltages [1]
- Plots are what you want to see: nice random noise



This is important because: The radiometer equation:

- Defines sensitivity to radio signals*
- Fundamental assumption: noise fluctuations behave randomly

$$S_{min} = \beta \frac{(S/N_{min})SEFD}{\sqrt{n_p \tau \Delta \upsilon}}$$

RFI is non-Gaussian

- Reduces sensitivity to signals
- Could obscure them entirely

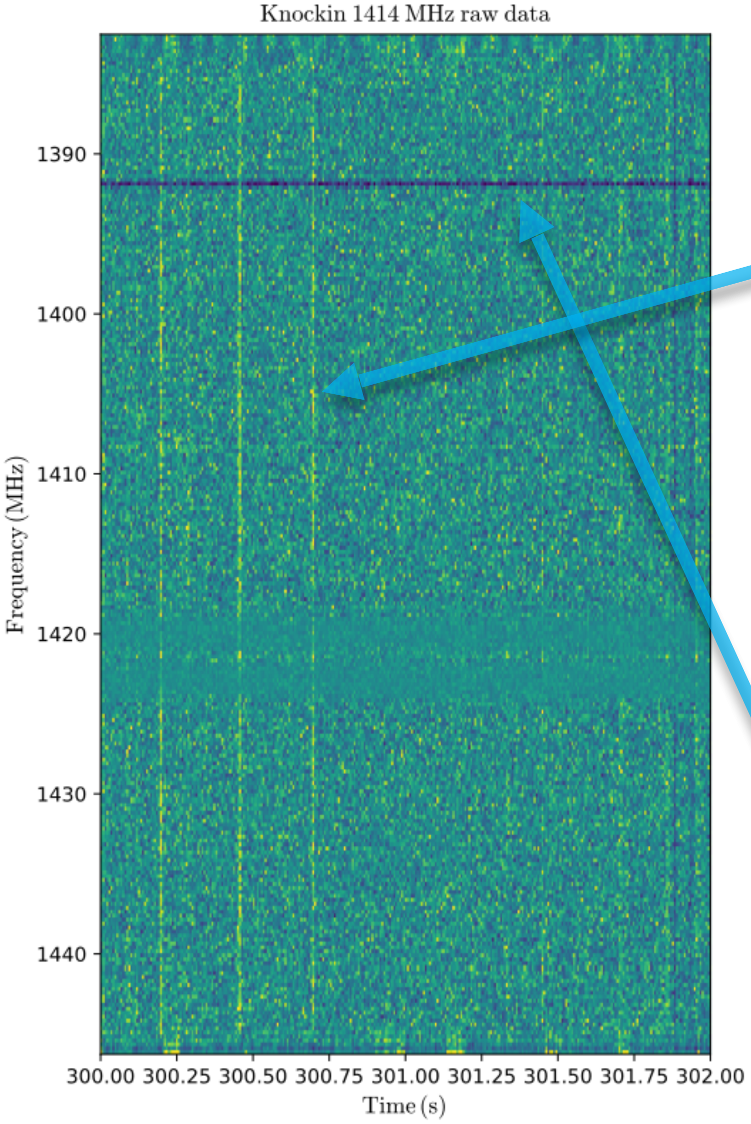
*Derivation found in [2]

So what is RFI???

- In essence, it is any ***unwanted*** non-Gaussian signal in your data
  - Astrophysical signals will be non-Gaussian too

Examples of non-Gaussian signals

Data taken using the e-MERLIN interferometer



Broad-band impulsive signals

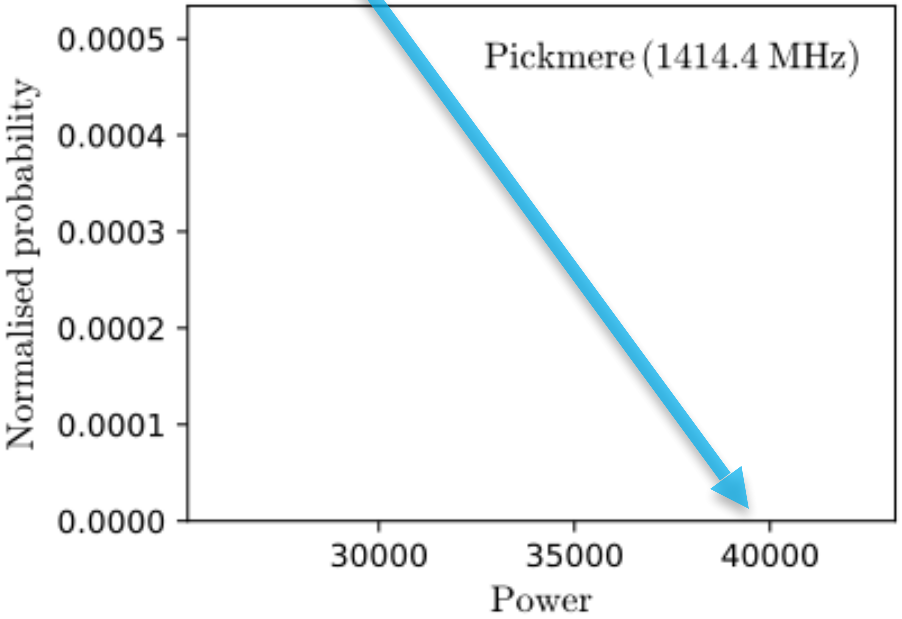
- Span entire observing bandwidth
- Something periodic?

Narrow-band persistent signals

- Span entire observation
- Something emitting at this frequency?

Data clearly not Gaussian

- E.g. saturated samples -- extra power



What causes RFI?

Unfortunately…

- All sorts of modern items of convenience
  - Airport radar
  - Petrol cars
  - Microwaves
  - Satellites
  - Mobile phones et al.…

In fact…

- Most of the radio spectrum is actively allocated to technology
  - **Next slide**: United States radio frequency allocations as of 2016 [4]
  - **Yellow** is protected for radio astronomy

UNITED STATES FREQUENCY ALLOCATIONS

THE RADIO SPECTRUM

RADIO SERVICES COLOR LEGEND

AERONAUTICAL MOBILE
INTER-SATELLITE
RADIO ASTRONOMY
AERONAUTICAL MOBILE SATELLITE
LAND MOBILE
RADIODETERMINATION SATELLITE
AERONAUTICAL RADIONAVIGATION
LAND MOBILE SATELLITE
RADIOLOCATION
AMATEUR
MARITIME MOBILE
RADIOLOCATION SATELLITE
AMATEUR SATELLITE
MARITIME MOBILE SATELLITE
RADIONAVIGATION
BROADCASTING
MARITIME RADIONAVIGATION
RADIONAVIGATION SATELLITE
BROADCASTING SATELLITE
METEOROLOGICAL
SPACE OPERATION
EARTH EXPLORATION SATELLITE
METEOROLOGICAL SATELLITE
SPACE RESEARCH
FIXED
MOBILE
STANDARD FREQUENCY AND TIME SIGNAL
FIXED SATELLITE
MOBILE SATELLITE
STANDARD FREQUENCY AND TIME SIGNAL SATELLITE

ACTIVITY CODE

FEDERAL EXCLUSIVE
FEDERAL/NON-FEDERAL SHARED
NON-FEDERAL EXCLUSIVE

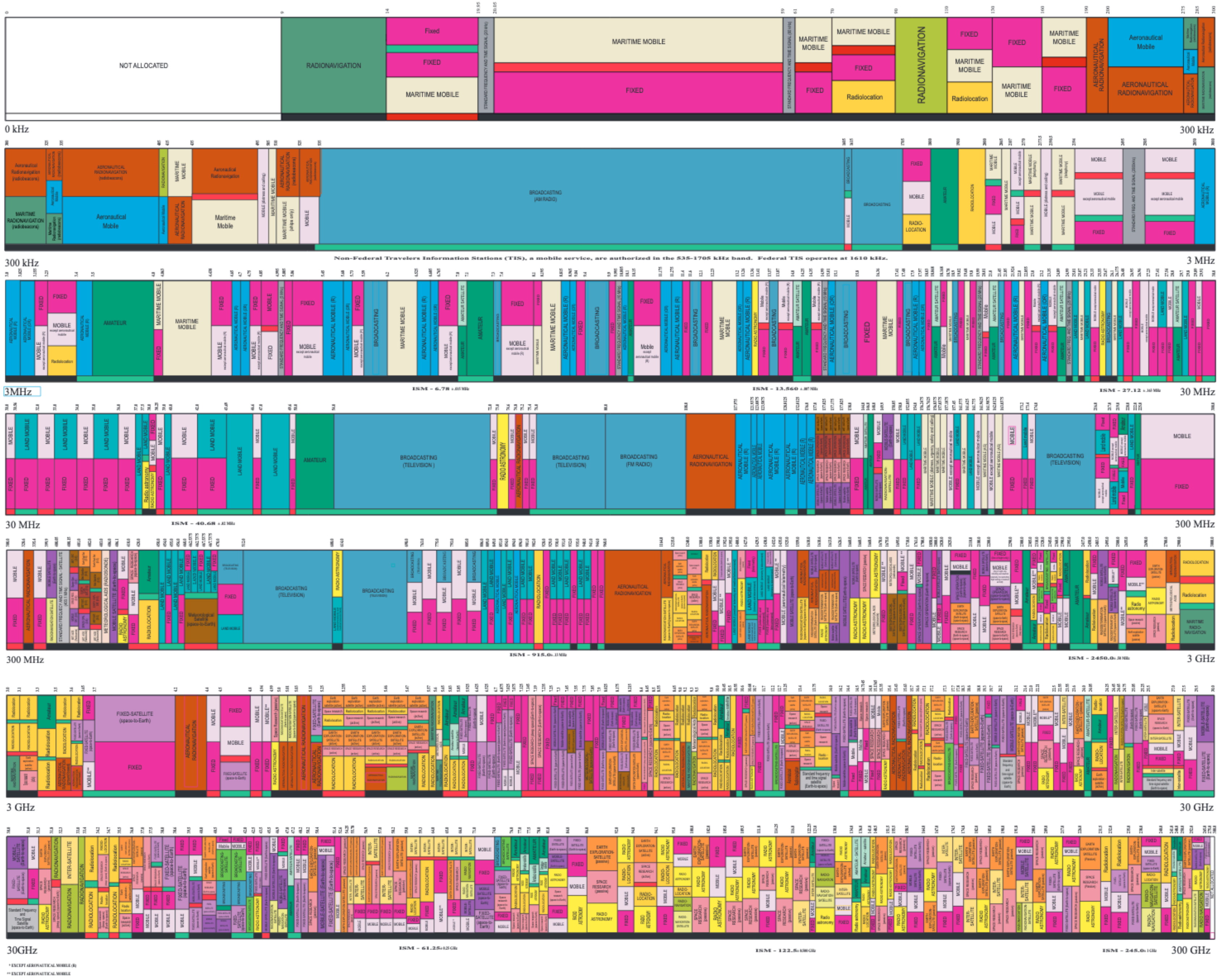
ALLOCATION USAGE DESIGNATION

SERVICE	EXAMPLE	DESCRIPTION
Primary	FIXED	Capital Letters
Secondary	Mobile	1st Capital with lower case letters

This chart is a graphic single-point-in-time portrayal of the Table of Frequency Allocations used by the FCC and NTIA. As such, it may not completely reflect all aspects, i.e. footnotes and recent changes made to the Table of Frequency Allocations. Therefore, for complete information, users should consult the Table to determine the current status of U.S. allocations.

U.S. DEPARTMENT OF COMMERCE
National Telecommunications and Information Administration
Office of Spectrum Management

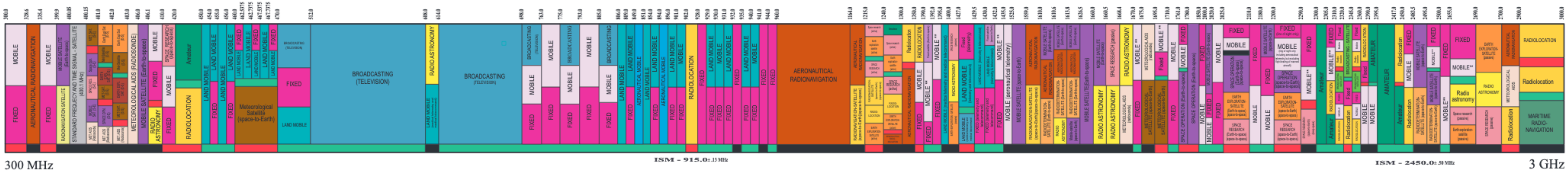
JANUARY 2016



For sale by the Superintendent of Documents, U.S. Government Printing Office

Enhance! (300 MHz – 3 GHz)

This figure doesn't need to be readable to show that the spectrum is a mess…

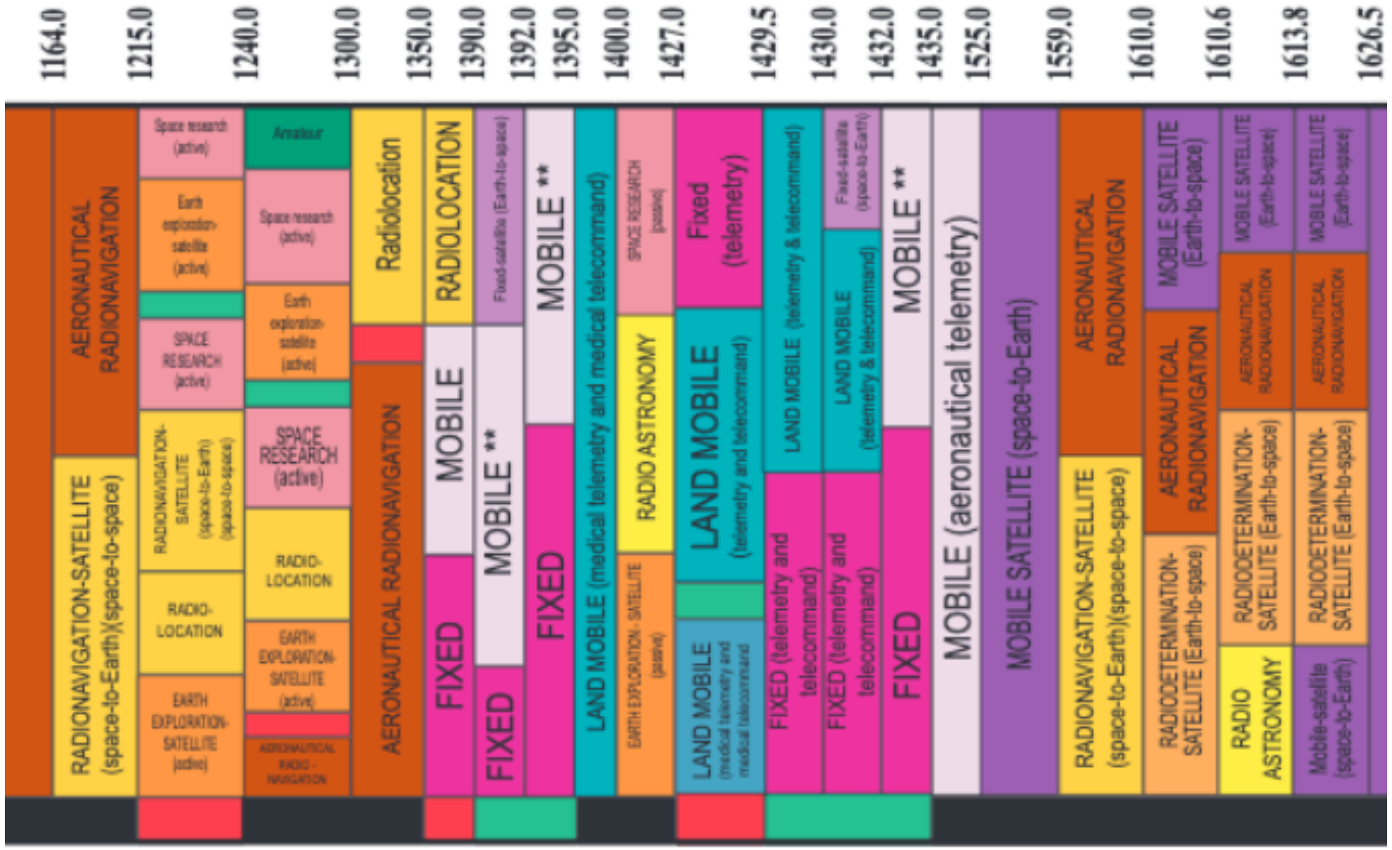


- Still very complicated
- Note lack of yellow…

Closer!

Zoom in to 1.4 GHz

- ~1.4 GHz is the 21cm Hydrogen line
- This is where many people observe pulsars
- Enemies on all sides!



A few choice examples…

Two RFI emitters:

- Airport radar
- Microwave ovens

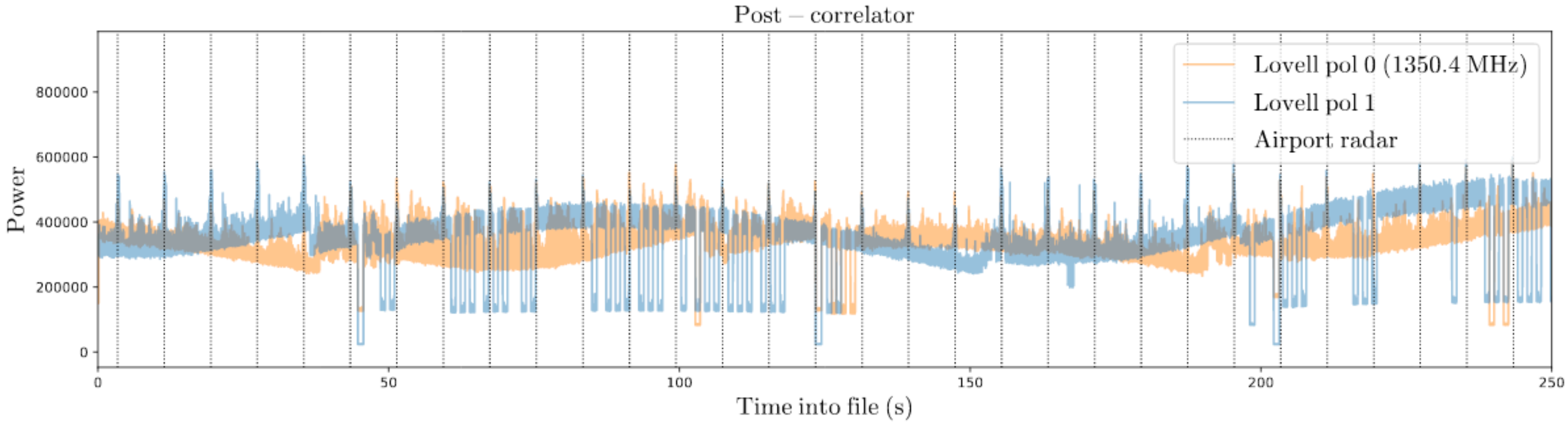
Why these?

- Some artificial signals can mimic astrophysical phenomena…

Airport radar

Observation at Jodrell Bank Observatory

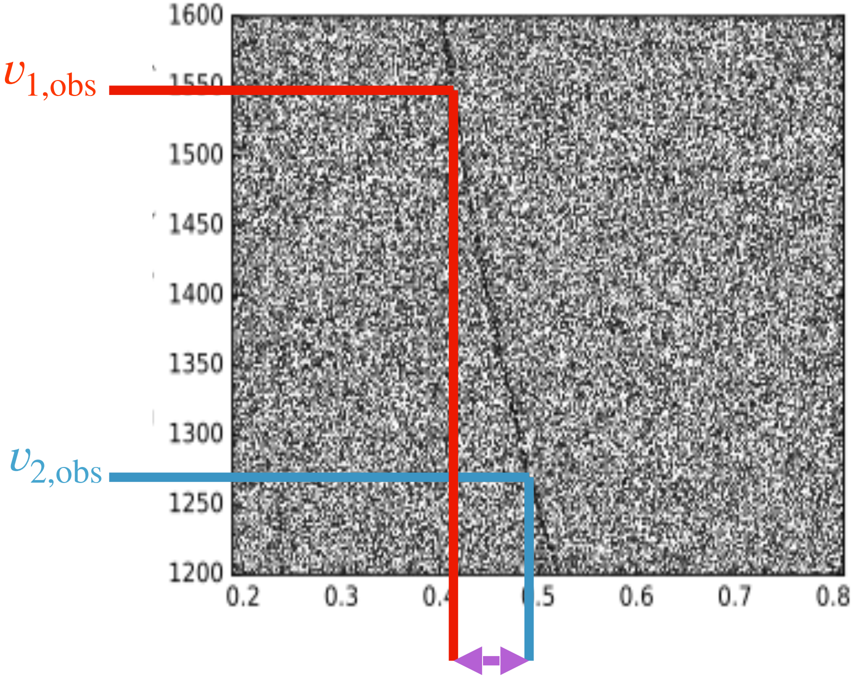
- Early e-MERLIN high time-resolution observations
- Data collapsed to timeseries
  - Spikes every 8 seconds (see dotted lines)
  - Periodic peaks in data… pulsar!?
  - Same cadence as local airport radar



Microwave ovens

Some weren't sure if FRBs were real for a while…

- Near the beginning of my PhD there were <10 FRBs
  - None had been localised
- There were also perytons
  - Frequency-swept signals
  - Nearly followed same frequency-time relation as FRBs…
  - …but not quite. They were artificial.
- Some wondered whether FRBs were real
  - Dr. Emily Petroff discovered the difference [3]
  - **Found**: If you open some microwave ovens before they're done cooking, the cyclotron emits a peryton
  - **Proved**: FRBs were a different population



$$\Delta t_{obs} \propto \mathrm{DM}_{obs} \times (\upsilon_{1,obs}{}^{-2} - \upsilon_{2,obs}{}^{-2}$$

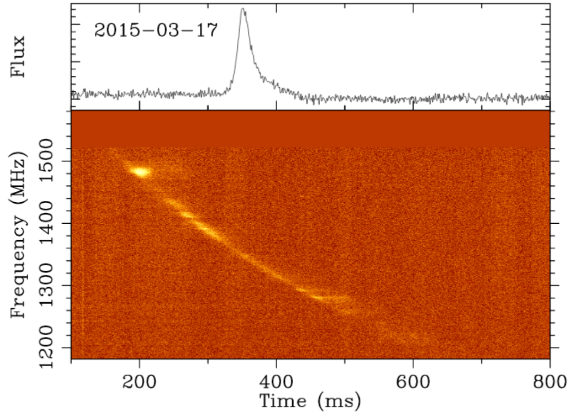


Figure 5. One of the bright perytons generated during the test on 17 March with DM = 410.3 cm^{-3} pc. RFI monitor data at the time of this peryton is shown in Figure 3.

RFI's effects on data

What we've seen so far…

- You've seen how RFI messes with data statistics
- RFI can mimic astrophysical signals
- How does it affect us finding real things?

Answer:

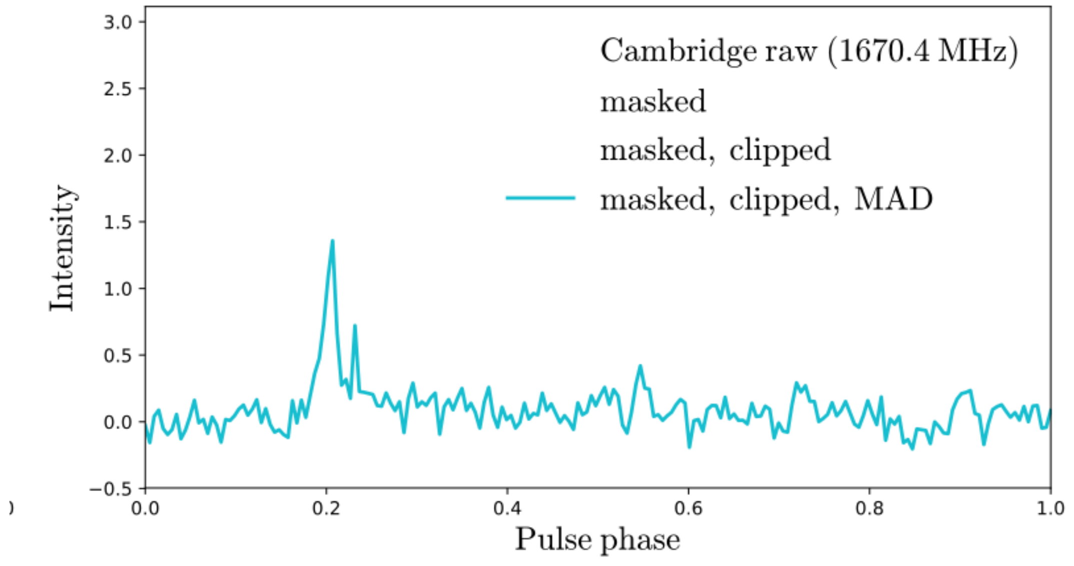
- It's not always pretty…

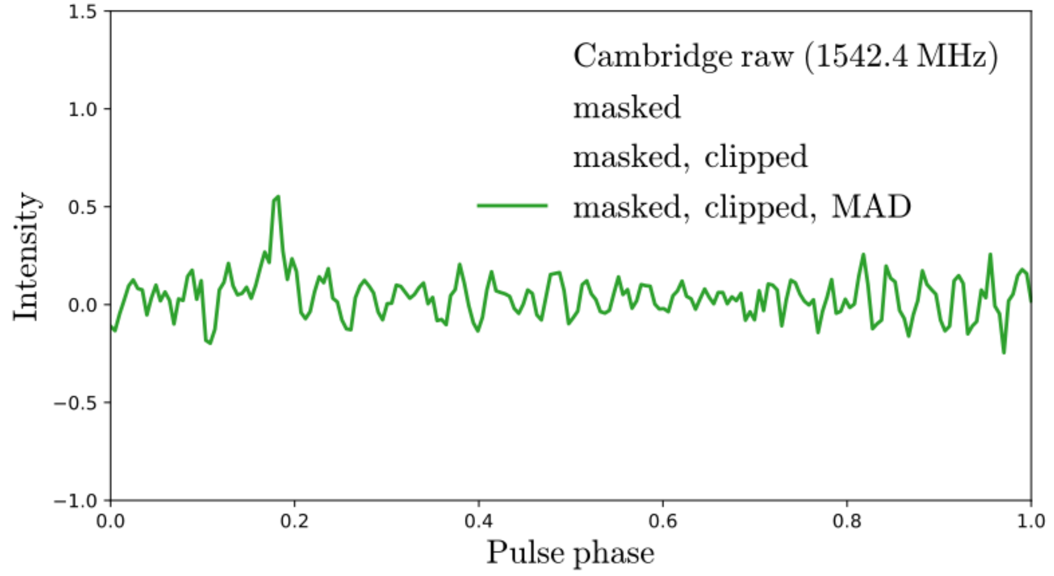
Pulse profile example

When observing pulsars…

- Observations of PSR B0329+54
  - At 1670.4 MHz and 1542.4 MHz
- Data taken using e-MERLIN's Cambridge dish
  - 64 MHz bandwidth, 10-minute observations
- **Thick faint lines**: folded pulse profile *before* RFI removal
  - Blue: Structure reduces pulsar S/N
    - Indicates periodic RFI
  - Green: No structure, but pulse profile invisible!
    - You wouldn't see this pulsar in a blind search

RFI can get in the way of you finding pulsars.



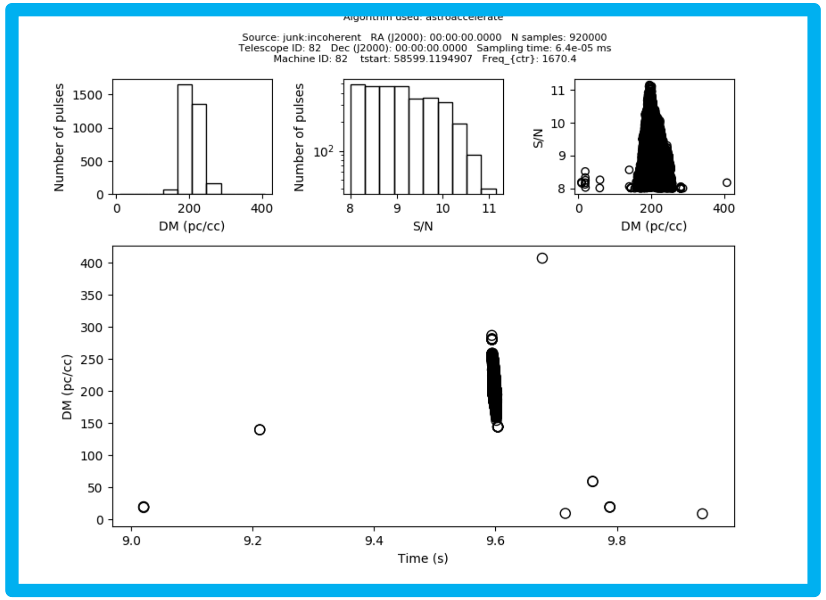


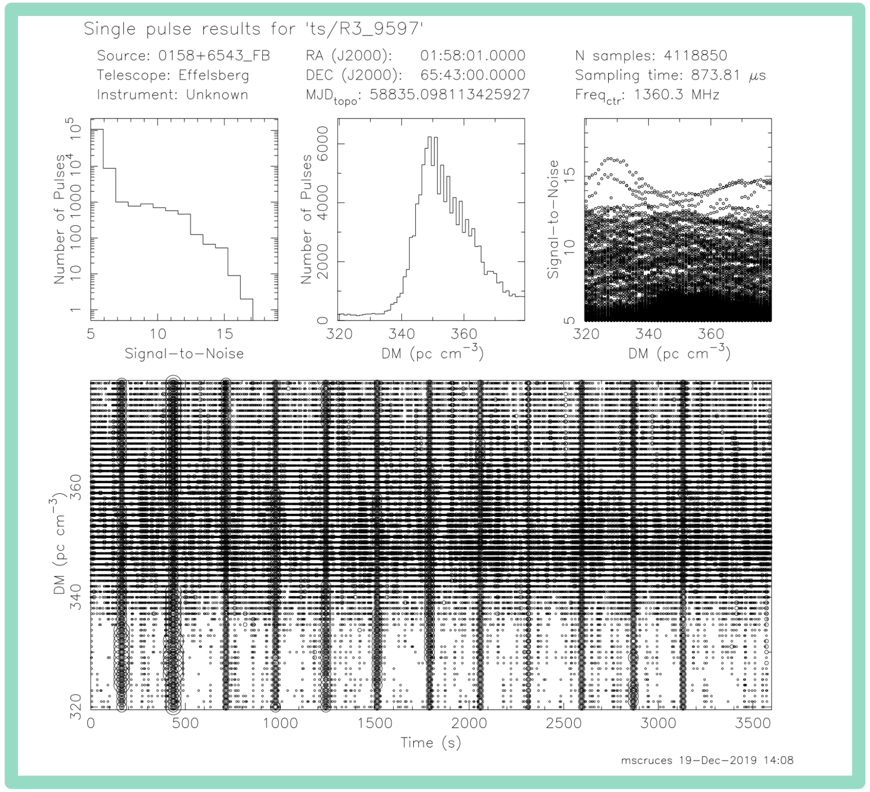
Story is the same for single pulses…

When searching for RRATs, giant pulses, FRBs…

- **Top plot:**
  - Typical single-pulse-search candidate file
    - Contains a DM=200pc/cc pulse at 9.6s
  - Main plot: DM dispersion trial vs Time
  - Subplots: #pulses vs DM, #pulses vs S/N, S/N vs DM trial
    - Note: behaviour of S/N vs DM plot
    - Shape indicitave of signal dedispersed around true DM
  - Not much RFI here…

- **Bottom plot:**
  - Candidate file with lots of RFI
  - Difficult to see a pulse in here.





RFI is bad. So can we fix it?

Two options…

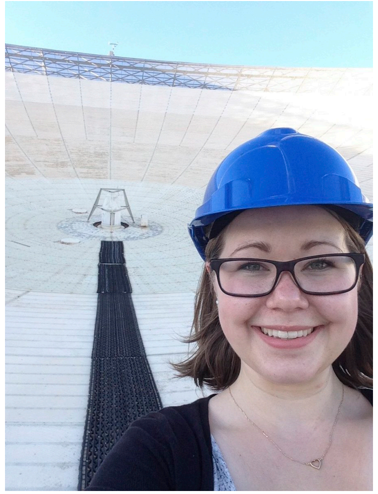
1. Fix the cause
2. Fix the symptoms

Ideally: Treat the cause

This solves the problem for everyone, forever

1. Easy things
   - Switch off phones at a radio observatory
   - Put your microwave in a Faraday cage

2. Be a detective!
   - Emily Petroff: Used Parkes to discover perytons were coming staff lunch room at noon… microwaves! [3]
   - Andrew Lyne: searched JBO with a little antenna to triangulate the location RFI… it was a defective lamp



3. Consult with your telescope engineers!
   - Work together

4. Sometimes this doesn't solve your issue.
   - Archival data
   - Unknown/unfixable cause of RFI

Mostly: Treat the symptoms

Sometimes this is all you can do

The method boils down to:

1. Identify the contaminated data
2. Supress, discard, or replace with something better
   - (e.g. Gaussian noise)

There is lots of literature!

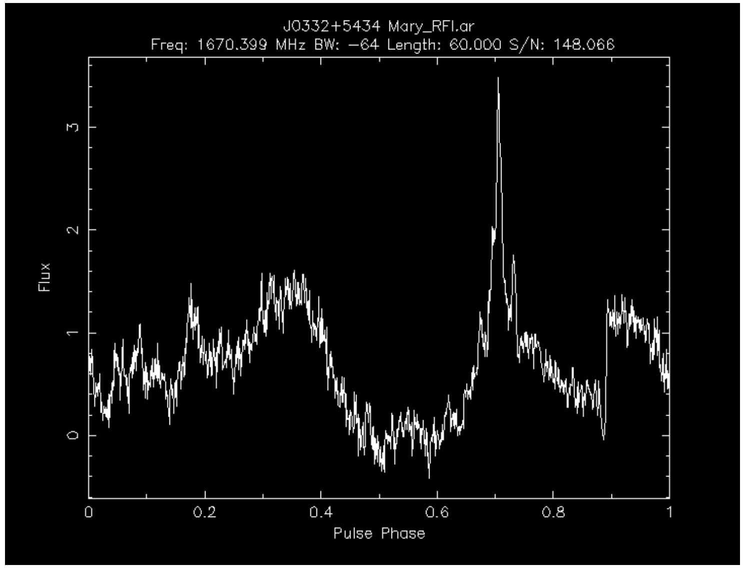
- Many different techniques…
- Many software packages…
- Each implemented differently...

RFI Mitigation examples

Let's imagine a scenario…

1. You've observed a known pulsar.
2. You've folded* your filterbank data.
3. You plot your data with `pav`.
4. You see this:

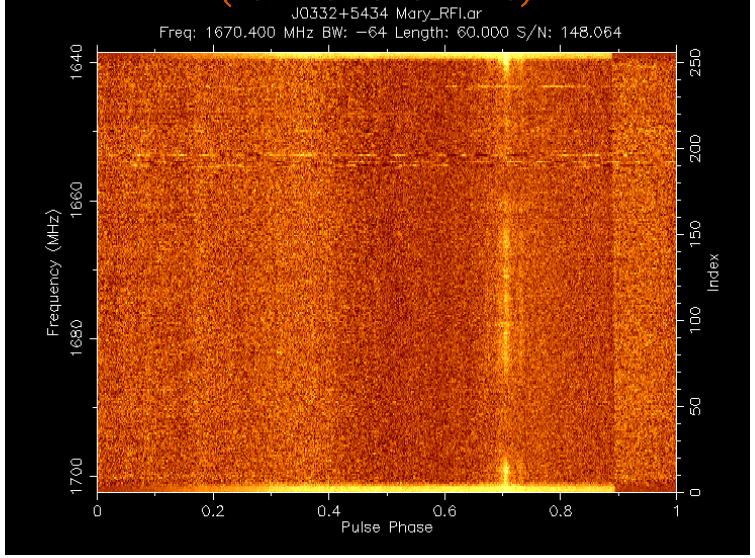
Pulse profile
(scrunch over time, frequency, polarisation)



Command:

`pav -DFTp filename.ar`

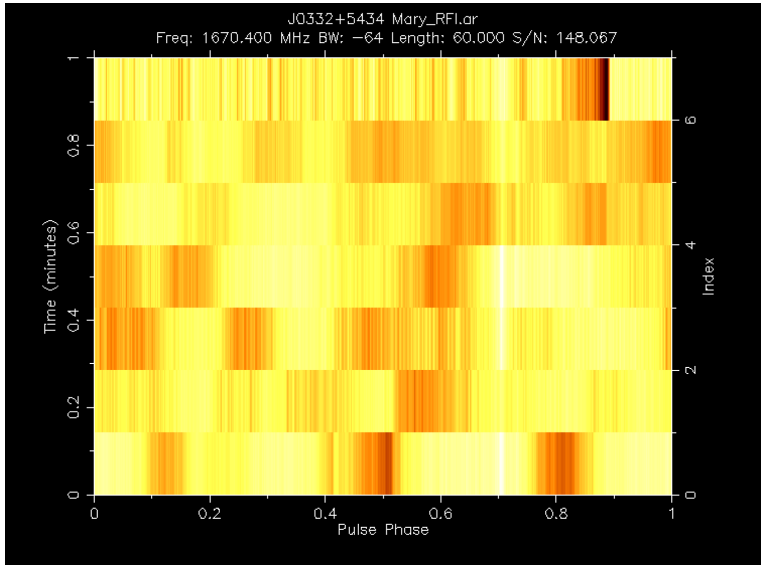
Pulse phase vs frequency
(scrunch over time)



Command:

`pav -GTd filename.ar`

Pulse phase vs time
(scrunch over frequency)



Command:

`pav -YFd filename.ar`

*e.g. using dspsr [11] [12]

Frequency structure

- Something's emitting here

Bright samples

- Saturated data?

Wobbly baseline

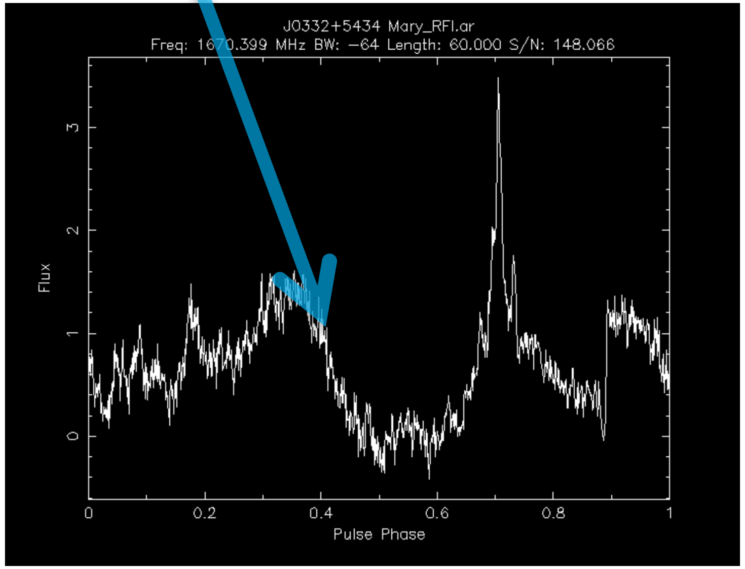
- Lowers S/N

… a mess.

- Can't see anything

Pulse profile

(scrunch over time, frequency, polarisation)

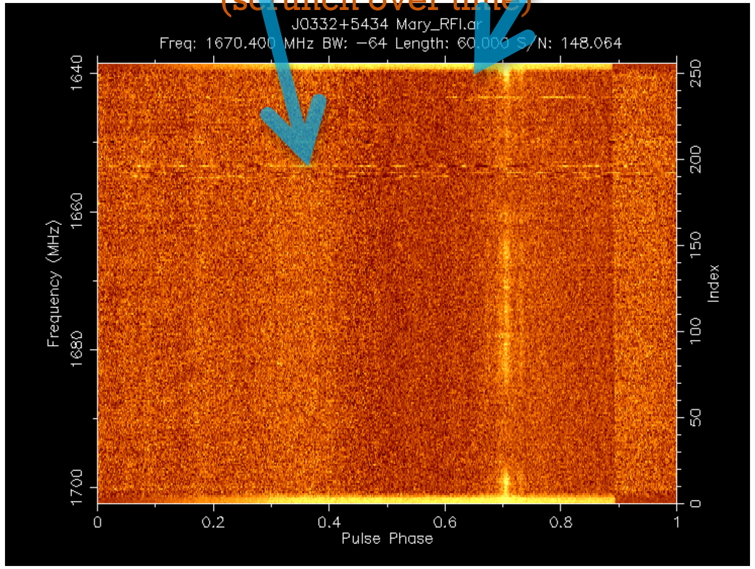


Command:

```
pav -DFTp filename.ar
```

Pulse phase vs frequency

(scrunch over time)

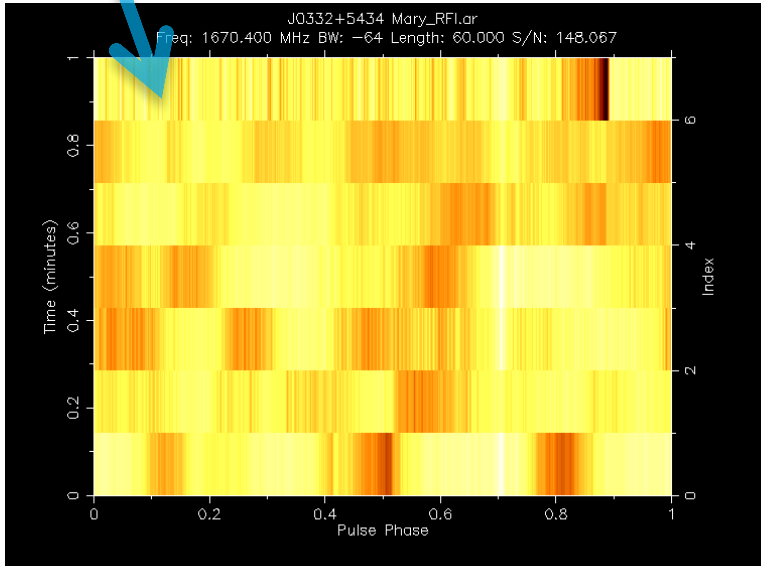


Command:

```
pav -GTd filename.ar
```

Pulse phase vs time

(scrunch over frequency)



Command:

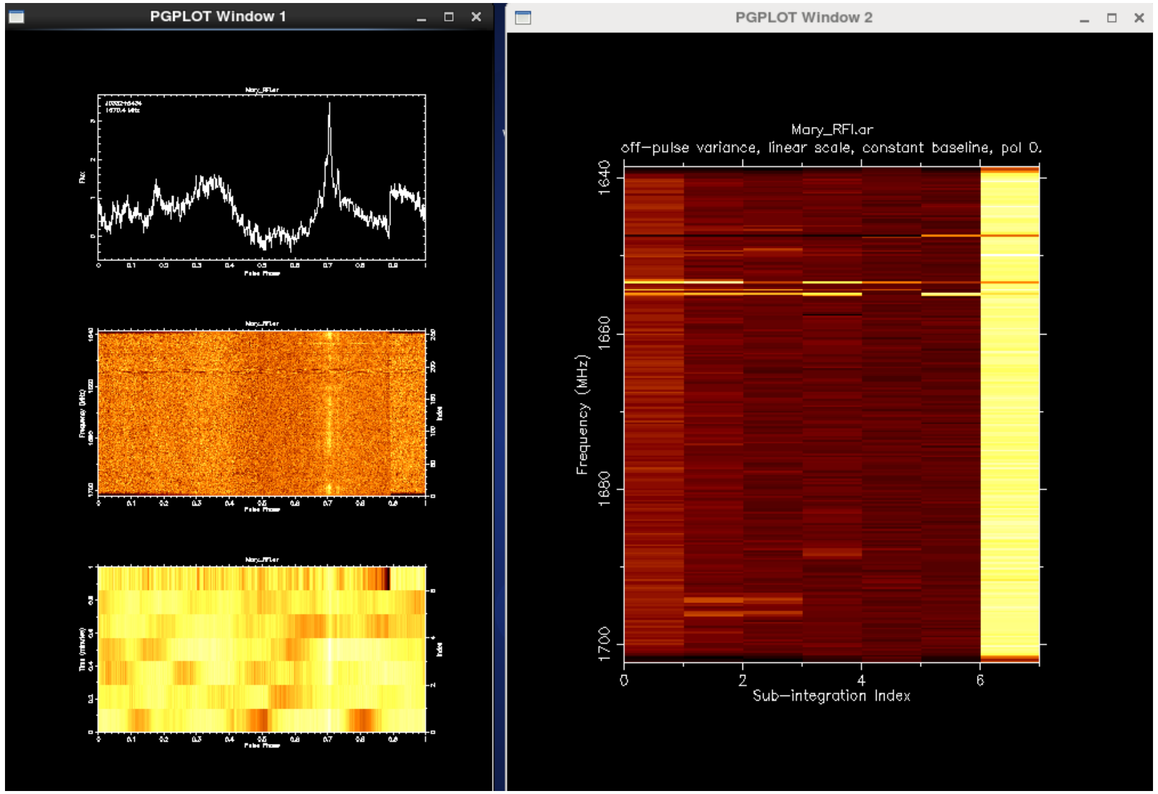
```
pav -YFd filename.ar
```

Example 1: **psrzap**
Flag your bad data by hand

- You can zap some of the RFI by hand using the PSRCHIVE `psrzap` tool [5][6]

- Command: `psrzap filename.ar` displays:
  a. Sub-integration vs frequency
  b. Pulse profile
  c. Pulse phase vs time
  d. Pulse phase vs frequency

Start zapping frequency channels and times to try to improve S/N

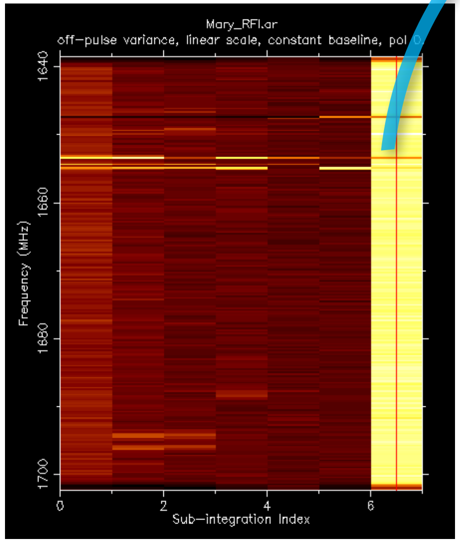


psrzap commands

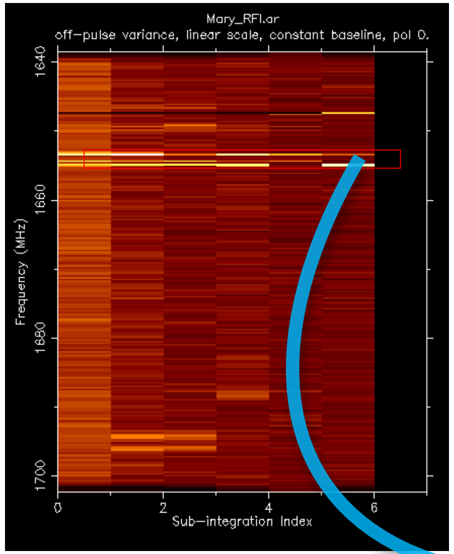
How to flag your bad data by hand

1) On the command line:

t
(for selecting time bins)



f
(for selecting frequency bins)



2) Right click:

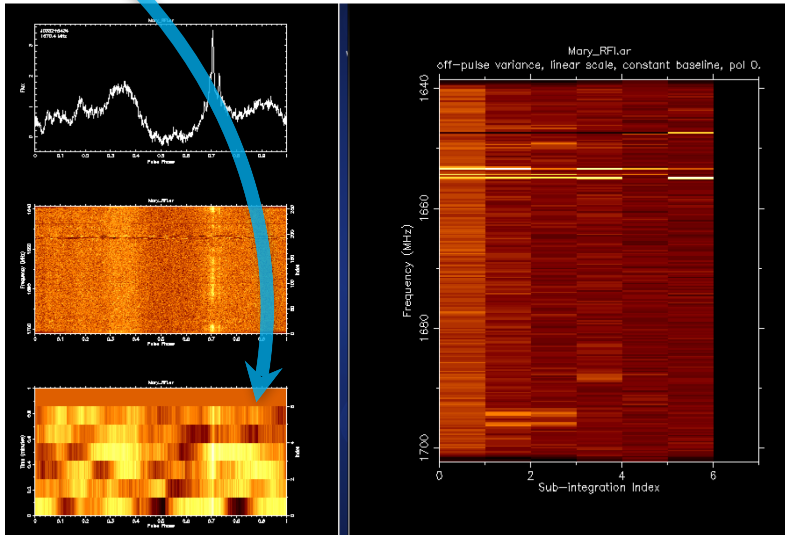
(start selection)

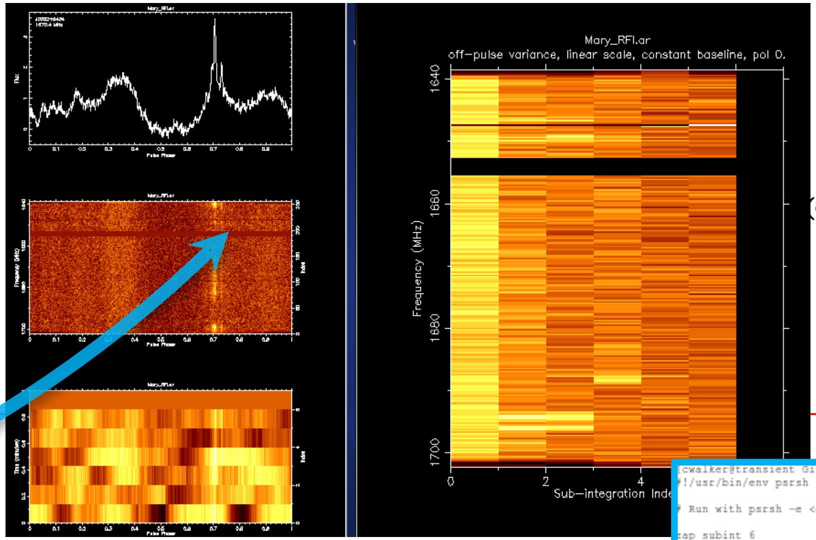
3) Move mouse

Right click
(end selection,
zoom)

4) Left click

(end selection, zap)





5) On the command line:

d
(update plots)

r
(reset zoom, update plots)

s
(save as zapped .ar file filename.zap)

w
(generate psrh script)
(useful for getting exact zapped channels/times)
(e.g. if you want to make a mask)

q
(quit)

```
[cwalker@transient GitHub_ClipPy_repo]$ more Mary_RFI.psh
#!/usr/bin/env psrsh

# Run with psrsh -e <ext> <script>.psh <archive>.ar

zap subint 6
zap chan 188-198
[cwalker@transient GitHub_ClipPy_repo]$
```

Example 2: Masking

If you know your bad channels, why not mask them from your entire dataset?

1. Identify bad channels
   - e.g. using psrzap's `w` command,
   - Or PRESTO's `rfifind` + `rfifind_stats.py` programs [7][8]
2. Save bad channels to list
   - Many search software can zap automatically from a list
   - Make sure list matches your search software's conventions!
     - Are channels zero indexed?
     - Does lowest number = lowest channel?
3. Include the list when running your search software
   - PRESTO's `prepsubband` : `-ignorechan`
   - sigproc's `dedisperse`* : `-filename`

Channel-zapping conventions for frequently used software

The list indexing is…

The lowest number represents the…

	Lowest channel	Highest channel
Zero-indexed	• PRESTO	
One-indexed		• dedisperse • SEEK • DESTROY

*`dedisperse` performs dedispersion, creating a timeseries for single-pulse searching with sigproc's `SEEK` [13][14] or Evan Keane's `DESTROY` [15]

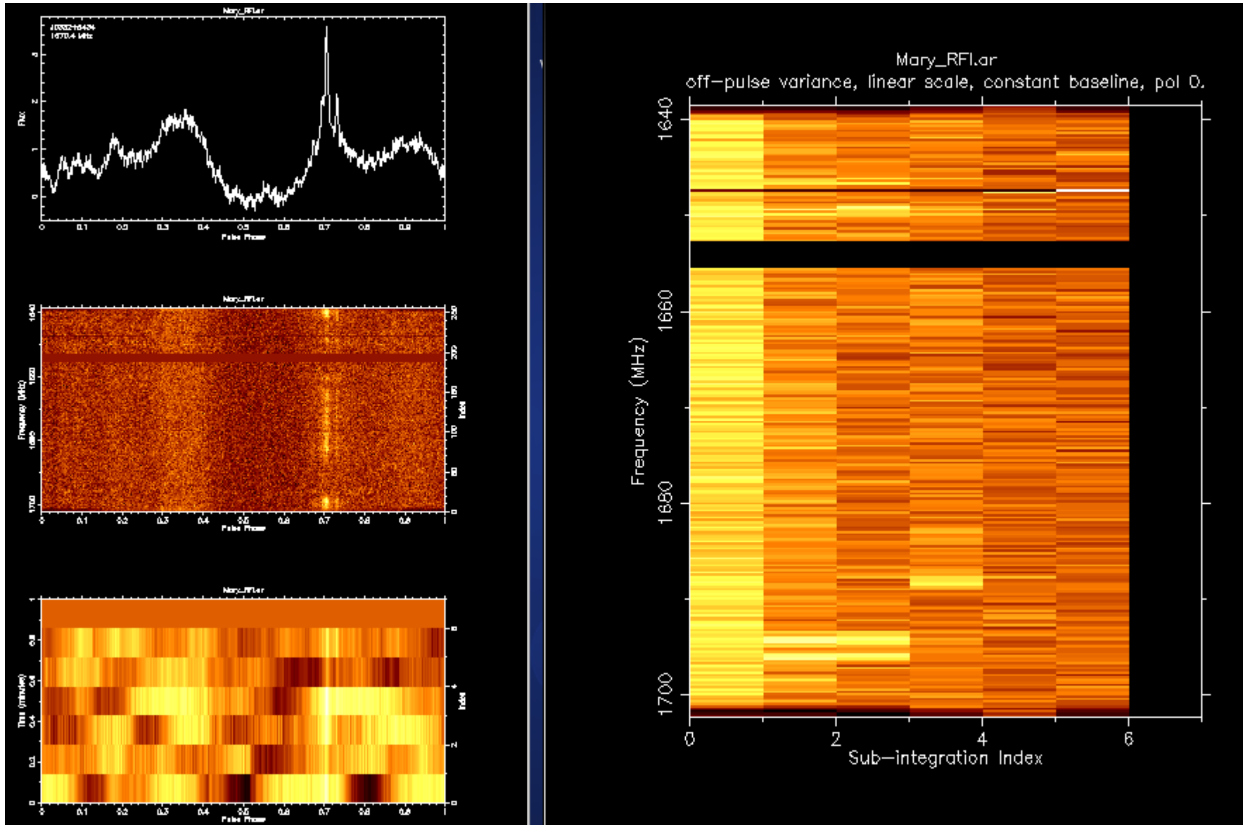
Masking isn't always perfect

It all depends on what you want to do

1. **Hand-flagging is labour intensive**
   - RFI may vary with time
   - May need to search hundreds (thousands) of plots!
   - Can't mask every candidate by hand

2. `psrzap` **is for pulsars**
   - what about FRBs?

3. `psrzap` **doesn't always work!**
   - Wobbly baseline still there

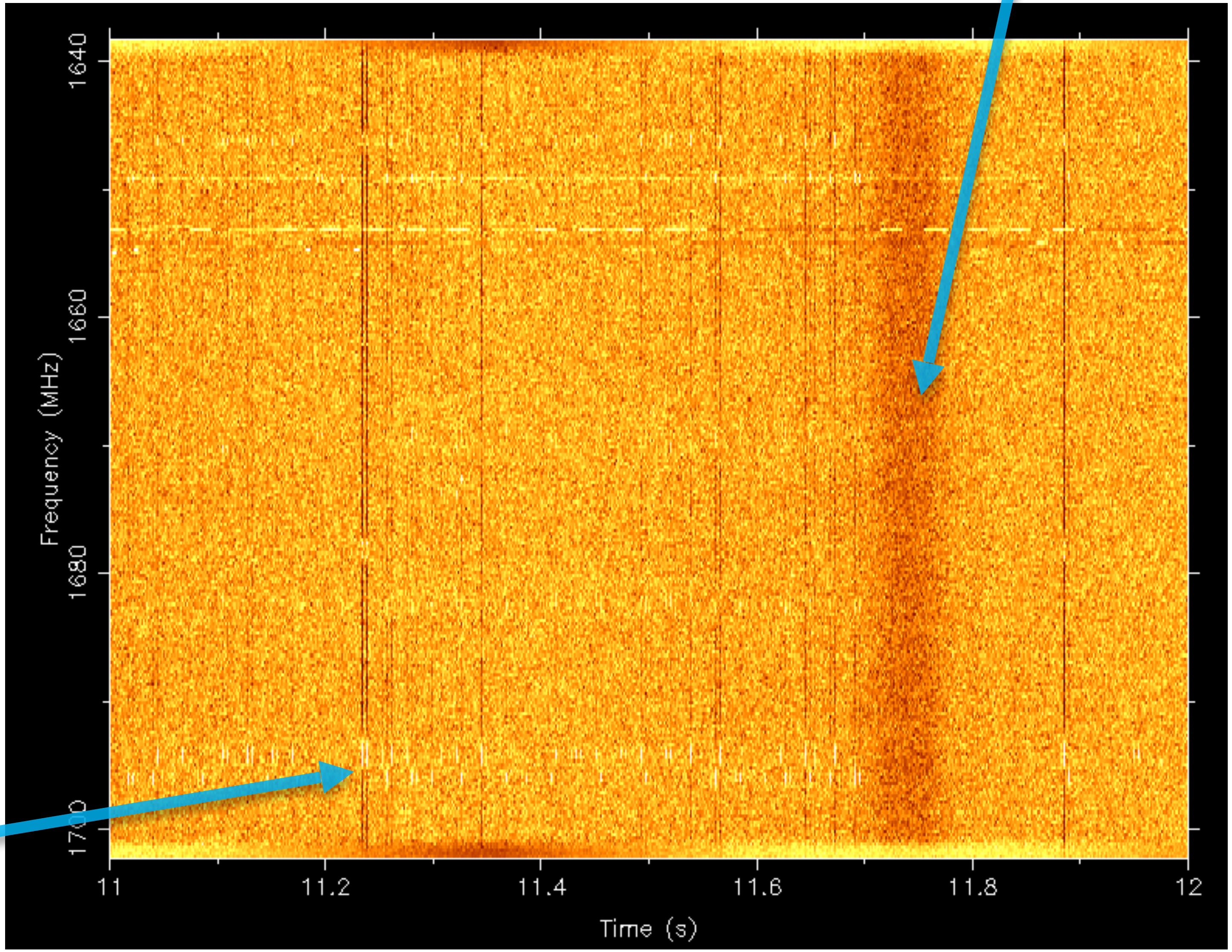


Lets look more closely

It always helps to look at your filterbank data

- What's causing that wobbly baseline?
  1. Look at the filterbank file
  2. Data displays broadband, impulsive RFI
  3. Causes a mess when folded

Also RFI. This is not the pulsar!!!

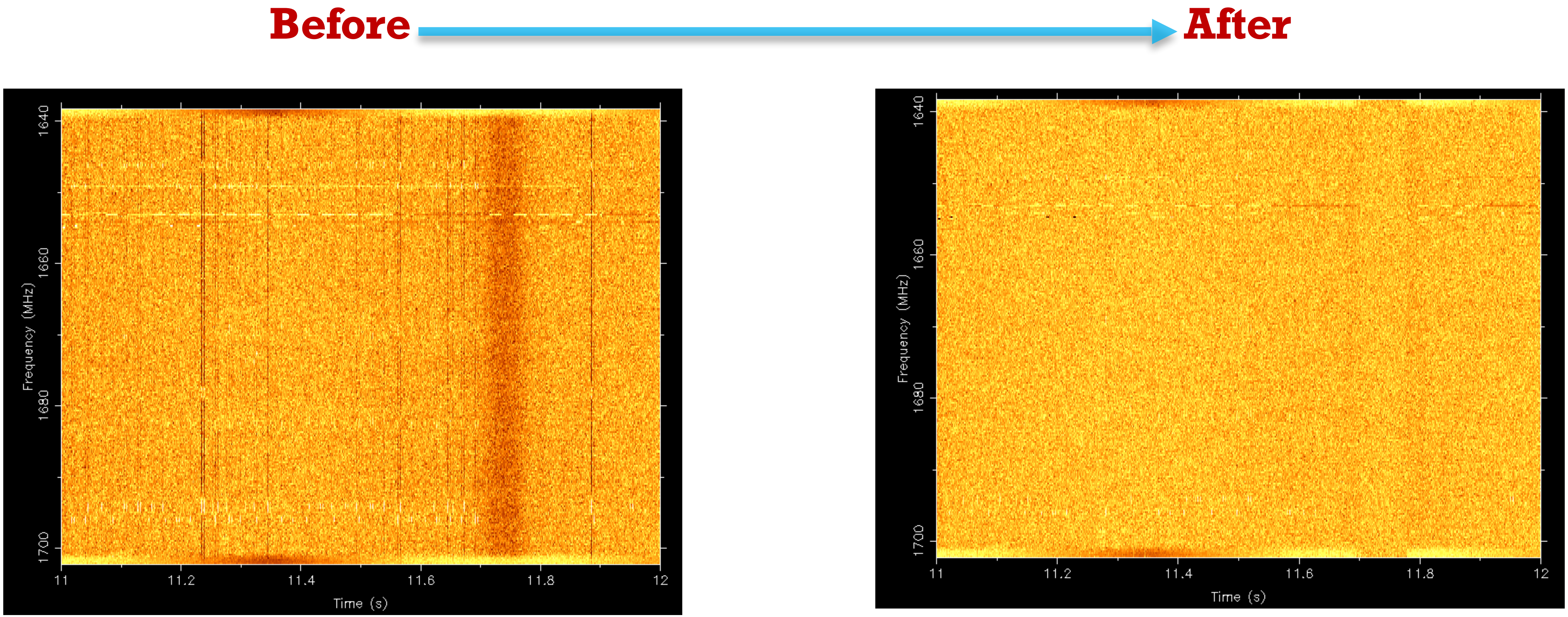


Bright signal saturates entire timesample

Example 3: ClipPy

Assume strong deviations are RFI. Clip them.

- Identify and replace the saturated timesamples
  - Using, e.g., `ClipPy` software [9]
  - **WARNING**: clipping could remove very strong low-DM signals! [16]



Mitigation example: ClipPy

Compare pre- and post-clipped observation

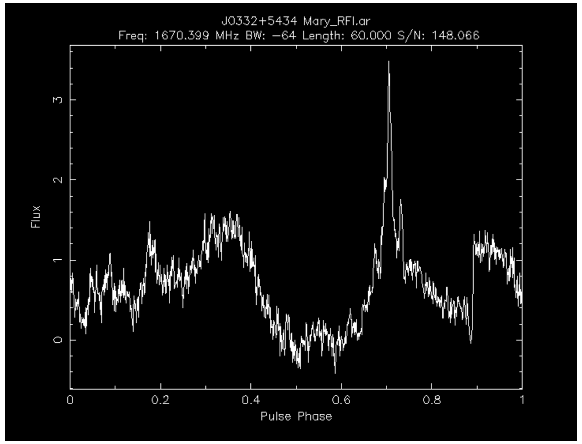
Before

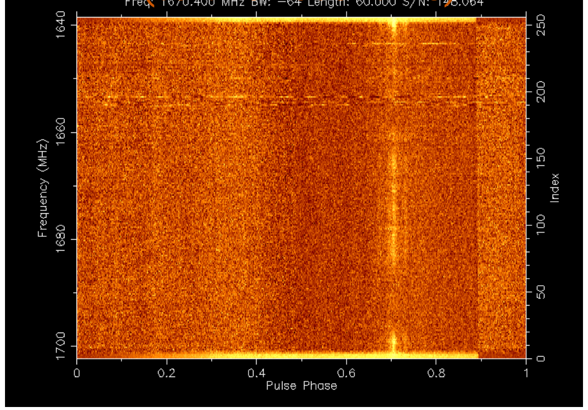
After

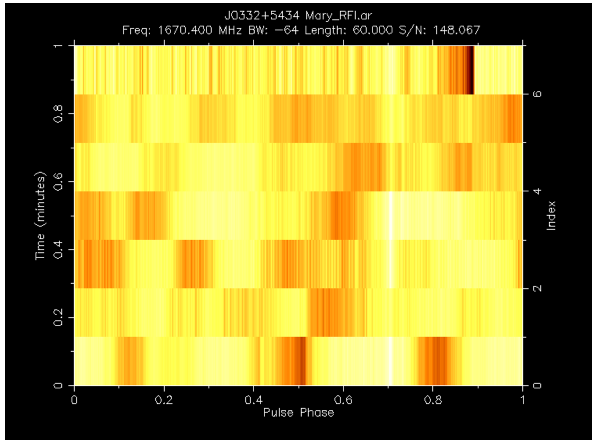
Pulse profile
(scrunch over time, frequency, polarisation)

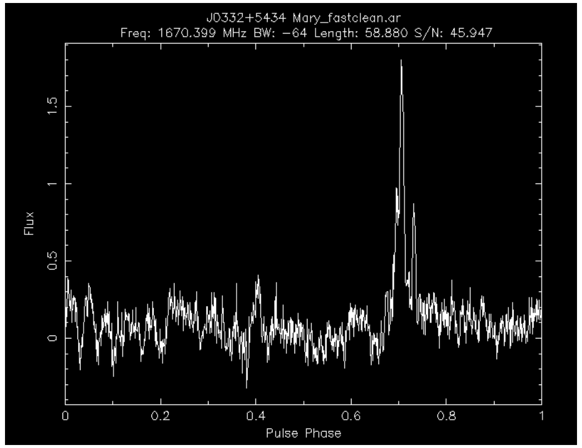
Pulse phase vs frequency
(scrunch over time)

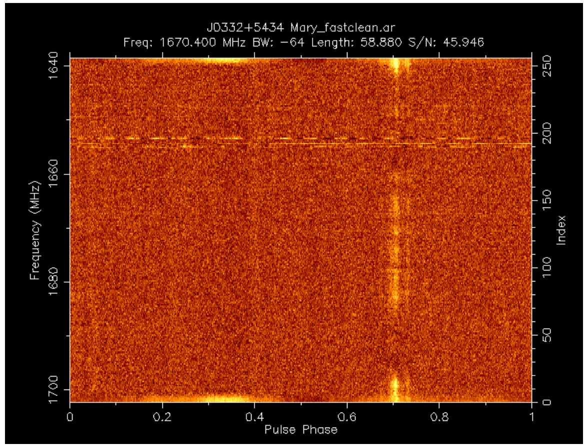
Pulse phase vs time
(scrunch over frequency)

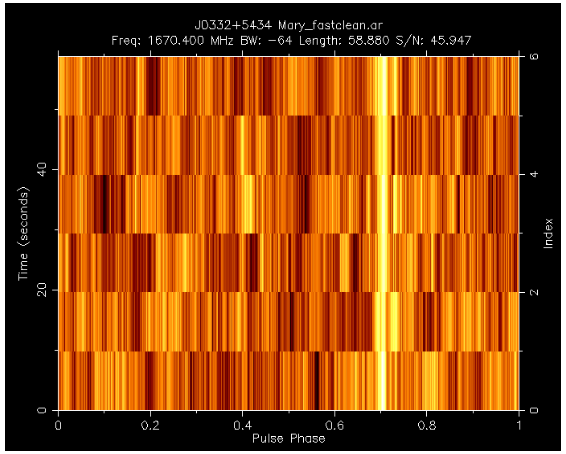








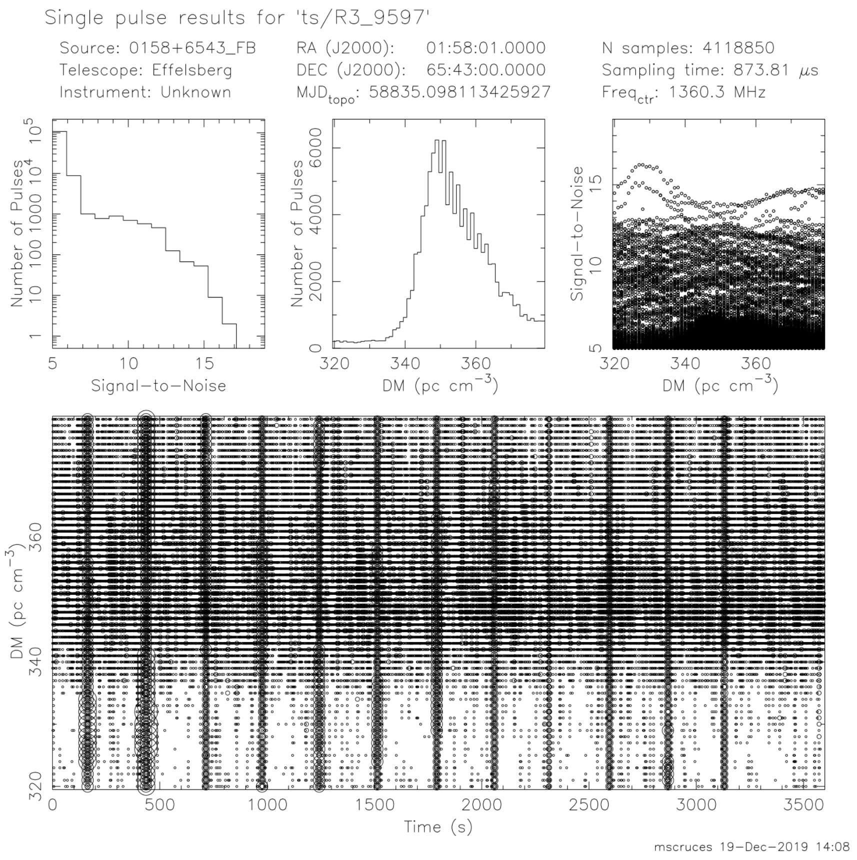


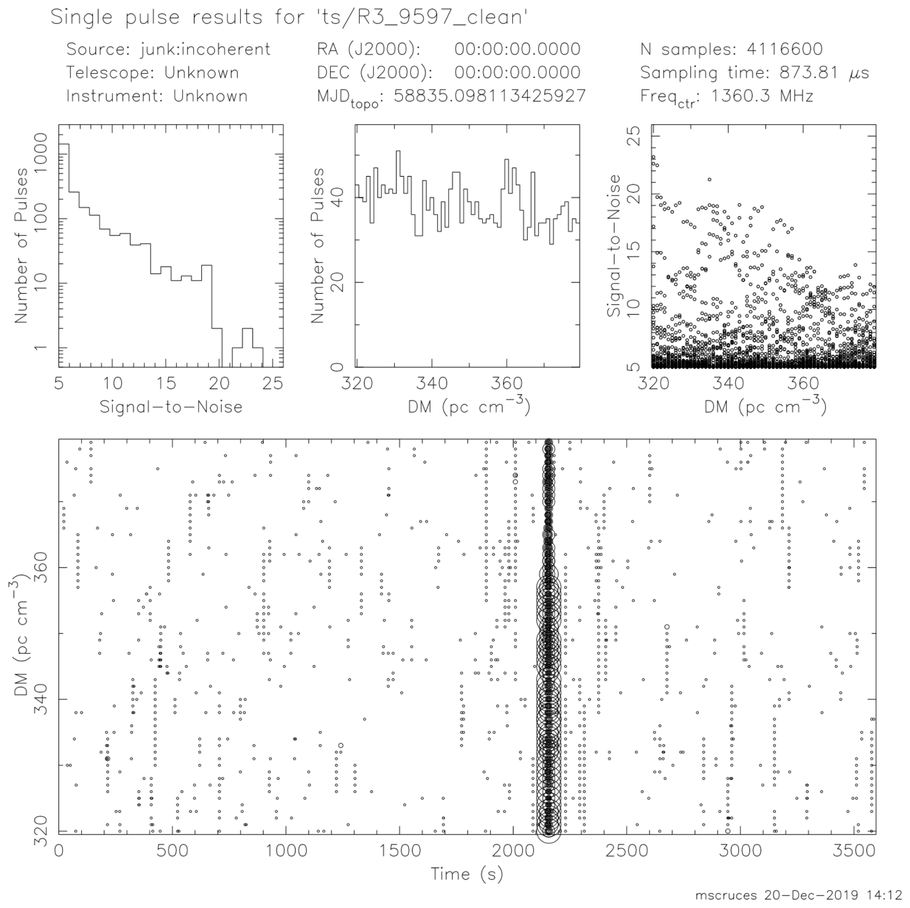


Mitigation example: ClipPy

Compare pre- and post-clipped single pulse candidates

Before → **After**





Food for thought: MAD

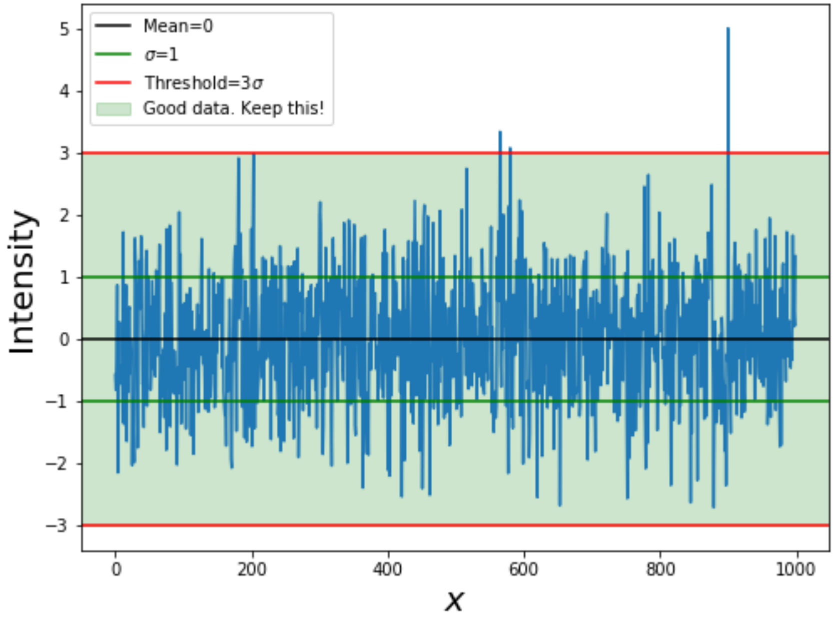
How do you decided what data to replace?

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(x_i - \langle x \rangle)^2}$$

1. Automated programs remove data using thresholds
   - Else, how do you distinguish RFI from astrophysical signal?

2. Algorithm:
   a) Calculate a **statistic** for your **dataset** $x = \{x_0, x_1, \ldots x_N\}$
   b) Define a **threshold**
   c) Replace samples with values > (**threshold** * **statistic**) away from the **mean**

3. Often the standard deviation, σ, is used
   - Large-valued outliers in datasets heavily weight σ due to the squared term
     - Might accept more rubbish
   - The Median Absolute Deviation (**MAD**) may be a more robust statistic
   - MAD has been used in, eg., [10]



$$MAD = median(\left| x_i - median(x) \right|)$$

Another technique: Zero-DM filtering

Could be preferable to clipping in some cases

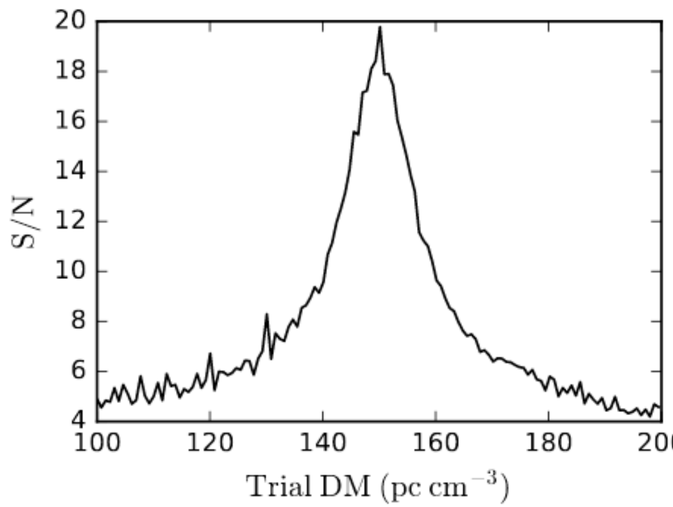
Principle behind this technique…

1. Astrophysical signals travel through ionised media
   - Acquire a dispersion measure (DM)

2. When searching, signal S/N will peak at true DM
   - RFI S/N will peak at DM = 0

3. The zero-DM filtering algorithm [16]:
   - Calculates DM=0 mean for a timesample
     - Scrunches across frequencies
   - Subtracts from frequency channels in timesample
     - Should get rid of RFI in data, and only slightly modify true pulse shape
   - **Pro:** doesn't remove strong astropysicals signals (like clipping)
   - **Con:** ineffective for low dispersion sources (e.g. in high frequency/low-bandwidth searches)

4. Often an option tou can select in programs
   - `PRESTO`: `-zerodm`
   - `dedisperse, SEEK, DESTROY` : `-subzero`
   - `dedisperse_all` : `-zerodm`
   - `Heimdall` [17]: `-rfi_no_broad`



No RFI: mean (ideally) = 0

RFI: mean != 0

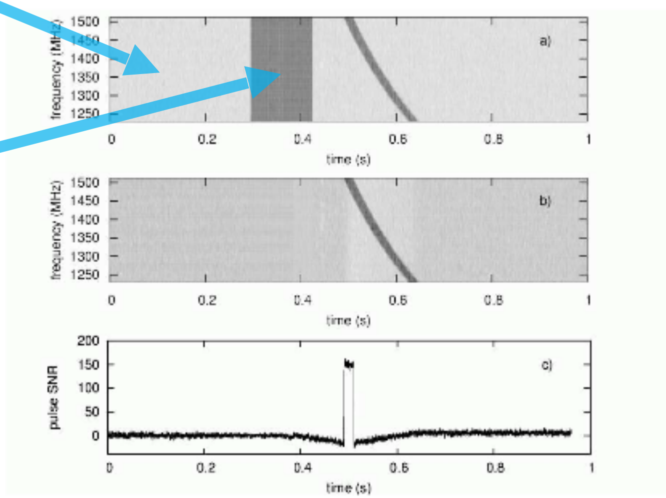
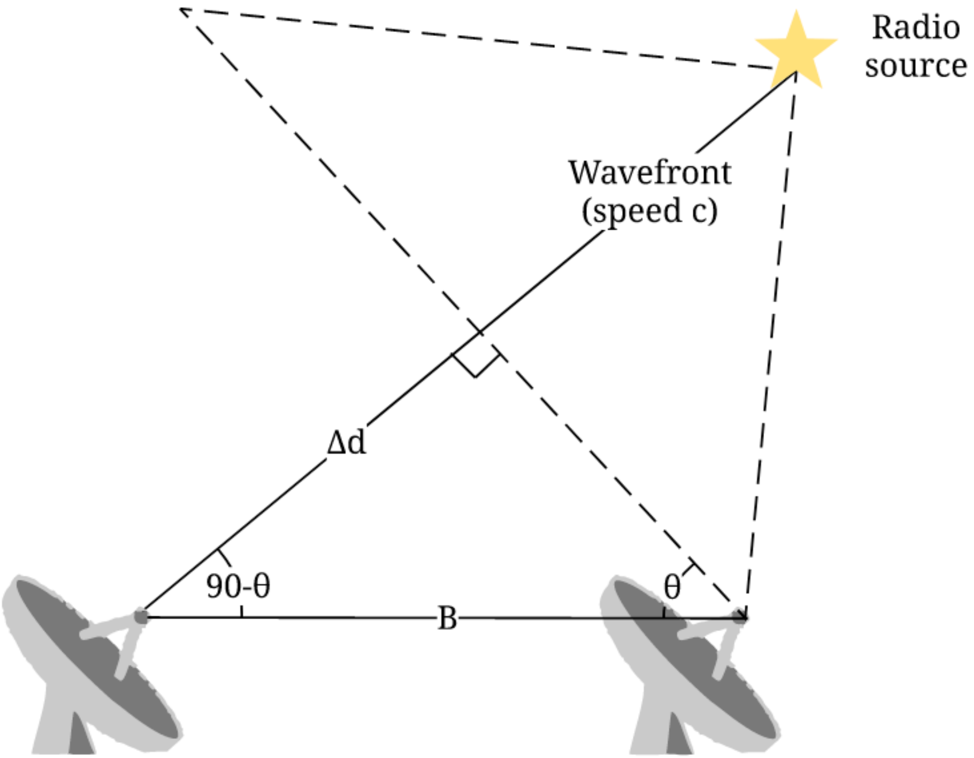


Figure 4. The top panel (a) shows a grey plot of simulated data of a 130-ms burst of broadband RFI with DM=0 cm^{-3}pc followed by a 20-ms dispersed pulse with DM=150 cm^{-3}pc across a 288MHz bandpass centered at 1374 MHz. The middle panel (b) shows the same data after application of the zero-DM filter. Note the negative non-pulse area. In panel (c), the data have been dedispersed for a DM=150 cm^{-3}pc and summed in frequency giving the pulse profile.

Finally: (Anti)Coincidencing

Leveraging multiple dishes

- Do you have multiple telescopes/beams?
  - Try coincidencing!
    - If beams pointed in same direction…
    - … and signal is seen in N telescopes…
    - … it is probably real! ☺️
      - Combine data! S/N should go up!
  - Or anti-coincidencing!
    - If beams are pointed in different directions…
    - … and signal is seen in N telescopes…
    - … it is probably local RFI ☹️
      - Ignore it



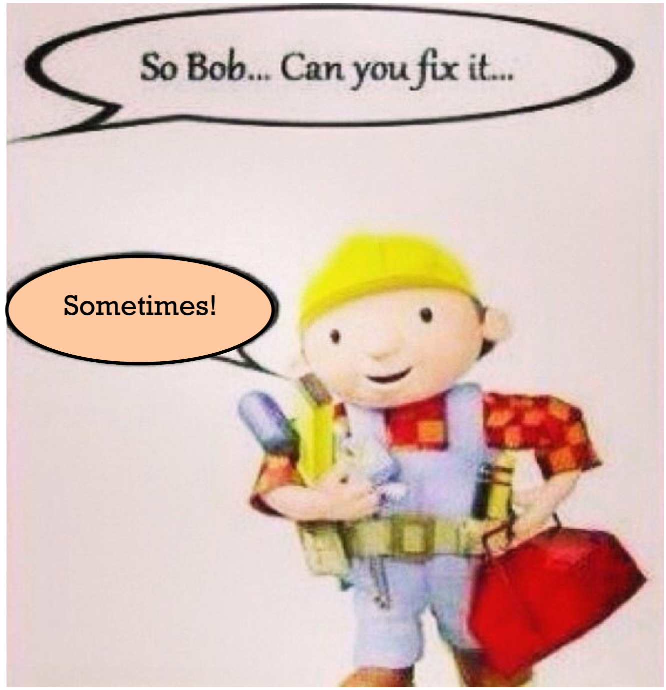
Conclusion

- To conclude the talk…

Conclusion

References are on next slide

1. RFI is *unwanted* non-Gaussian signal in your data
   - There is lots of it
   - Situation is probably getting worse
2. Ideally treat the cause of RFI…
   - Find sources of emission, stop them!
   - Usually challenging, sometimes impossible
3. … You might have to treat the symptoms.
   - Techniques include:
     - Masking, flagging, clipping, MAD, zero-DM, coincidencing
   - Many software solutions:
     - `PSRCHIVE`, `PRESTO`, `ClipPy`, many single-pulse search options…



Thank you!
Questions (or suggestions?)

References

- [1] Dicke, R. H., 1946: https://ned.ipac.caltech.edu/level5/Wall/Wall4.html
- [2] Lorimer D. R., Kramer M., 2012, Handbook of Pulsar Astronomy
- [3] Petroff, E., et al., 2015: https://arxiv.org/abs/1504.02165
- [4] https://en.wikipedia.org/wiki/Spectrum_management
- [5] http://psrchive.sourceforge.net/
- [6] van Straten et al., 2012: https://arxiv.org/abs/1205.6276
- [7] https://www.cv.nrao.edu/~sransom/presto/
- [8] Ransom, S., 2011: https://arxiv.org/abs/1703.05581
- [9] https://github.com/mbcxqcw2/
- [10] Cameron, A. et al., 2017: https://arxiv.org/abs/1703.05581
- [11] http://dspsr.sourceforge.net/
- [12] van Straten, W., Bailes, M., 2010: https://arxiv.org/abs/1008.3973
- [13] http://sigproc.sourceforge.net/
- [14] Lorimer, D., R., 2011: https://ui.adsabs.harvard.edu/abs/2011ascl.soft07016L/abstract
- [15] https://github.com/evanocathain/destroy_gutted
- [16] Eatough, R., P., et al, 2009: https://arxiv.org/abs/0901.3993
- [17] https://sourceforge.net/p/heimdall-astro/wiki/Home/